On the Relation between Acceptability and Frequency

Aria Adli

The aim of this work is to lay out the relation between the two main sources of linguistic evidence, acceptability judgments and frequency of occurrence in spontaneous speech. Tying acceptability to grammar and frequency to usage, the empirical relation between these data sources is seen as a manifestation of the underlying relation between grammar and usage. Contrasting with the functionalist view, I argue in favor of a clear distinction between grammar and usage, and contrasting with the standard generative view, I propose that both grammar and usage are constitutive parts of human syntax. They are related in a very specific way and their interplay is crucial in syntactic variation and the process of grammatical change. A systematic experimental study has been conducted in which acceptability and frequency data are obtained from the same subjects and on the same linguistic phenomenon: which is the well-known interaction between the semantic role of the *wh*-element and the preferred position of the subject (preverbal vs. postverbal) in Spanish. The results show that acceptability is not a sufficient but a necessary condition for usage. Constructions that are acceptable but not or hardly used are called *latent constructions*. It is suggested that latent constructions can occur as an intermediate step in the process of syntactic change.

Keywords: grammar, usage, Spanish, *wh*-questions, subject position

1. Introduction

The issue of linguistic evidence is one of the elements that characterize the paradigmatic divide between linguistic subfields. While variationist sociolinguists consider spontaneous speech data as the most reliable source of evidence, syntacticians traditionally see introspective data as the most reliable reflection of human language faculty. Labov's (1996) "validity principle" expresses the critical attitude of many sociolinguists towards judgment data (p. 83): "When the use of language is shown to be more consistent than introspective judgments, a valid description of the language will agree with that use rather than with intuitions". He lists five conditions that "promote the failure of linguistic intuitions", namely (i) social intervention, (ii) physical collapse, (iii) semantic suspension, (iv) cognitive strategies and

(v) pragmatic opacity. Schütze (1996) has shown that introspective data do come along with a number of methodological challenges and potential artifacts. However, general conclusions on the quality of the data as such are in my opinion too radical. The present work proposes a strict experimental protocol to deal with these challenges, applying a gradient acceptability judgment test.

According to a classic position in generative syntax, grammatical intuition is the most direct reflection of competence (Chomsky 1965). At the same time, Chomsky (1965) tied the notion of competence to a binary notion of grammaticality, while gradience is tied to his notion of acceptability and is supposed to incorporate aspects of performance. The exclusion of gradience from core syntax (which he had still admitted in Chomsky 1964) has never been really operational and led to what Sorace & Keller (2005: 1498) describe as an "artificial classification into acceptable and unacceptable examples". Gradience came back through the backdoor expressed by symbols like "?" or "??" as we can already witness in Ross (1967). Both gradient judgments and language use have been seen as less appropriate approaches from the early days of generative syntax. Paradoxically, the issue on the relation between introspection and production brought me to work with exactly these two types of data.

However, the boundaries have become somewhat more permeable in the last years. We witness an increasing use of spontaneous speech data in grammar research, both in the functionalist and generative tradition (e.g. Kroch 1989; Payne 1990; Bybee & Hopper 2001; Elsig 2007; Enghels 2009), and also some studies using judgments in sociolinguistics or at the cross-border of syntax and sociolinguistics (Wilson & Henry 1998; Adli 2004; Cornips & Poletto 2005; Gervain & Zemplén 2005; Henry 2005).

The relation between introspection and spontaneous speech data, or between judgment and production, is a core methodological issue in linguistics. Many scholars will probably agree that a promising approach consists

in data complementarity, i.e. in the combination of various data sources (Kepser & Reis 2005). Furthermore, many will probably agree that the two main data sources, introspection and production, are correlated. However, the question on the precise nature of this correlation in syntax is still open: For example: Are introspection and judgment correlated in a straightforward, linear manner, meaning that the more acceptable a construction the more often it is produced? Do we find a qualitative split between acceptable und unacceptable constructions, maybe of the sort that only acceptable constructions show a correlation between judgment and production, while suboptimal constructions are not produced in spite of our capacity to reliably express nuances in acceptability among them? And if so, what about the "grey zone" of marginal acceptability? Studies in phonology have shown that low-frequency phonotactic sequences that do not seem to violate any prominent rule are also low in acceptability (e.g. Pierrehumbert 1994; Bailey & Hahn 2001). A similar correlation has also been reported by Bybee & Eddington (2006) for lexical verb+adjective combinations in Spanish. However, we lack more thorough knowledge on this correlation in syntax, in particular we lack empirical designs in which we compare both data sources obtained from the same speakers, which is what this study intends to provide.

The questions on the nature of the correlation are related to another methodological issue: We have been observing in the last years the emergence of several theoretical approaches to *gradience*. They acknowledge and formalize the observation that we do not only make a binary distinction between acceptable and inacceptable constructions, but that we can observe nuanced, though systematic differences between degrees of acceptability. The nature of frequency data is inherently metrical (it conveys information on *how often* a particular construction occurs, not just if it is present or absent). In order to make a meaningful comparison with introspective data, one needs to work with an equally metrical data source, i.e. gradient judgments.

A new direction of work has been developing in the last years in which gradience is considered as an integral part of syntax and for which one needs to account on a formal level (e.g. Müller 1999; Keller 2000; Coetzee 2004; Gutiérrez-Bravo 2006). Many of these authors work within a modified OT syntax framework. Another theoretical variant, namely stochastic OT (Boersma & Hayes 2001) also integrates data from usage. There are also authors who do collect empirical gradient data, but interpret the results within the standard generative, binary view of grammaticality (e.g. Lotfi 2003; Trutkowski et al. 2003; Kim & Yoon 2009).

The larger readiness to make use of less traditional types of data in syntax has been accompanied by a blurring of the distinction between language knowledge (or competence) and language use, together with the rise of what Langacker (1987) called *usage-based model*. The blurring between grammar and usage goes back to Labov's (1969) concept of variable rules as well as to ideas developed by generative semanticians (Lakoff 1974). The increasing availability of electronic corpora data has been recently contributing to this trend.

However, there are also voices critical to this trend. Newmeyer (2003) wants to revitalize the divide and argues that grammar and usage have to be clearly distinguished ("grammar is grammar and usage is usage", p. 702). He argues against the stochastic nature of grammar and against the functionalist view that grammar is a fragile, temporary object. He writes: "No generative grammarian ever claimed that sentences generated by the grammar should be expected to reveal directly what language users are likely to say. This must be true for an obvious reason, namely, that knowledge of grammatical structure is only one of many systems that underlie usage (p. 692)." We have recently witnessed a vivid debate on this issue in the journal *Language* which I cannot repeat for reasons of space (Gahl & Garnsey 2004; Clark 2005; Guy 2005; Laury & Ono 2005; Meyer & Tao 2005; Newmeyer

2005; Bybee 2006; Gahl & Garnsey 2006; Newmeyer 2006a, b; Guy 2007). What is clear is that this issue is highly controversial among linguists.

The point I want to make in this paper is that it does make sense to maintain a clear distinction between grammar and usage. However, we also need more data from experiments in order to advance in this discussion. Knowledge and usage are constitutive elements of human language, none of them being more "epiphenomenal" than the other. Furthermore, I assume that the process of syntactic change cannot be satisfactorily modeled without taking into account both concepts. What is needed, is an empirically corroborated model for the relation between grammar and usage (see also the discussion in Boye & Engberg-Pedersen 2010). Assuming that the grammar/usage pair parallels the competence/performance pair, a strict division of labor between competence-oriented and performance-oriented linguistic research unnecessarily limits our capability of dealing with some fundamental issues of linguistic evidence, which can best be dealt with in a synergetic way across linguistic subfields.This is a large enterprise and it will require a number of studies on various syntactic phenomena in which judgment and production are correlated. The present paper intends to take a few steps in this direction and to formulate initial hypotheses. To this end, results from both judgment and production on the relation between the semantic role of the *wh*-element and the position of the subject in Spanish interrogatives are presented. Before laying out the essentials of this syntactic phenomenon, I want to present the methodology, which is a crucial element in relating judgment and production.

2. Combining various data sources: the *sgs* database

The data is taken from the Spanish part of the *sgs* database (***s***peech production - ***g***rammaticality judgments - ***s***ocial data), an ongoing project in

the scope of which I am building up a multilingual corpus presently covering French, Persian, Spanish, and Catalan. It includes and combines three elements: syntactically tree-annotated transcriptions of spontaneous speech, gradient grammaticality judgments, and social background information of speakers. All three data types were collected with each subject in an integrated test-suite. With regard to the Spanish part, the sample consists of 54 native speakers of the Spanish variety of Catalonia (see Payrató 1985; Sabater 1991; various papers in Sinner & Wesch 2008). The sample is approximately balanced between women (60%) and men (40%). Participants have at least completed compulsory education, but most of them have a level of 3 or higher on the International Standard Classification of Education (UNESCO 2006), which roughly corresponds to two additional years after compulsory education (Spanish *Ciclos Formativos* or *Bachillerato*). Their age ranges from 17 to 48 (mean: 28 years). Excluding the data of one person who aborted the test session prematurely, the analysis is carried out with N = 53.

2.1 Syntactically annotated spontaneous speech

Particular attention has been paid to obtaining a corpus with a substantial proportion of interrogative sentences. The problem is that interviewees hardly ask questions in standard interview situations.[1] In order to elicit interrogatives, a game task was developed, in which the subject is *required* to ask questions (he/she investigated a fictive murder case). As a result, approximately 80% of the constructions are interrogatives. In a second step, the recordings have been fully transcribed and sentence-wise time-stamped. In a third step, the transcriptions were syntactically annotated. Some theoretical entities such as essential instances of movement are included in the annotation, which corresponds to a reduced tree structure.[2]

The 53 subjects produced in sum 10778 sentences (defined as one matrix clause with all its embedded clauses), of which 6195 are non-elliptical

constructions. They, in turn, contain 1183 true information questions with local *wh*-movement. Finally, 384 of them are questions with an overt (non-*wh*) subject, which is the set of target constructions (a few instances of *wh*-in-situ constructions and constructions with long-distance subject topicalization have been excluded beforehand).

2.2 Gradient acceptability judgment test

The experimental procedure was a computer-based version of the paper and pencil-based gradient grammaticality judgment test in Adli (2005). It has been coded in the programming language *php* and runs in a normal browser window. Subjects read a test sentence and gave a nuanced judgment on a *graphic rating* scale, which allowed for the expression of fine nuances in a very intuitive way. The computer-based version of the test is easier to handle than drawing lines with ruler and pencil. Subjects were placed in a quiet room in front of a notebook computer.

They drew a line on the computer screen by dragging a cross on a horizontal slider, which ran from 0 to 100 (the actual value was always displayed under the slider). The length of the line corresponded to the degree of acceptability. At the onset, subjects judged a reference sentence.

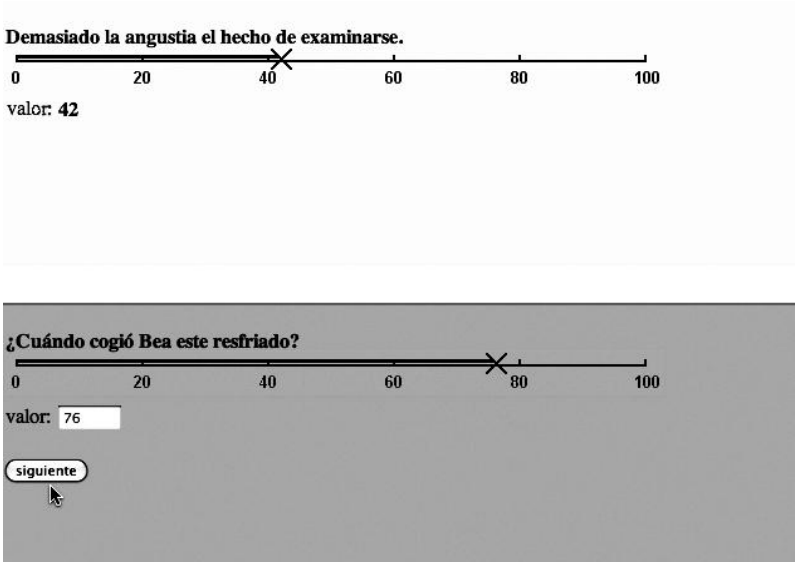

Figure 1: Gradient acceptability judgment test

The participants went one by one through the test sentences presented in a randomized order in the lower part of the screen, while the judgment of the reference sentence remained visible in the upper part of the screen. The following sentence was displayed upon clicking on the "next" button, and they could not go back. The reference sentence was a suboptimal construction that mostly received a judgment in the intermediate scale range.[3] It represented an intermediate scale anchor, which generally helped to improve the precision of the instrument. Thus, subjects could find the degree of acceptability of each test sentence by estimating its distance from the left (clearly unacceptable) and right (obviously acceptable) endpoint, as well as comparing it to the previously given judgment to the reference sentence.

The test began with an instruction and training phase during which they learned and practiced the notion of gradience (as opposed to a categorical distinction between grammatical and ungrammatical). They also learned to exclude from their considerations aspects that are supposed to be irrelevant to the syntax of the construction (e.g. broader pragmatic plausibility, opaque ideas of "elegance", etc.).

On average, the instructions lasted 15 minutes and the experimental phase 20 minutes (the experiment included constructions other than *wh*-interrogatives which are however not subject of this paper).

3. Analyzing the data

3.1. Judgments

The experimental material consists of simple *wh*-questions with an overt subject.

(1a) ¿A quién estos libros le gustan?
to whom these books DAT-CL $like_{3pl}$
(1b) ¿A quién le gustan estos libros?
to whom DAT-CL $like_{3pl}$ these books
'Who likes these books?'

(2a) ¿Qué Juan compró ayer?
what Juan bought yesterday
(2b) ¿Qué compró Juan ayer?
what bought Juan yesterday
'What did Juan buy yesterday?'

(3a) ¿Cuándo Marina hizo esta tarta?
when Marina made this cake
(3b) ¿Cuándo hizo Marina esta tarta?
when made Marina this cake
'When did Marina bake this cake?'

(4a) ¿Por qué Pedro cerró la tienda?
why Pedro closed the shop
(4b) ¿Por qué cerró Pedro la tienda?
why closed Pedro the shop
'Why did Pedro close down the shop?'

The design contains two independent variables. Variable A distinguishes between four different semantic roles of the *wh*-element (EXPERIENCER, THEME, TIME, REASON), henceforth called *wh*-role. Variable B differentiates preverbal and postverbal subject position.[4] The relevant constructions are given below. Note that there is no repetition of nouns and verbs in the experimental material. Since each construction is presented in two lexical variants, every subject has to give 8 x 2 = 16 different judgments. The dependent variable is the arithmetic mean of the acceptability values of the two lexical variants.

A: semantic role of *wh*-element / B: subject position	EXPERIENCER	THEME	TIME	REASON
preverbal	(1a)	(2a)	(3a)	(4a)
postverbal	(1b)	(2b)	(3b)	(4b)

Table 1: ANOVA design with variables A "*wh*-role" and B "subject position"

One well-known property in Spanish is the fact that preverbal subjects cannot occur with certain kinds of constructions such as *wh*-questions, or they are the marked variant compared to postverbal subjects, as with wide focus declaratives with unaccusative verbs (Arnaiz 1997), or with sentences with narrow information focus on the subject (Bolinger 1955; Contreras 1976). Numerous observations have been made with respect to the restrictions of preverbal subjects in Spanish *wh*-questions, to which I cannot do justice here (Meyer 1972; Torrego 1984; Goodall 1991; Suñer 1994; Suñer & Lizardi 1995; Uribe-Etxebarria 1995; Olarrea 1996; Inclán-Nichol 1997; Ordóñez 1997; Baauw 1998; Ordóñez 2000; Toribio 2000; Zubizarreta 2001; Gutiérrez-Bravo 2006; Beas 2007). The discussion turns around the question what types of *wh*-constructions allow preverbal subjects and what types do not. In essence, markedness of preverbal subjects in *wh*-questions depends (i) on properties of the *wh*-element (see below), (ii) on the respective variety or dialect, and (iii) on the position of the *wh*-clause, whether it is a matrix or a subordinate construction. Some authors have argued that the relevant property of the *wh*-element is its semantic role (Bakovic 1998; Gutiérrez-Bravo 2006). I adopt this idea in this work (alternative approaches are centered on other aspects of the wh-element, such as the argumental vs. adjunct status, Torrego 1984; or its specificity, Beas 2007).

Semantic roles are organized in a hierarchy with some roles being more prominent than others, affecting numerous grammatical properties such as case marking, mapping on argument structure, word order, and also subject position in certain types of Spanish constructions including *wh*-questions (Larson 1988; Speas 1990; Pesetsky 1995; Bakovic 1998; Gutiérrez-Bravo 2006). With regard to the semantic role of the *wh*-element, Bakovic (1998: 37) formulated the hierarchy in (5). Roughly, the less prominent a semantic role, i.e. the more it appears to the right in (5), the "easier" it is to place the subject in preverbal position.

(5) ARGUMENT > LOCATION > MANNER > REASON

The idea is that each variety has its own "cut-off point" on the scale. For example, if a variety allows preverbal subjects in questions with *wh*-object arguments (which, according to Suñer 1994 is the case in Puerto Rican Spanish), then it allows preverbal subjects with all other *wh*-elements. Or if a variety allows preverbal subjects with *wh*-location elements but not with *wh*-object arguments, then it will also allow them with *wh*-manner and *wh*-reason elements (as has been suggested by Torrego 1984 for Iberian Spanish). Moreover, Bakovic (1998) observes that within the same variety the cut-off point can be lower for subordinate clauses than for matrix clauses, meaning that subordinate clauses can be more permissive with respect to preverbal subjects but not vice-versa. For example, Mexican Spanish, which is on the non-permissive side (Gutiérrez-Bravo 2006: 148), does not allow any *wh*-construction with a preverbal subject, except for subordinate constructions with *wh*-reason elements.

Gutiérrez-Bravo (2006: 75) proposes a slightly more differentiated thematic hierarchy given in (6), which is a compilation of the (essentially similar) hierarchies proposed by Larson (1988: 382), Speas (1990: 16), and Bakovic (1998: 37). All these hierarchies have in common that EXPERIENCER is more prominent than THEME, THEME more prominent than LOCATION, TIME, or MANNER, and any semantic role more prominent than REASON.

(6) AGENT > EXPERIENCER > THEME > LOCATION > MANNER|TIME > REASON

The experimental design in Table 1 (variable A) is based on a presumed hierarchy as in (7), which is a reduced version of the hierarchy proposed by Gutiérrez-Bravo (2006).

(7) EXPERIENCER > THEME > TIME > REASON

The following research questions are asked: First, where is the cut-off point for the Spanish variety of Catalonia with regard to the hierarchy in (7)? Second, do we find a simple binary separation of the judgments for the constructions with preverbal subject into two groups (acceptable vs. unaccepta-

ble)? Or do we find a monotonic slope along the hierarchy itself, i.e. constructions with *wh*-EXPERIENCER having the lowest and constructions with *wh*-REASON having the highest degree of acceptability? The general issue behind this second point concerns the debate whether grammar produces a binary distinction (corresponding to the standard view both in generative syntax and OT syntax), or whether grammar produces gradience and markedness.

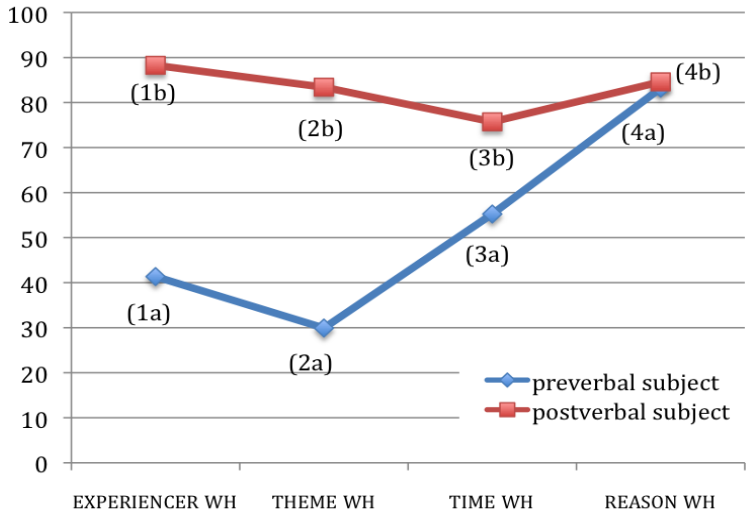

Figure 2: Gradient acceptability values on a scale from 0 to 100

Figure 2 shows the mean values, higher values on the y-axis show higher degrees of acceptability. The results of a two-way analysis of variance reveal (i) a significant difference between the four *wh*-roles, (ii) a significant difference between preverbal and postverbal subject, and more importantly in the context of this discussion, (iii) a significant interaction between *wh*-role and subject position.[5] The differences between the *wh*-roles are subtle in constructions with postverbal subject, which can all be considered as unmarked. However, the differences are, as expected, salient with preverbal subject.

In a next step, the difference between the variant with preverbal and the one with postverbal subject is analyzed for each *wh*-role in simple main effect tests. In addition, the effect size is indicated by the partial η^2 value

(Cohen 1973), which expresses the amount of variance explained. This value is a more precise indicator than the difference between the mean values, because it also takes into account the dispersion of the individual judgments. These effect sizes can be interpreted as the "costs" that come along with preverbal subject. All *wh*-roles show an effect of subject position, except for the sentence pair (4a)/(4b) with REASON-*wh*. The size of the effect between the sentence pairs (1a)/(1b) with EXPERIENCER-*wh*, and (2a)/(2b) with THEME-*wh* is essentially identical (partial η^2 values are 0.878 and 0.881 respectively). On the other hand, the effect size for the pair (3a)/(3b) is less salient, though significant (partial $\eta^2 = 0.474$).[6]

Based on these results, it is difficult to support the idea of a cut-off point on the thematic hierarchy scale dividing ungrammatical from grammatical constructions in a binary manner. Rather we find degrees of markedness concerning constructions with preverbal subject. (4a) with a REASON-*wh* is fully unmarked. (3a) with a TIME-*wh* is somewhat marked. Finally, (1a) and (2a) are highly marked. (8) summarizes the results and shows the degrees of markedness correlated with the *wh*-role for constructions with preverbal subject.

(8) EXPERIENCER|THEME > TIME > REASON

3.2. Spontaneous speech data

Among the target set of 384 *wh*-questions with an overt (non-*wh*) subject extracted from the Spanish treebank of *sgs*, 237 (61.7%) have the subject in postverbal position (WH-V-S) as in (9a). But only as few as 2 (0.5%) have the subject in non-topicalized preverbal position (WH-S-V). These two instances are embedded questions (which allow preverbal subjects "more easily"). What is surprising is therefore the fact that *wh*-elements of the REASON type with preverbal subject are not more marked in the judgment test, but are basically not produced. However, 145 (37.8%) *wh*-questions have the subject in topicalized position (S-WH-V) as in (9c), which is thus

the only other option for subject placement that speakers make use of. A word order like (9c) was not part of the gradient acceptability test.

(9a) ¿Y qué te ha contado su compañera de piso?
and what you-CL has told his flat mate
'And what did his flat mate tell you?'

(9b) ¿Me podrías decir [a qué hora el electricista subió]?
me could say at what hour the electrician came up
'Could you tell me at which time the electrician came up?'

(9c) ¿Y esta chica cómo vino aquí al piso?
and this girl how came here to-the appartment
'And how did this girl come here to the apartment?'

Given that preverbal subjects can be considered as absent, I pursue the analyses with constructions with postverbal subject, and compare the data from spontaneous speech with the gradient judgment data with regard to the semantic roles EXPERIENCER, THEME, TIME, REASON.

A major challenge when working with spontaneous speech is the classification of arguments and adjuncts in categories of semantic or thematic roles, given that one encounters many examples that are anything but prototypical. The difficulties are two-fold. First, we are far from any consensus on a classification system, which may not be surprising given that numerous authors of various persuasions have discussed the issue of thematic roles. The roots of these notions go back to structuralists (e.g. Blake 1930). They have been taken up by Gruber (1965), Fillmore (1966) (in the context of his semantic deep cases), and they have been further developed by Jackendoff (1972, 1987). Identification of TIME and REASON is less intricate. These two roles are limited to *wh*-adjuncts such as *cuándo* 'when', *a qué hora* 'at which time/hour' and alike, or to *wh*-adjuncts of the type *por qué* 'why', *por qué razón* 'for which reason'. The greater difficulty concerns the notion THEME, requiring careful consideration of the semantic entailments of the verb. I use the notion THEME in the sense of Dowty's (1991) prototypical patient or P-Patient.[7] He proposes a system, which he reduces to two proto-

typical roles namely P-Agent and P-Patient. They are defined by a cluster of semantic entailments (which do not always have clear-cut boundaries). He suggests the following entailments with regard to P-Patient: (i) undergoes change of state, (ii) incremental theme, (iii) causally affected by another participant, (iv) stationary relative to movement of another participant, (v) does not exist independently of the event, or not at all (see Dowty 1991: 572-575 for examples). The clearest cases are direct objects of what Dowty (1991: 601) calls "highly transitive verbs", i.e. verbs which have a high number of P-Patient entailments for the object (and P-Agent entailments for the subject). All objects for which P-Patient entailments cannot be unambiguously identified are excluded from the set of tokens in the present study.[8] EXPERIENCER is an argument of what Postal (1970) calls psychological predicates. In Dowty's (1991: 577) cluster concept EXPERIENCER is a subtype of the P-Agent, namely "sentience without volition or causation" as in *gustar* 'like', or *encantar* 'delight'. Therefore, it has no overlap with the P-Patient definition.

Among the 237 *wh*-questions with the (non-*wh*) subject in postverbal position, 32 have a *wh*-element with the semantic role THEME, 27 with the role TIME, 3 with the role REASON, and none with the role EXPERIENCER. These results show that one of the semantic roles, namely REASON, occurs very scarcely, and that another one, EXPERIENCER, is fully absent in the relevant constructions.

4. Comparing judgment and spontaneous speech data

Let us now turn to the relation between acceptability and frequency, beginning with *wh*-questions with preverbal subject of the types (1a), (2a), (3a), and (4a). They are basically absent in spontaneous speech. The constructions (1a), (2a), and (3a) are (in different degrees) suboptimal. These results can hint at a principle that would essentially block suboptimal con-

structions from language production. However, we would at least expect the occurrence of (4a), the only unmarked *wh*-construction with preverbal subject. Nevertheless, there is just a single hit of (4a). These results contrast with two of the *wh*-questions with postverbal subject of the types (2b) and (3b), which do occur in spontaneous speech. Constructions of the type (4b) also occur, but scarcely (3 tokens). This finding is not surprising, given that *wh*-constructions with postverbal subject are all acceptable. The puzzling question is why constructions of the type (1b), (4a), and (4b) do not occur at all or occur very scarcely.

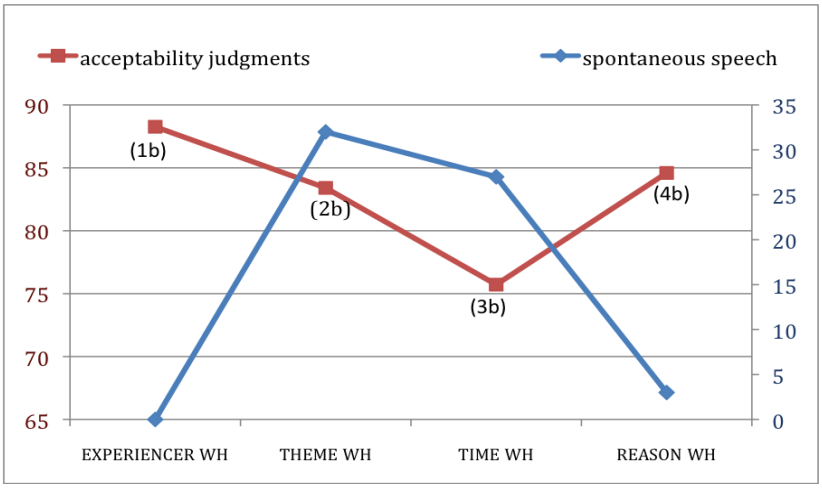

Figure 3: Acceptability and frequency for *wh*-questions with postverbal subject

The results for the constructions with postverbal subject are visualized in the double y-axis chart in Figure 3, in which the degree of acceptability and the frequency of occurrence are superposed. The y-axis on the left indicates the acceptability values concerning the line with square markers; the y-axis on the right indicates the frequency concerning the line with diamond-shaped markers.

Figure 3 seems to indicate a correlation between acceptability and frequency for the two constructions that occur in spontaneous speech: The (2b) type has a slightly higher degree of both acceptability and frequency compared to the (3b) type. Only future research with a wider range of phenome-

na can show whether nuances in acceptability are consistently reflected in nuances in frequency of occurrence.

However, Figure 3 shows a clear mismatch between the acceptability of constructions with an EXPERIENCER or a REASON *wh*-element like (1b) and (4b) on the one hand, and the scarcity or even total absence of such constructions in spontaneous speech, on the other. We find a total of 35 constructions with EXPERIENCER arguments within the set of non-elliptical sentences, but not one of them is a *wh*-element. With regard to REASON *wh*-elements, we find 6 tokens with the subject either topicalized or postverbal, but only one with preverbal (non-topicalized) subject.[9]

So far, we can state that whenever a construction is judged suboptimal as in (1a), (2a), and (3a), it is basically absent in language use. All constructions occurring in the spontaneous speech data are judged acceptable. However, constructions can receive high acceptability scores but hardly occur in spontaneous speech as is the case for (1b), (4a), and (4b). The more general question is why some acceptable constructions are basically absent in spontaneous speech data. Due to the low number of tokens, conclusions have to be drawn very carefully. It cannot be ruled out that constructions of the type (1b), (4a), and (4b) *are* in use, but could simply not be captured in the *sgs* database. One should bear in mind that the issue preverbal vs. postverbal subject is accounted for by the semantic role of the *wh*-element, i.e. the constructions differ in meaning. Therefore, one cannot exclude that the respective constructions are not detected in the database because there was no discursive necessity to make use of them during the interviews. This issue is part of a more general challenge in corpus-based syntactic research, which is that many theoretically relevant constructions are hard to observe. Compared to corpus-based research in many other linguistic domains (e.g. phonology), one needs very large corpora for drawing sound generalizations, especially on the absence from usage (see Keller & Lapata 2003, who try to make use of the world wide web as a corpus). The present picture is no ex-

ception: Out of 10778 annotated utterances in *sgs*, only 64 correspond to one of the eight types (1a) to (4b), although the database contains a substantial proportion of interrogatives. In my opinion, the numbers are not high enough to count as conclusive evidence. However, the observed tendencies can serve as a basis for formulating more precise hypotheses to be tested in future research, with very large databases. In this sense, I propose the hypothesis in (10).

(10) Acceptability is a necessary but not a sufficient condition for usage.

(10) contradicts Bybee & Eddington's claim (2006: 352) that "acceptability is a function of experience", because the latter does not capture the observed fact that some constructions are not used although they are acceptable.[10] Why should there be a relation between judgment and production in terms of (10)? One possible account is a cut-off model along the lines of Coetzee (2004). According to this idea, there is a cut-off point in acceptability, below which a construction is never produced. Although Coetzee's (2004) proposal deserves further inquiry, the mismatch between identical acceptability and different frequency with regard to REASON *wh*-questions with preverbal versus postverbal subject seems to indicate to a different direction (suggested with the same caution due to very scarce numbers). Taking up the distinction between grammar in the sense of competence and usage, I suggest a different hypothesis: Not all possibilities provided by grammar are used. I call a form that is available grammar-wise but not used a *latent construction*. If this idea is on the right track, it has also interesting implications for language change. A possible scenario that could account for certain cases of diachronic change is given in (11).

(11) (i) Construction *X* is not available in the grammar.
(ii) *X* becomes available, but is not used.
(iii) *X* is used as part of a set of optional syntactic variants.
(iv) Cases of unstable optionality are dissolved, i.e. only *X* survives.

The process in (11) can also be incomplete, i.e. it can stop at any stage and eventually regress. Stage (ii) means that *X* co-occurs (or competes) with one or more alternative structural variants. In case of a presumably stable optionality, the process remains stable at stage (iii). If *X* appears only as a temporary phenomenon and is not established, then the process runs from (i) to (iii) and back to (i). The idea of latent constructions can also be applied to the disappearance of certain forms, i.e. a form first disappears from usage before it disappears from grammar.

Meisel (in press) highlights the genuine link between language acquisition and diachronic change. The notion of latent construction is somewhat problematic from an acquisitional perspective, because it is by definition not present in the Primary Linguistic Data of the language learning child. I leave this issue for future research.[11]

Although (10) and (11) are still tentative, I think that they formulate questions that point to a promising direction of research. Obviously much more can be said on the scenario in (11), e.g. concerning the issue under which conditions each of the steps (ii), (iii), or (iv) are triggered, whether they show sensitivity to frequency and if so, how this sensitivity manifests.

5. Conclusion

The hypothesis in (10) and the idea of latent constructions is in line with Newmeyer's (2003) criticism of Lakoff & Johnson (1999) who assume that any generalization about usage has to follow from grammar. Newmeyer (2003: 692) argues that "knowledge of grammatical structure is only one of many systems that underlie usage". It is important to clearly differentiate the grammar system from usage, something that usage-based models ignore.

These two notions relate to different sources of linguistic evidence, namely judgment and frequency, and we need to further understand the relation between them. Future research should also analyze data on syntactic phenomena that can be considered as (semantically similar or equivalent) optional variants.

Should a threshold in acceptability be assumed that could account for presence and absence of constructions in usage? Is the correlation between acceptability and frequency linear or non-linear (e.g. S-curve-shaped)? By both distinguishing and correlating grammar and usage, we can also get closer to a satisfactory answer to the presumed contradiction between the stability of grammar systems on the one hand, and the temporary and provisional state of affairs in usage on the other – one of the enigmas of syntactic change.

References

Adli, Aria (2004). *Grammatische Variation und Sozialstruktur.* Berlin: Akademie Verlag.

Adli, Aria (2005). Gradedness and Consistency in Grammaticality Judgments. In Kepser, S. & Reis, M. (eds.), *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives.* Berlin: Mouton de Gruyter. 7-25.

Arnaiz, Alfredo R. (1997). An overview of the main word order characteristics of Romance. In Siewierska, A. (ed.), *Constituent Order in the Languages of Europe.* Berlin, New York: Mouton de Gruyter. 47-74.

Baauw, Sergio (1998). Subject-Verb Inversion in Spanish Wh-questions: Movement as Symmetry Breaker. In Bezooijen, R. v. & Kager, R. (eds.), *Linguistics in the Netherlands.*

Bailey, Todd & Hahn, Ulrich (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44: 568-591.

Bakovic, Eric (1998). Optimality and inversion in Spanish. In Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M. & Pesetsky, D. (eds.), *Is the Best Good Enough?* Cambridge: MIT Press. 35-58.

Beas, Omar (2007). *Agreement on the left edge: The syntax of left dislocation in Spanish*. University of Southern California.

Behnstedt, Peter (1973). *Viens-tu? Est-ce que tu viens? Tu viens ? Formen und Strukturen des direkten Fragesatzes im Französischen.* Tübingen: Narr.

Belletti, Adriana & Rizzi, Luigi (1988). Psych-verbs and θ-theory. *Natural Language and Linguistic Theory* 6(3): 291–352.

Blake, Frank R. (1930). A semantic analysis of case. In Hatfield, J. T., Leopold, W. & Zieglschmid, A. J. F. (eds.), *Curme volume of linguistic studies (Language monographs 7).* Baltimore: Linguistic Society of America. 34-49.

Boersma, Paul & Hayes, Bruce (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32(1): 45–86.

Bolinger, Dwight (1955). Meaningful word order in Spanish. *Boletín de Filología de la Universidad de Chile* 7: 45-56.

Boye, Kasper & Engberg-Pedersen, Elisabeth (eds.) (2010). *Language usage and language structure*

Berlin, New York: Mouton de Gruyter.

Bybee, Joan L. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language* 82(4): 711-733.

Bybee, Joan L. & Eddington, David (2006). A Usage-based Approach to Spanish Verbs of 'Becoming'. *Language* 82(2): 323-355.

Bybee, Joan L. & Hopper, Paul J. (eds.) (2001). *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins.

Chomsky, Noam (1964). *Current Issues in Linguistic Theory.* Den Haag: Mouton.

Chomsky, Noam (1965). *Aspects of the Theory of Syntax.* Cambridge: MIT Press.

Clark, Brady (2005). On Stochastic Grammar (Discussion Notes). *Language* 81(1): 207-217.

Coetzee, Andries W. (2004). *What it Means to be a Loser: Non-Optimal Candidates in Optimality Theory*. PhD dissertation, University of Massachusetts.

Cohen, Jacob (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement* 33: 107-112.

Contreras, Helen (1976). *A theory of word order with special reference to Spanish.* Amsterdam: North Holland.

Cornips, Leonie & Poletto, Cecilia (2005). On Standardising Syntactic Elicitation Techniques. PART 1. *Lingua* 115(7): 939-957.

Coveney, Aidan B. (1996). *Variability in Spoken French. A Sociolinguistic Study of Interrogation and Negation.* Exeter: Elm Bank.

den Dikken, Marcel (1997). The syntax of possession and the verb 'have'. *Lingua* 101(3-4): 129-150.

Dowty, David (1991). Thematic Proto-Roles and Argument Selection. *Language* 67(3): 547-619.

Elsig, Martin (2007). *Diachronic aspects of syntactic variation in the interrogative system of Québec French*. PhD dissertation, University of Hamburg.

Enghels, Renata (2009). The syntactic position of the perceived participant as indicator of the internal structure of the Spanish and French infinitival complement. *Linguistics* 47(3): 759-791.

Fillmore, Charles (1966). Toward a modern theory of case. *The Ohio State University project on linguistic analysis, Report.* Columbus: Ohio State University. 1-24.

Gahl, Susanne & Garnsey, Susan Marie (2004). Knowledge of Grammar, Knowledge of Usage: Syntactic Probabilities Affect Pronunciation Variation. *Language* 80(4): 748-775.

Gahl, Susanne & Garnsey, Susan Marie (2006). Knowledge of Grammar Includes Knowledge of Syntactic Probabilities (Discussion Notes). *Language* 82(2): 405-410.

Gervain, Judit & Zemplén, Gábor (2005). Focus raising: A paradigmatic example of the treatment of syntactic variation. In Cornips, L. & Corrigan, K. P. (eds.), *Syntax and Variation: Reconciling the Biological and the Social.* Amsterdam, Philadelphia: John Benjamins. 123-145.

Goodall, Grant (1991). Spec of IP and Spec of CP in Spanish Wh-Questions. In Mithun, M., Perissinotto, G. & Raposo, E. P. (eds.), *Linguistic Perspectives on the Romance Languages.* Amsterdam, Philadelphia: John Benjamins. 199-209.

Gruber, Jeffrey S. (1965). *Studies in lexical relations*. Indiana University Linguistics Club.

Gutiérrez-Bravo, Rodrigo (2006). *Structural Markedness and Syntactic Structure: A Study of Word Order and the Left Periphery in Mexican Spanish.* New York, London: Routledge.

Guy, Gregory R. (2005). Grammar and usage: A variationist response (Letters to Language). *Language* 81(3): 561-563.

Guy, Gregory R. (2007). Grammar and usage: The discussion continues (Letters to Language). *Language* 83(1): 2-4.

Henry, Alison (2005). Idiolectal variation and syntactic theory. In Cornips, L. & Corrigan, K. P. (eds.), *Syntax and Variation: Reconciling the Biological and the Social.* Amsterdam, Philadelphia: John Benjamins. 109-122.

Inclán-Nichol, Sara (1997). *Absence of Verb Inversion and Specificity in Peninsular Spanish Wh-questions*. UCLA.

Jackendoff, Ray (1972). *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.

Jackendoff, Ray (1987). The status of thematic relations in linguistic theory. *Linguistic Inquiry* 18: 369-411.

Keller, Frank (2000). *Gradience in Grammar. Experimental and Computational Aspects of Degrees of Grammaticality*. PhD dissertation, University of Edinburgh.

Keller, Frank & Lapata, Mirella (2003). Using the Web to Obtain Frequencies for Unseen Bigrams. *Computational Linguistics* 29(3): 459-484.

Kepser, Stephan & Reis, Marga (eds.) (2005). *Linguistic Evidence: Empirical, Theoretical, and Computational Perspectives.* Berlin: Mouton de Gruyter.

Kim, Ji-Hye & Yoon, James H. (2009). Long-distance bound local anaphors in Korean--An empirical study of the Korean anaphor caki-casin. *Lingua* 119(5): 733-755.

Kroch, Anthony (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199-244.

Labov, William (1969). Contraction, deletion, and inherent variability of the English copula. *Language* 45(4): 716-762.

Labov, William (1996). When Intuitions fail. In McNair, L., Singer, K., Dolbrin, L. & Aucon, M. (eds.), *Papers from the Parasession on Theory and Data in Linguistics. Chicago Linguistic Society 32.* Chicago: Chicago Linguistic Society. 77-106.

Lakoff, George (1974). Interview. In Parret, H. (ed.), *Discussing language: Dialogues with Wallace L. Chafe, Noam Chomsky, Algirdas J. Greimas [and others]*. The Hague: Mouton. 151-178.

Lakoff, George & Johnson, Mark (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* New York: Basic Books.

Langacker, Ronald W. (1987). *Foundations of cognitive grammar, vol. 1: Theoretical prerequisites.* Palo Alto: Stanford University Press.

Larson, Richard K. (1988). On the double object construction. *Linguistic Inquiry* 19: 335-391.

Laury, Ritva & Ono, Tsuyoshi (2005). Data is Data and Model is Model: You Don't Discard the Data that Doesn't Fit Your Model! (Discussion Notes). *Language* 81(1): 218-225.

Lotfi, Ahmad R. (2003). Persian Wh-Riddles. In Boeckx, C. & Grohmann, K. K. (eds.), *Multiple Wh-Fronting.* Amsterdam: John Benjamins. 161-186.

Masullo, Pascual J. (1993). Two types of quirky subjects: Spanish versus Icelandic. *Proceedings of NELS 23.* Amherst: University of Massachusetts. 303-317.

Meisel, Jürgen M. (in press). Bilingual language acquisition and theories of diachronic change: Bilingualism as cause and effect of grammatical change. *Bilingualism Language and Cognition*.

Meyer, Charles F. & Tao, Hongyin (2005). Response to Newmeyer's 'Grammar is Grammar and Usage is Usage' (Discussion Notes). *Language* 81(1): 226-228.

Meyer, Paula (1972). Some observations on constituent order in Spanish. In Casagrande, J. & Saciuk, B. (eds.), *Generative Studies in Romance Languages.* Rowley: Newbury House Publishers. 184-195.

Müller, Gereon (1999). Optimality, markedness, and word order in German. *Linguistics* 37(5): 777–818.

Newmeyer, Frederick J. (2003). Grammar is Grammar and Usage is Usage. *Language* 79(4): 682-707.

Newmeyer, Frederick J. (2005). A Reply to the Critiques of 'Grammar is Grammar and Usage is Usage' (Discussion Notes). *Language* 81(1): 229-236.

Newmeyer, Frederick J. (2006a). Grammar and usage: A response to Gregory R. Guy (Letters to Language). *Language* 82(4): 705-708.

Newmeyer, Frederick J. (2006b). On Gahl and Garnsey on Grammar and Usage (Discussion Notes). *Language* 82(2): 399-404.

Olarrea, Antxon (1996). *Pre and Post-verbal Subject Position in Spanish: A Minimalist Account*. University of Washington.

Ordóñez, Francisco (1997). The inversion construction in interrogatives in Spanish and Catalan. In Lema, J. & Teviño, E. (eds.), *Theoretical Analyses on Romance Languages.* Amsterdam, Philadephia: John Benjamins. 328-350.

Ordóñez, Francisco (2000). *The Clausal Structure of Spanish: A Comparative Study*. New York: Garland.

Payne, Doris L. (1990). *The pragmatics of word order: Typological dimensions of verbinitial languages.* Berlin: Mouton de Gruyter.

Payrató, Lluís (1985). *La interferència lingüística. Comentaris i exemples català - castellà.* Barcelona: Montserrat (Curial edicions catalanes).

Pesetsky, David (1995). *Zero syntax.* Cambridge: MIT Press.

Pierrehumbert, Janet (1994). Syllable structure and word structure: A study of triconsonantal clusters in English. In Keating, P. (ed.), *Phonological structure and phonetic form (Papers in laboratory phonology 3).* Cambridge: Cambridge University Press. 168–190.

Postal, Paul M. (1970). On the surface verb 'remind'. *Linguistic Inquiry* 1: 37-120.

Ross, John Robert (1967). *Constraints on variables in syntax*. Massachusetts Institute of Technology.

Sabater, Ernest (1991). *Ni Heavy, ni Light!: Català modern.* Barcelona: Empúries.

Schütze, Carson T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology.* Chicago: University of Chicago Press.

Sinner, Carsten & Wesch, Andreas (eds.) (2008). *El castellano en las tierras de habla catalana.* Madrid, Frankfurt am Main: Iberoamericana / Vervuert.

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

Speas, Margaret (1990). *Phrase Structure in Natural Language.* Dordrecht: Kluwer.

Suñer, Margarita (1994). V-movement and the licensing of argumental *wh*-phrases in Spanish. *Natural Language and Linguistic Theory* 12(2): 335-372.

Suñer, Margarita & Lizardi, Carmen (1995). Dialectal Variation in an Argumental/Non-Argumental Asymmetry in Spanish. In Amastae, J., Goodall, G., Montalbetti, M. & Phinney, M. (eds.), *Contemporary Research in Romance Languages: Papers from the 22nd Linguistic Symposium on Romance Languages El Paso-Ciudad Juarez 1992.* Amsterdam: John Benjamins.

Toribio, Almeida Jacqueline (2000). Setting parametric limits on dialectal variation in Spanish. *Lingua* 110(5): 315-341.

Torrego, Esther (1984). On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15(1): 103-129.

Trutkowski, Ewa, Zugck, Marco, Blaszczak, Joanna, Fanselow, Gisbert, Fischer, Susann & Vogel, Ralf (2003). Superiorität in europäischen Sprachen. Zwischenbericht zu einer Datenerhebung. In Fischer, S., Vijver, R. v. d. & Vogel, R. (eds.), *Linguistics in Potsdam 21*. Potsdam: Institut für Linguistik, Universität Potsdam. 123-137.

UNESCO (2006). International Standard Classification of Education 1997 (ISCED). Geneva: Unesco Institute for Statistics.

Uribe-Etxebarria, Maria (1995). On the Nature of SPEC/IP and Its Relevance for Scope Asymmetries in Spanish and English. In Amastae, J., Goodall, G., Montalbetti, M. & Phinney, M. (eds.), *Contemporary Research in Romance Languages: Papers from the 22nd Linguistic Symposium on Romance Languages El Paso-Ciudad Juarez 1992.* Amsterdam: John Benjamins. 355-366.

Wilson, John & Henry, Alison (1998). Parameter setting within a socially realistic linguistics. *Language in Society* 27: 1-21.

Zubizarreta, Maria Luisa (2001). The constraint on preverbal subjects in Romance interrogatives: a Minimality effect. In Hulk, A. & Pollock, J.-Y. (eds.), *On Romance Inversion.* Oxford: Oxford University Press.

[1] Applying a standard interview technique (more precisely a so-called sociolinguistic interview), Coveney (1996: 116) obtained (after excluding rhetorical or echo-questions) in average as few as 4.25 yes/no-questions and one *wh*-question per speaker, although the interviews had a mean duration of 36 minutes. He obtained this sparse outcome despite his attempts to create interview situations that should motivate the interviewee to ask questions back. The outcome in Behnstedt (1973: 217&222) is merely better.

[2] In short, the annotation carries (i) on properties of the entire structure (e.g. sentence modality, coordination, ellipsis, cross-sentential dependency, question bias, specifics such as cleft or pseudo-cleft, etc.), (ii) on the grammatical function of embedded clauses (subject, object, PP-complement, adverbial, relative, adverbial, peripheral-adverbial, sentential rela-

tive, predicative, DP complement), (iii) on each major type of XP (subject, verb, direct object, indirect object, prepositional complement / locative adverbial complement, NP-complement, VP-adjunct, IP-adjunct, CP-adjunct, complementizer, attribute, vocative), and (iv) on the base and landing position for major types of displacements (*wh*-movement, topicalization, dislocation, focus-movement). These XPs are further characterized. For example, the annotation of verbs carries on finiteness, ±modal, verb class (copular, unaccusative, impersonal, unergative, raising, etc.), clitical affixation, person, number, ±negation, and voice.

[3] The reference sentence is given in (i). The word order in combination with the elided matrix verb is clearly marked.

(i) ?Demasiado la angustia el hecho de examinarse.
too much the anxiety the fact of taking an exam
'There is too much anxiety for taking an exam.'

[4] I leave the question open as to whether the subject in transitive constructions and the oblique experiencer in psych-verb constructions occupy the same structural position, a position defended by Masullo (1993) and Gutiérrez-Bravo (2006: 51) (see Belletti & Rizzi 1988 for more details on psych-verbs).

[5] All three tests are highly significant. Main effect A '*wh*-role': $F(3,156) = 108.5$, $p < 0.000$; main effect B 'subject position': $F(1,52) = 289.9$, $p < 0.000$; interaction A x B: $F(3,156) = 144.2$, $p < 0.000$.

[6] The results of the simple main effect tests read as follows:
(1a) vs. (1b) (simple main effect $B|a_1$): $F(1,52) = 374.8$, partial $\eta^2 = 0.878$, $p < 0.000$.
(2a) vs. (2b) (simple main effect $B|a_2$): $F(1,52) = 383.3$, partial $\eta^2 = 0.881$, $p < 0.000$.
(3a) vs. (3b) (simple main effect $B|a_3$): $F(1,52) = 46.9$, partial $\eta^2 = 0.474$, $p < 0.000$.
(4a) vs. (4b) (simple main effect $B|a_4$): $F(1,52) = 0.7$, partial $\eta^2 = 0.014$, $p < 0.398$.

[7] THEME and PATIENT only differ minimally in Dowty's (1991) system. The distinctive element is whether the argument is causally affected by another participant or not (p. 577).

[8] Are excluded: lexical doublets which are ambiguous with regard to P-Agent and P-Patient (e.g. *comprar* 'buy'*, vender* 'sell'), verbs that have a PATIENT or GOAL subject and an AGENT or CAUSE object (e.g. *recibir* 'receive'*, someterse* 'undergo'), verbs for which the two arguments cannot be clearly distinguished (e.g. predicative constructions with *ser/estar* 'be'), verbs that involve volition on both subject and object (such as *x difiere de y* 'x is different from y, *x discute algo con y* 'x discusses the matter with y') and verbs which do not have neither P-Agent nor P-Patient entailments (e.g. complements have been traditionally described as SOURCE or GOAL, often realized as prepositional complements). Predicative constructions with *ser/estar* 'be' (69 tokens) constitute a large part among excluded sentences. The same holds for the possessive verb *x tiene y* 'x has y' (35 tokens), see den Dikken (1997), and the verb *x se llama y* 'x is called y' (12 tokens).

[9] Note that I do not claim that a construction that does not occur in a corpus of spontaneous speech is *never* used. One should refrain from such conclusions which are non-falsifiable claims.

[10] Bybee & Eddington (2006) also point out that a low-frequency item can be judged as acceptable – however, if and only if it is semantically similar to other high-frequency items. Their claim does not build on syntactic phenomena but on possible verb+adjective combinations for which it is less problematic to argue in terms of semantic similarity.

[11] Two hypotheses require further inquiry: First, the availability of a latent construction is a "secondary effect" of a parametric change (implying that if a latent construction occurs, it represents a minor aspect of the entire process of change). Second, it represents a non-parametric change, presumably at the interface between grammar and usage, which can be, however, the precursor of more important steps.