Aria Adli, University of Cologne

What you like is not what you do: Acceptability and frequency in syntactic variation

Abstract: An interesting property of French WH-questions is the presence of several syntactic variants. This study discusses a quantitative analysis of WH-adjunct and WH-object questions, drawing on gradient acceptability judgments as well as frequency of occurrence in spontaneous speech. Both types of evidence are collected with the same set of speakers in order to allow direct comparisons. The results show interesting mismatches between acceptability and frequency. First, the preferred variants differ: Speakers make use of the WH-in-situ variant (the most frequent form), the variant with the particle *est-ce que* and the WHSV form, but not the WHVS, SVWHO and WH-cleft variants. However, the judgment data show that all of these variants are acceptable (with formal variants being more acceptable than colloquial ones). Second, the contrasts between WH-adjunct and WH-object questions are clear-cut in spontaneous speech, while they are mild in acceptability. In particular, we see that the particle *est-ce que* is only used with the WH-object *que*, although WH-adjunct with *est-ce que* is also considered equally acceptable. Thirdly, fine-grained corpus analyses distinguishing between different categories of WH-adjuncts and WH-objects highlight a differential behavior among WH-words: WH-REASON-adjuncts are essentially precluded with the WH-in-situ order. Furthermore, the particle *est-ce que* is limited to the inanimate WH-object *que* ('what'), i. e. it is not used with the animate WH-object *qui* ('whom'). The mismatches between acceptability and frequency prove to be less pronounced in fine-grained analyses. With regard to the larger methodological picture, we see the interesting potential that lies in combined studies drawing on acceptability as well as frequency. In particular, acceptability judgments – which traditionally have a bad standing in sociolinguistics – can help to reveal normative effects and play a crucial role in the circumscription of the full envelope of variation.

1 Relation between acceptability and frequency

The two most important sources of evidence in grammar research are acceptability judgments and corpus data. They are closely associated with specific theoreti-

cal frameworks and traditions. Acceptability judgments are seen as the royal path in generative grammar. They are considered a direct reflection of the real object of interest: ILanguage or competence. Corpus data of actual language use are in the center of interest in sociolinguistics and usage-based approaches. Apart from theoretical differences between the frameworks, various approaches can also be incompatible at the methodological level, illustrated by the following quotations. On the one hand, Chomsky (1965: 191) has already claimed half a century ago: "To maintain, on grounds of methodological purity, that introspective judgments of the informant (often, the linguist himself) should be disregarded is, for the present, to condemn the study of language to utter sterility". This position is still up-to-date for today's generative syntacticians. On the other hand, Labov (1996: 83) states that "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". The majority of sociolinguists share his critical stance towards introspection.

Yet, these antagonistic positions have slightly softened – a development sustained by generative studies on diachronic syntax and language acquisition. Many linguists would agree that the choice of the type of evidence depends more on the research question than on some inherent quality criterion of the data type itself. Just as corpus data can be extracted in a more or less meaningful way, introspection can be collected in a more or less convincing manner. Whatever method we use, focus should be given to careful methodology and data handling. Nevertheless, one important issue remains open: What is the relation between introspection and language use and how can we model it? It is very important to find out whether both empirical sources lead to the same answer on one theoretical question. In many circumstances it is certainly advantageous to corroborate, if possible, theoretical hypotheses in linguistics by means of different types of empirical data. However, the road to such a complementary approach needs to be better paved. More specifically, we need to have more precise knowledge on the relation between acceptability and frequency to better interpret the results of a study working with both types of data.

In order to do so, this study presents empirical findings on syntactic variation in French WH-questions, using frequency as well as gradient acceptability data. The unique aspect of the present study is the fact that both types of data were collected from the same speakers.

The relation between acceptability and frequency is an under-studied issue. We have only few empirical studies thus far.

Backus and Mos (2011) compare gradient acceptability judgments and corpus data, however not with regard to word order but to two ways of expressing potentiality, namely by a derivational morpheme equivalent to English *-able* or

by a copula construction. They observe a good match between introspection and production.

Stefanowitsch (2008) discusses from a cognitive usage-based perspective the correlation between negative evidence in corpora and acceptability judgments with respect to typical uses of verbs with the ditransitive/dative contrast in English. He essentially suggests that the quantitative difference between the expected and observed frequency of the co-occurrence of linguistic constructions, which the speakers presumably are capable of calculating "subconsciously", correlate with degrees of unacceptability.

Bybee and Eddington (2006) is another study on the lexicon in the usage-based framework. The authors analyze the correlation between high- vs. low-frequency verb+adjective expressions in Spanish. They show that high-frequency items are also judged as more acceptable, suggesting that "grammaticality or acceptability judgments are heavily based on familiarity, that is, the speaker's experience with language" (Bybee and Eddington 2006: 349).

Featherston (2005) analyzes several variants concerning three phenomena discussed in the theoretical literature on German syntax (discourse-linking, parenthesis vs. extraction, object coreference). He shows that only those constructions occur in corpora that have a relatively high degree of acceptability. In order to explain mismatches between acceptability and frequency, he suggests that human grammar contains both a cumulative and a probabilistic component.

Kempen and Harbusch (2008) compare gradient experimental judgments and frequency counts from corpora with respect to word order variants in German finite subordinate clauses. They observe that only those constructions which scored high in the judgment test were also detectable in the corpus. However, the constructions showed a large amount of variation with respect to their frequency of occurrence, although they all received high acceptability values. Based on their results, Kempen and Harbusch (2008) propose a two-factor theory in order to explain this mismatch: First, a construction must exceed a frequency minimum to be learnable, i. e. to be included into the child's grammar.[1] Second, repair mechanisms can give a positive bias to ungrammatical constructions, leading to marginal results, which "should not be mistaken for an authentic grammaticality rating" (Kempen and Harbusch 2008: 190). However, we do not think that their two-factor theory is the whole story: If scarcity of a construction was an indicator that it is not part of the grammar, we would end up with a fairly restrictive grammar. Such a grammar would not include many non- or hardly-

1 The criterion of learnability replaces a grammaticality threshold previously proposed in Kempen and Harbusch (2005) but now dismissed by the authors.

occurring items (e. g. multiple questions) which native speakers nevertheless can give surprisingly stable and even nuanced, gradient judgments about. What is more, many constructions that are qualified by most speakers as natural and fully acceptable can be fairly scarce in usage (e. g. wh-indirect object questions, see below; see also the discussion in Sampson 2007).

Bader and Häussler (2010) also compare gradient judgments of corpus data with respect to the order of subject and object and to verb-cluster linearization in German. They observe a similar mismatch as Kempen and Harbusch (2008) and Adli (2011c), namely that constructions with a high level of acceptability can greatly vary with respect to frequency (this "ceiling effect" has previously been underestimated by Featherston 2005). At the same time, extreme scarcity of a construction does not allow us to predict its level of acceptability.

We propose in Adli (2011c) the concept of a *latent construction* to refer to those fully acceptable but extremely scarce or non-occurring constructions. Furthermore, we propose a possible scenario for certain types of diachronic change, involving the following steps: (i) A construction X is not available in grammar, (ii) a construction X becomes available in grammar but is not used, (iii) X is used as part of a set of optional syntactic variants, (iv) cases of unstable optionality are dissolved, leaving only X (Adli 2011c: 398).

We also mention the early studies by Greenbaum (1976, 1977). He showed a correlation between acceptability judgments and judgments on the assumed frequency of the same constructions. He showed that native speakers believe that the more acceptable a construction is, the more often it occurs. Interestingly, this is a misbelief of the speakers, as we know today. One could also take Greenbaum's (1976, 1977) result as an indication that speakers are probably mostly unaware of the large degree of variation in frequency among acceptable constructions.

Given the challenge to explain why certain constructions are acceptable but hardly occur, there is a close link between the issue of the relation between acceptability and frequency and the issue of data scarcity of specific constructions in corpora. In this context, we also mention Pullum (2007), who discusses rarity in corpora. Similarly, Foster (2007) and Ayres-Bennett (1994) discuss negative evidence in corpora (see also Stefanowitsch 2008). The issue of rare typological features from a generative point of view has been discussed by Newmeyer (2010) and Rijkhoff (2010). The underlying problems raised above are not new (though they have rarely been discussed in methodological terms): The question to ask is whether rare constructions are also marked constructions. This has been explicitly stated by Baayen et al. (1997: 14), and goes back to Greenberg (1966) and Trubetzkoy (1939). On this matter, Haspelmath (2006: 33) pleads in favor of using directly the notions "rare" or "frequent" instead of the fairly polysemous notions of marked or unmarked.

When comparing two measures, a standard practice in empirical methodology is to collect these measures with the same subjects. However, apart from Adli (2011c) (and disregarding Greenbaum 1976, 1977, who does not study acceptability), all studies mentioned above compare a sample of speakers who gave introspective judgments with known corpora, i. e. one measure taken from sample 1 (acceptability judgments) is compared to another measure taken from sample 2 (authors or speakers recorded in corpora). This approach is understandable from a practical point of view due to a lack of appropriate data. Having said this, the innovative aspect of the present study is the fact that spontaneous speech, acceptability data and social information were collected from the same set of speakers, compiled in the database *sgs* described below. We can therefore rule out that any difference or similarity observed between acceptability and frequency might be due to the (social, individual, dialectal, text- or discourse-specific...) differences between the respective samples. We believe that this approach offers more reliable results on the relation between both data types. The present study thus extends the research program which we started with Adli (2011c) on Spanish. We now turn to another language (and another sample), namely French and analyze the phenomenon of French WH-questions. Another aspect of the present study worth mentioning is that it analyzes an envelope of variation (standard practice in quantitative sociolinguistics but not in syntax), both with acceptability and frequency data.

2 Acceptability and frequency in linguistic variation

Before presenting the constructions that are going to be compared in terms of acceptability and frequency, we give a brief overview of syntactic variation in French WH-syntax. Since we aim at taking into account the entire set of variants (called "circumscription of the envelope of variation" in variationist terminology), this is an important preliminary step. This step builds on the principle of accountability (Labov 1982: 30) which states that the variants belonging to the same variable must be specified by the total number of occurrences and the potential occurrences or non-occurrences in the variable environment.

The following list illustrates the large repertory of syntactic variants of WH-questions in French with (a) examples being WH-adjunct questions and (b) examples being WH-object questions.

(1a) *tu fais le dessin quand ?* [WH-in-situ]
you make the drawing when
'When do you make the drawing?'

(1b) *tu vois qui devant la fenêtre ?* [WH-in-situ]
you see who in front of the window
'Who do you see in front of the window?'

(2a) *quand est-ce que tu fais le dessin ?* [WH-ESQ]
when EST-CE QUE you make the drawing

(2b) *qui est-ce que tu vois devant la fenêtre ?* [WH-ESQ]
who EST-CE QUE you see in front of the window

(3a) *quand tu fais le dessin ?* [WHSV]
when you make the drawing

(3b) *qui tu vois devant la fenêtre ?* [WHSV]
who you see in front of the window

(4a) *quand fais -tu le dessin ?* [WHVS$_{clit}$ (=clitic inv.)]
when make -you the drawing

(4b) *qui vois -tu devant la fenêtre ?* [WHVS$_{clit}$ (=clitic inv.)]
who see -you in front of the window

(5) *tu fais quand le dessin ?* [SVWHO]
you make when the drawing

(6a) *quand les enfants font -ils le dessin ?* [complex inversion]
when the children$_i$ make -they$_i$ the drawing
'When do the children make the drawing?'

(6b) *qui les enfants voient -ils devant la fenêtre ?* [complex inversion]
who the children$_i$ see -they$_i$ in front of the window
'Who do the children see in front of the window?'

(7a) *c'est quand que tu fais le dessin ?* [WH-in-situ cleft]
it is when that you make the drawing

(7b) *c'est qui que tu vois devant la fenêtre ?* [WH-in-situ cleft]
it is who that you see in front of the window

(8a) *quand c'est que tu fais le dessin ?* [WH-cleft]
when it is that you make the drawing

(8b) *qui c'est que tu vois devant la fenêtre ?* [WH-cleft]
who it is that you see in front of the window

(9a) *quand est-ce que c'est que tu fais le dessin ?* [WH-ESQ cleft]
when EST-CE QUE it is that you make the drawing

(9b) *qui est-ce que c'est que tu vois devant la fenêtre ?* [WH-ESQ cleft]
who EST-CE QUE it is that you see in front of the window

The list distinguishes on a first level between non-clefted WH-questions, (1a) to (6b), and clefted WH-questions, (8a) to (9b) (Lambrecht 2001; Dufter 2008). We can find both WH-in-situ constructions as in (1a) and (1b) (Adli 2006; Hamlaoui 2011; Déprez et al. 2013), which are not restricted to echo questions (Reis 1991; Escandell-Vidal 2002; Sobin 2010), and WH-fronted constructions. In the fronted variants, the initial WH-word can be followed by the interrogative particle *est-ce que* as in (2a) and (2b), or by (non-inverted) subject and verb as in (3a) and (3b), or by inverted subject and verb as in (4a) and (4b).[2] Furthermore, French allows clefted WH-questions which also exhibit part of the already mentioned variation: WH-clefts can appear with WH-in-situ as in (7a) and (7b), or with a fronted WH--element as in (8a) and (8b), or with the particle *est-ce que* as in (9a) and (9b).

Two of the non-clefted questions are rather restricted: First, (5) is a marked variant of the WH-in-situ question in which the direct object is postposed. This option exists with WH-adjunct questions containing a transitive or bitransitive verb. It resembles the WH-in-situ question with right-dislocated object (*tu le fais quand le dessin*), apart from the fact that the object clitic (*le*) is missing. Second, (6a) and (6b) are so-called complex inversions in which the full subject is doubled by a coreferential inverted clitic.

3 Methodology

3.1 Overview of the *sgs* database

Sgs is a multilingual database that we have been constructing since 2004 (see Adli 2011b). It contains data on four languages – French, Spanish, Catalan and Persian – that have been collected using the same methodological protocol. Every person was first recorded, then participated in a gradient acceptability judgment test, and finally filled out an extensive social questionnaire. Spontaneous speech data were obtained by recording interviewer and interviewee while they played a specifically designed game. Essentially, the interviewee had to solve a fictive murder case by speaking freely with the (native and well-trained) interviewer. Most interviewees chose a non-formal, rather colloquial register, encouraged by a previous warm-up or "ice-breaker" phase. We favored this game task over

2 The construction with an inverted weak subject pronoun as in (4a) and (4b) is often called subject-clitic inversion (Auger 1994; Elsig 2009). A construction with an inverted non-pronominal subject (e. g. *Quand fait Jean le dessin?*) is often referred to as stylistic inversion (Kayne and Pollock 1978; Drijkoningen and Kampers-Manhe 2008).

the standard sociolinguistic interview because it elicits a substantial number of declarative and interrogative sentences, while sociolinguistic interviews are sentence-type-restricted in the sense that interviewees hardly ever produce questions. It would have been much more difficult (and costly) to realize the same study on WH-questions with data obtained with the classic sociolinguistic interview.

The recordings were transcribed, time-stamped, and most importantly, syntactically annotated. We annotated type and function of all major constituents, including tree-relations of subordination and coordination in complex sentences.

The French part of *sgs* contains 27 hours of recordings or 44,231 main lines, 51 % of which are produced by 101 interviewees and 49 % by the interviewers. Interviews were carried out in the summer of 2005 in Paris with French native speakers between the ages of 19 and 49 (mean age: 29). The sample is essentially gender-balanced (56 % women and 44 % men). Only data of these 101 interviewees (and not of the interviewers) are taken into consideration in the following analyses. According to the transcription and annotation guidelines, one main line corresponds to one (full or elliptical) sentence, or to single interactional markers such as a single *oui* ('yes'), *non* ('no'), or to pragmatic phenomena such as false starts or interrupted, unfinished sentences. The corpus contains all in all 10,943 full, i. e. non-elliptical, sentences produced by the interviewees. Among them, we find 1,721 root WH-questions. To my knowledge, this is the largest set of WH-questions extracted from a single corpus. Yet, one should take also note of Elsig (2009: 147), who extracted 1,055 tokens from the Ottawa-Hull corpus of modern Canadian French (Poplack 1989). Coveney and Dekhissi (2013) extracted roughly 1,070 true-information (i. e. non-rhetorical) WH-questions from a corpus based on selected contemporary French films/screenplays (Dekhissi in prep.). Druetta (2008: 37) worked with 395 WH-questions from the G. A. R. S. corpus, recorded essentially in southeastern France. Aside from that, other studies on WH-questions in spoken language work with rather small numbers (Behnstedt 1973; Coveney 1996).

3.2 Towards an envelope of variation

3.2.1 Descriptive overview

We will start by presenting descriptive details on the 1,721 extracted full, root WH-questions from *sgs* in Table 1. Please recall that this set neither includes embedded WH-questions such as *tu sais quand il est parti à Paris?* 'Do you know when he

left for Paris?' nor elliptical WH-questions such as *quand?* or *quand ça?* 'When?'. Furthermore, it only includes true information questions.[3]

Table 1: Number of tokens of different word order variants of WH-questions

VARIANT	N	PERCENT	
WH-in-situ	944	56.2 %	(see (1a)/(1b))
WH-ESQ	281	16.7 %	(see (2a)/(2b))
WHSV	256	15.2 %	(see (3a)/(3b))
WHVS	167	9.9 %	(see (4a)/(4b))
SVWHO	6	0.4 %	(see (5))
WH-in-situ cleft	17	1.0 %	(see (7a)/(7b))
WH-cleft	6	0.4 %	(see (8a)/(8b))
WH-ESQ cleft	0	0.0 %	(see (9a)/(9b))
complex inversion	2	0.1 %	(see (6a)/(6b))
multiple WH-question	1	0.1 %	

Please note that the cells in the table are not fully comparable because they do not represent an envelope of variation. Most notably, WH-subject questions (e. g. *qui fait le dessin* 'who does the drawing'), which can be assigned due to the surface order to the WH-fronted as well as the WH-in-situ category, were by definition not assigned to the WH-in-situ category. Furthermore, stylistic inversion (e. g. *quand dort Jean* 'when does Jean sleep') and subject-clitic inversion (e. g. *quand dort-il* 'when does he sleep') are aggregated into one category because the table does not differentiate between pronominal and non-pronominal subjects. Finally, it includes one multiple WH-question.

Yet, Table 1 provides insights into the frequency of different WH-variants in spontaneous speech: First, we observe that only four variants are really productive: inversion questions (see (4a)/(4b)), the form with initial WH-element followed by subject and verb (see (3a)/(3b)), the form with the *est-ce que* particle (see (2a)/(2b)), and – by far the most frequent variant – the WH-in-situ form (see (1a)/(1b)). Second, we see that complex inversion (see (6a)/(6b)) is basically absent – which is less surprising due to its high level of formality. Thirdly, we observe that WH-cleft constructions (see (7a) to (9b)) are extremely scarce.

3 Yet, there is no echo question and only nine rhetoric questions in *sgs* – rhetoric in the sense of utterances that are pragmatically equivalent to declaratives with the speaker knowing the answer (see Prieto and Rigau 2007).

Our goal is to obtain a comparable set of constructions for the following analyses. To this end, we further restrict the overall set of WH-questions in Table 1 by limiting ourselves to sentences with a pronominal subject. We know that sentences with pronominal and lexical subjects are analyzed quite differently in French: Weak pronouns in spontaneous French are clitics that can be analyzed as mere verbal affixes (under this assumption colloquial French might in fact be attributed properties of a null subject language, see e. g. Culbertson 2010). Furthermore, limiting ourselves to pronominal subjects removes cases of stylistic inversion from the WHVS order and excludes cases of complex inversion like in (6a)/(6b) (as a result, postverbal subjects will only occur as subject-clitic inversions such as (4a)/(4b)). The result of this restricted set is shown in Figure 1, in which the non-occurring WH-ESQ cleft constructions such as (9a)/(9b) and multiple WH-questions are no longer represented. Figure 1 shows the number of tokens for each word order variant in the French part of *sgs*, also further distinguishing between WH-adjunct and WH-object questions. One should bear in mind that the SVWHO order with WH-objects is not a zero frequency but an empty cell, because this order is only defined for WH-adjunct questions (with transitive or ditransitive verbs).

Figure 1 reveals a very clear distributional difference between WH-adjuncts and WH-objects, which will be discussed in more detail further below.

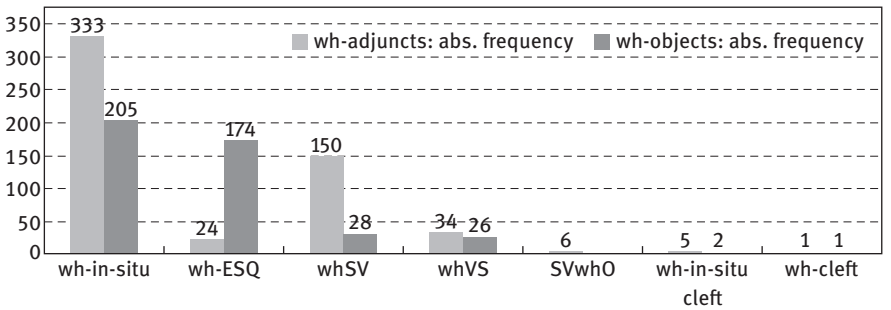


Figure 1: Number of tokens of different word order variants of WH-adjunct and WH-object questions with pronominal subject

3.2.2 Calculating relative frequencies

In the next step, we calculate the proportion (percentage between 0 and 100 or relative frequency between 0 and 1) of each of the seven variants of WH-object questions and of each of the six variants of WH-adjunct questions shown in Figure 1. Essentially, there are two ways to calculate this measure: We can first add up all

occurrences of each variant in the entire corpus and then calculate their proportions. This means that we treat the entire corpus as a single unity, disregarding the level of individual speakers. This measure is called "single-text-value" in Adli (2011a: section 6.2). Or we can first calculate the proportions for each speaker and then calculate the mean value of the proportions for the sample (called "speaker-sample-value" in Adli 2011a: section 6.2). The differences are shown in (10a) and (10b). For example, if we calculate the relative frequency of our target-variant WHSV among the $x = 7$ variants of WH-adjunct questions of Figure 1 as a single-text-value, we would first add up all occurrences of our target variant across all $n = 101$ speakers and then divide this number by the sum of all $x = 7$ variants across all $n = 101$ speakers. However, if we want to work with the speaker-sample-value, we would first add up the relative frequencies of all $n = 101$ speakers for our target-variant WHSV and divide this number by $n = 101$.

(10a) relative frequency (single-text-value):

$$\frac{\sum_{i=1}^{n} N_{\text{TARGET-VARIANT}i}}{\sum_{i=1}^{n}\left(\sum_{j=1}^{x}\left(N_{\text{VARIANT}ji}\right)\right)}$$

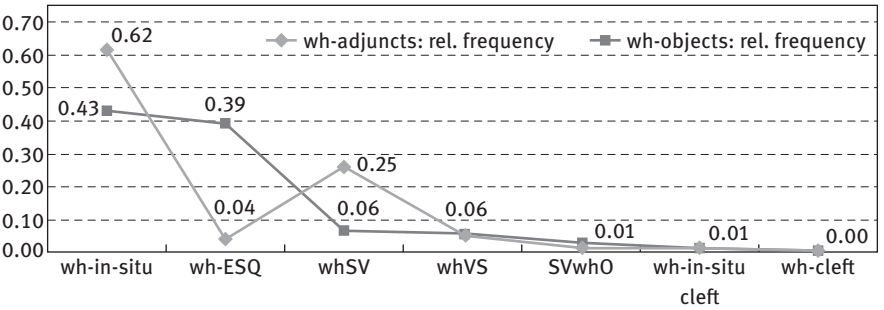
(10b) relative frequency (speaker-sample-value):

$$\frac{\sum_{i=1}^{n}\left(\frac{N_{\text{TARGET-VARIANT}i}}{\sum_{j=1}^{x}\left(N_{\text{VARIANT}ji}\right)}\right)}{n}$$

In order to work with speaker-sample-values, a corpus must adhere to stricter conditions: Data must be collected from each speaker under comparable and controlled conditions – which is the case in *sgs*. We will work in the following analyses with speaker-sample-values according to (10b), because they are more robust than single-text-values, especially when dealing with scarce data: Whenever we have a distribution where many speakers have produced very few tokens of a target variant and a few speakers have produced a comparatively high number of tokens, single-text-values can overestimate the results, sometimes even drastically distort them.

Figure 2 quantifies the same constructions as in Figure 1, but it shows relative frequencies as speaker-sample-values. In a next step, these relative frequencies can be compared to the gradient acceptability judgments, since (i) the judgments will also be mapped on a linear scale from 0 to 1, and (ii) they are also mean scores of individual values. Since we will later add judgment scores to the diagram, we use lines and not bars in Figure 2 (multiple lines are often more readable than multiple bars, especially for spotting interactions).

Figure 2: Relative frequency (speaker-sample-value) of different word order variants of WH-adjunct and WH-object questions with pronominal subject



3.3 Fine-grained rating of acceptability

3.3.1 Gradient acceptability judgments on a visual analogue scale

Introspective data were collected experimentally using a gradient acceptability judgment test, developed in Adli (2004: chapter 3) and already applied in various studies (e. g. Adli 2010a). This instrument measures acceptability in a gradient manner. Unlike the magnitude estimation technique (Bard et al. 1996), it is based on a graphic rating or a visual analogue scale (Freyd 1923; Funke 2010). The scale has two endpoints (totally unacceptable and fully acceptable). Subjects rate the perceived degree of acceptability by drawing a line with a pen: The longer the line the more acceptable the sentence (see Adli 2011b for a computer-based version of the test). Figure 3 shows an example page taken from the test material. Two sheets that can be turned independently from each other are placed in a letter size (A4) binder. A (suboptimal) reference sentence that subjects have judged at the end of the training phase is printed on the upper sheet. It remains visible until the end of the experiment, providing a self-chosen intermediate scale anchor. Thus subjects can calibrate their ratings by means of both endpoints and this anchor. The experimental sentences are printed on the white sheets on the lower part of the binder. Once the subjects have rated the sentences on a white sheet, they turn it to continue with the next test sentences (see the Appendix for a list of all experimental items as well as the sentences from the instruction and training phase).

The experiment starts with a thorough instruction phase (see Adli 2004: chapter 3 for details), during which subjects learn the concept of gradience – as opposed to binary good/bad, tripartite good/intermediate/bad judgments, etc. Furthermore, they are instructed to judge syntactic well-formedness and not irrel-

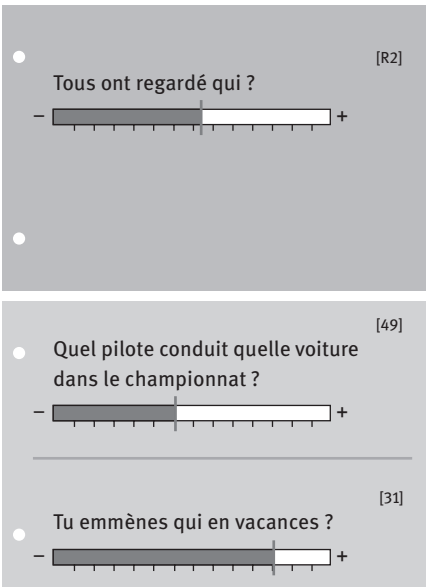


Figure 3: Gradient acceptability judgment test

evant extra-grammatical aspects (such as pragmatic plausibility). They are asked to leave aside normative considerations and to refer to spoken, colloquial everyday language.

During the instruction phase, subjects become accustomed to judging acceptable, marginal and unacceptable constructions. Finally, their knowledge of the instrument is verified in a training phase before starting with the actual experimental items. Each of the constructions shown in Figure 4, namely (1a), (1b), (2a), (2b), (3a), (3b), (4a), (4b), (5), (8a), and (8b), were presented in three lexical variants (see Appendix) in order to obtain a more valid test score. Among the three cleft forms (WH-in-situ cleft, WH-cleft, WH-ESQ cleft), the WH-cleft variant (8a)/(8b) was included in the test. To give an example for calculating the dependent variables, a subject's experimental score for the WH-object question (1b) is the arithmetic mean of her/his judgments of the respective three variants of (1b). Except for the WH-element and the pronominal subject, lexical repetitions between the test sentences were avoided. The order of the sentences was randomized. On average, the instruction and training lasted 15 minutes and the actual experimental phase 20 minutes. Please note that the acceptability test included other constructions that are not at the center of the present paper but which can be considered filler sentences with respect to the present set of experimental items.

3.3.2 Descriptive overview of the ratings

Figure 4 shows the gradient acceptability values (two lines in the upper part) and the relative frequencies of Figure 2 (two lines in the lower part).

The value points of WH-adjunct questions are linked by a broken line, and the value points of WH-object questions by a continuous line. Since both scales run from 0 to 1 (i. e. the domain of definition for both measures is [0,1]), they can be mapped on the same diagram. Yet, it is important to keep in mind that these are two qualitatively different measures that cannot be set in a numerical relation to each other (for example, a statement such as "the acceptability of WH-in-situ object questions is nearly twice as high as the frequency of WH-in-situ object-questions" would not make sense). Also one should keep in mind that frequency, unlike acceptability, is represented in a relative way, namely as a relative frequency or the proportion of one WH-variant among all other WH-variants. The best way to understand Figure 4 is to see it as a superposition of two sheets of tracing paper, one with a diagram on frequency and the other on acceptability.

Unsurprisingly, the judgment values of all sentences are within the range of acceptable constructions. In terms of comparison the judgment test also included several ungrammatical or suboptimal constructions, which obtained visibly lower scores. For example, a multiple WH-question with superiority violation such as *qu'achète qui ce soir?* 'what buys who tonight?' received the average acceptability value of 0.27 (this is not shown in Figure 4).

The comparison of the frequency and acceptability values in Figure 4 reveals several facts outlined in the following section.

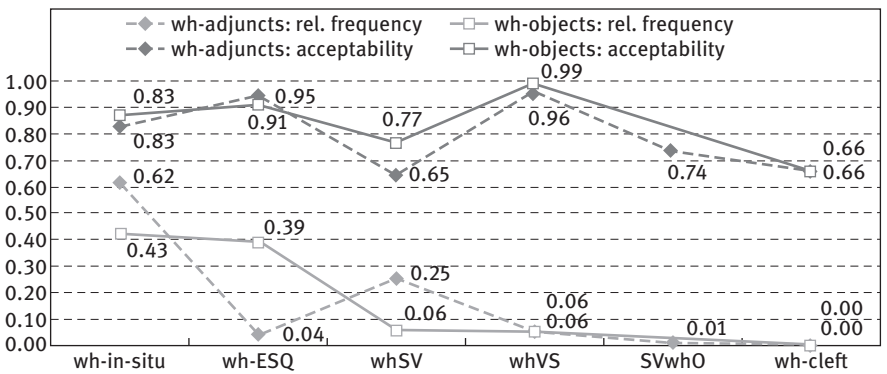


Figure 4: Relative frequency and gradient acceptability of different word order variants of WH-adjunct and WH-object questions with pronominal subject

4 Comparison of frequency and acceptability of French WH-variants

4.1 Gradience in acceptability, (near-)zero frequency

While all WH-variants score high enough in the judgment test to be considered within the range of acceptable constructions (i. e. neither ungrammatical nor marginal), with differences being gradual in nature, we do find categorical differences on the frequency side: We can distinguish occurring from (nearly) non-occurring forms.

WH-clefts such as (8a) and (8b) and the marked WH-in-situ order SVWHO such as (5) essentially do not occur at all. Several other constructions occur very rarely, namely the subject-clitic inversion WHVS such as (4a) and (4b), the WHSV order with WH-objects such as (3b) and the WH-ESQ form with WH-adjuncts such as (2a). Table 1 and Figure 2 have shown that the preferred order in usage is WH-in-situ. This observation is unambiguous for ordinary, non-clefted questions, and it also seems to apply to clefted questions (though the very low numbers of clefts makes this last claim somewhat speculative).

Yet, the frequencies suggest two hypotheses to be pursued in future research: The extreme scarcity of WH-cleft questions (of any type) is somewhat puzzling. Either contrastive focus itself is a very scarce phenomenon in spontaneous speech or contrastive focus is mainly expressed by prosodic and not syntactic means in French WH-questions. What we can observe is that a syntactic device, namely clefting, exists in French grammar but is hardly ever put to use. If we assume that contrastive focus as such is not an extremely scarce phenomenon in spontaneous speech interrogatives, we have to conclude that WH-cleft constructions are not a standard form of expressing contrastively focused WH-questions in French, contradicting Zubizarreta and Vergnaud (2005). These questions call for research at the syntax-phonology interface where the context of each sentence would be carefully analyzed, too. The question of whether contrastive focus in French WH-questions is marked by prosodic rather than by syntactic cues remains an open one.

The scarcity of the marked WH-in-situ order SVWHO with postposed object as in (5), which – as the judgments show – is within the range of acceptable constructions, is also somewhat surprising. One possible analysis would be that a construction like (5), repeated as (11a), is derived from a construction with a right-dislocated object as in (11b) by omitting the coreferential clitic pronoun in a process similar to topic drop.

(11a)	*tu*	*fais*	*quand*	*le*	*dessin?*
	you	make	when	the	drawing

(11b) *tu* *le*$_i$ *fais* *quand* *le* *dessin*$_i$*?*
you CL make when the drawing

A follow-up analysis reveals that the frequency of occurrence of right-dislocated objects among all WH-questions is also very scarce: We only find 6 occurrences. Thus, the scarcity of (5) is not surprising under the assumption that the SVWHO order is derived from right-dislocated objects.

The scarcity of subject-clitic inversion as in (4a) and (4b) in spontaneous speech – a phenomenon already observed by Coveney (1996) and Culbertson (2010) – is in line with a clear distinction between "standard" French, the variety at pace with normative considerations (and also employed for writing) and colloquial French. This distinction can be expressed by a model of diglossia (Zribi-Hertz 2010) or generalized bilingualism (Meisel et al. 2011) of French native speakers.

4.2 Gradience in both acceptability and frequency

We observe clear contrasts between WH-adjunct questions and WH-object questions in frequency for those three word orders that occur somewhat regularly or that at least are not very scarce (namely WH-in-situ (1a)/(1b), WH-ESQ (2a)/(2b) and WHSV (3a)/(3b). However, these contrasts are very subtle (WH-ESQ and WHVS) or non-existent (WH-in-situ) in acceptability. One reason for this observation is the already-mentioned and surprising fact that WH-ESQ adjunct questions and WHSV-object questions hardly occur in usage. What is more, the adjunct-object asymmetry has opposite directions in frequency and acceptability for the WH-ESQ and the WHSV order. We will therefore proceed to follow-up analyses of the adjunct-object-asymmetry in frequency, which should provide some answers to these puzzling facts.

4.3 Different preferences in acceptability and frequency

The preferences for certain WH-variants revealed by the judgments (recall that all are nuances within the range of acceptable constructions) do not match the pattern in usage. This overall acceptability-frequency mismatch is most salient for the WHVS form (4a)/(4b) (subject-clitic inversion receives the highest acceptability scores and hardly occurs in usage), and is also fairly clear for the WH-in-situ form (1a)/(1b) (its very high frequency is not reflected in the acceptability scores). Interestingly, these two variants have a "non-neutral" register or style value, with the WH-in-situ form being [+colloquial] and the subject-clitic-inversion [+formal].

To make the picture complete: WHSV (3a)/(3b) and SVWHO (5) are also [+colloquial], while WH-ESQ (2a)/(2b) is often described as "neutral" (Behnstedt 1973: 104; Coveney 1996: 98) in the sense that it fits into several registers. I represent the register-neutrality of WH-ESQ by the presence of both [+colloquial] and [+formal]. Dufter (2008) shows that *c'est* clefts occur 2.5 times more often in corpus data of spoken French compared to corpus data of written French. My interpretation of his result is that WH-clefts, as scarce as they may be, are [+colloquial] (or better, they tilt towards the [+colloquial] side). Hence, the two variants with the highest acceptability values (WH-ESQ (2a)/(2b) and WHVS (4a)/(4b)) are precisely those forms which bear the [+formal] feature. What does this result mean for the relation between acceptability and frequency? It seems that speakers cannot *not* take the normative perspective into consideration when making acceptability judgments. Please recall that subjects were thoroughly instructed to leave aside the normative perspective and to rely on colloquial language. I come back to this point in Section 5.

4.4 The issue of granularity and the analysis of frequency

In order to understand the adjunct-object-asymmetries described in Sections 4.2 and 4.3, we will increase the level of granularity in the corpus queries. So far, we have analyzed full root-WH questions with a weak pronominal subject, comparing (the aggregation of all lexical/grammatical types of) WH-adjuncts with (the aggregation of all lexical/grammatical types of) WH-objects. Please note that it is useful and often necessary to aggregate lexical/grammatical subtypes: It allows us to cover a range of constructions and thus to further generalize the findings, and what is more, it helps to reduce the problem of data scarcity. To put it in methodological terms: There is always a trade-off between internal validity and consistency (fine-grained query, i. e. fewer data) and external validity and feasibility (coarse-grained query, i. e. more data); for further discussion on the grain problem, see Manning 2003 and Crocker and Keller 2006.

Granularity is a minor issue for the acceptability judgment test. The test sentences were constructed – in line with standard experimental methodology – in a consistent manner (see Appendix): In the present study, the WH-adjunct questions are all *quand* questions ('when' questions) and the WH-object questions are all *qui* questions ('who(m)' questions).

Even though a very fine-grained control of constructions is often undesirable for a number of syntactic corpus queries, here we need to further match the constructions in the corpus with the test sentences of the acceptability judgment. To this end, we have split up WH-adjunct questions into three categories: WH-

REASON questions, WH-TIME questions and other WH-adjunct questions. Each category was further subdivided into questions with (phonologically lighter) simple WH-words (e. g. *pourquoi* 'why', *quand* 'when', *où* 'where', *comment* 'how') and (phonologically heavier) discourse-linked and/or prepositional WH-expressions (e. g. *pour quelle raison* 'for which reason', *dans quelle pièce* 'in which room', *de quelle manière* 'which way'). The separation by these categories was motivated as follows: First, WH-TIME questions match the test sentences of the acceptability judgments. Furthermore, there are good reasons to believe that TIME adjuncts are placed higher in the syntactic tree than many other adjuncts (e. g. manner, place) (see e. g. Rigau 2002, who adjoins the former to IP and the latter to VP). Second, WH-REASON adjuncts show a particular behavior in many languages: For example, only WH-REASON questions allow preverbal subjects (as opposed to unmarked postverbal subjects) in all Spanish varieties (Torrego 1984; Gutiérrez-Bravo 2006; Adli 2010b). Stepanov and Tsai (2008) argue in a cross-linguistic study that WH-REASON (and WH-PURPOSE) questions differ from other WH-adjuncts by their very high position in the tree. They place them in a high layer of the CP-system. Third, all other WH-adjuncts remain aggregated in order to minimize problems of data scarcity.

With regard to WH-objects, we distinguished (as with WH-adjuncts) between simple WH-words and D-linked and/or prepositional WH-expressions. In addition, we distinguished between [+human] and [-human] WH-objects. Please note that the fine-grained analysis of WH-objects is based on the data of a subsample of N=48 speakers – unlike the rest of our quantitative analyses which builds on the results of 101 speakers: Animacy of referential expressions was later added to the annotations, but only for roughly half of the sample. Nevertheless, 48 speakers still represent a sufficient sample size. Moreover, we compare further arithmetic means of individual relative frequencies below, thus the different sample sizes are also unproblematic from this technical perspective. We find 196 instances of [-humain] WH-object questions within this subsample. However, [+human] WH-objects (e. g. *qui* 'who' or *quelle personne* 'which person') only occur three times and cannot be analyzed due to data scarcity.[4] The extreme frequency discrepancy between [+human] and [-human] WH-object questions might be – for reasons not yet fully understood – a general yet surprising property of spontaneous speech: our findings are in line with the distribution in the Ottawa-Hull corpus, where *que/quoi* 'what' were identified by Elsig (2009: 157) 434 times compared to only 14 instances of *qui* 'who(m)'.

4 We would like to add that WH-indirect questions would not be analyzable either because they occur only three times in the entire corpus

The fine-grained analysis outlined above allows us to construct an exact match between the test sentences of the acceptability judgments and the data from spontaneous speech with regard to WH-adjunct questions (with *quand*). However, it does not allow a match with regard to WH-object questions because of the scarcity of [+human] WH-objects in the corpus.

Relative frequencies according to (10b) are calculated based on all eight subtypes within a category, i. e. four word orders and two weights of the WH-element.[5] For example, many WH-time questions can, at least theoretically, be expressed by one of the four word orders and either by *quand* or a complex WH-expression.[6]

4.5 Results based on the fine-grained corpus queries

Table 2 reveals all in all 37 WH-REASON questions, most of them with *pourquoi*. We also count 171 WH-TIME questions, most of which are realized by non-simple forms (e. g. à quelle heure 'at which hour'). Furthermore, there are 221 other WH-adjuncts: The most frequent one in this category is the MANNER adjunct *comment* 'how', the second most frequent one is the PLACE adjunct *où* ('where'). There are also some non-simple WH-elements, mostly PLACE (e. g. à quel étage, dans quelle pièce, dans quelle domaine) and MANNER adjuncts (e. g. *de quelle manière, dans quelles circonstances, en quels termes*).

Table 2: Relative and absolute frequencies of different types of WH-adjunct and WH-object questions

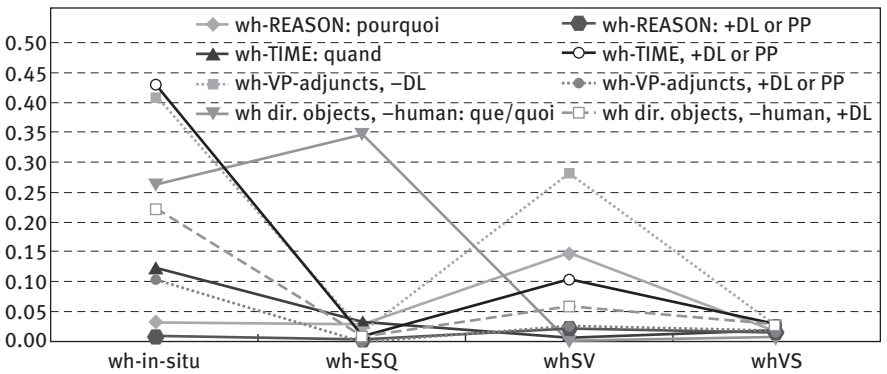
	WH-REASON: *pourquoi*	WH-REASON, +DL/PP	WH-TIME: *quand*	WH-TIME, +DL/PP	WH-VP-adjuncts, –DL	WH-VP-adjuncts, +DL/PP	WH dir. objects, –human: *que/quoi*	WH dir. objects, –human, +DL
WH-in-situ	0.03 (5)	0.01 (1)	0.12 (24)	0.43 (94)	0.41 (94)	0.1 (29)	0.26 (54)	0.22 (42)
WH-ESQ	0.03 (3)	0 (0)	0.04 (11)	0.01 (3)	0.02 (7)	0 (0)	0.35 (79)	0 (1)
WHSV	0.15 (19)	0.02 (3)	0 (1)	0.1 (26)	0.28 (73)	0.02 (7)	0 (0)	0.06 (13)
WHVS	0.01 (5)	0.01 (1)	0 (0)	0.03 (10)	0.03 (6)	0.02 (3)	0.01 (3)	0.03 (4)

5 The rows for the SVwhO, WH-cleft and WH-in-situ cleft variants can be disregarded: With just four tokens across all WH-types, their relative frequencies are mostly 0.00 (in two cells they are 0.01).

6 This reasoning in terms of true variation of course does not apply to the category of other WH-adjuncts. The envelope in this case is more approximate.

The coarse-grained analysis in Figure 4 above reveals that the WH-in-situ order is the most frequent variant for both WH-adjunct and WH-object questions. However, the fine-grained analysis in Figure 5 now exhibits two constructions with a different pattern, having a dispreference for WH-in-situ: (i) the WH-REASON question with *pourquoi* is preferred with the WhSV form and (ii), the WH-object question with *que/quoi* is preferred with the WH-ESQ form.

Figure 5: Relative frequencies of different types of WH-adjunct and WH-object questions



With regard to (i) we notice a clear dispreference of the REASON-adjunct *pourquoi* for the WH-in-situ order. This finding can be seen as corpus-linguistic evidence in support of the claim made by Stepanov and Tsai (2008) according to which REASON-why is base-generated in a very high CP-position. Under their assumption we should not find any instance of in-situ REASON-why, yet Table 2 contains some sporadic cases. A new inspection of these sporadic cases reveals that all except for one (*et ils étaient louches pourquoi?* 'and why were they indecent?') are ambiguous between a REASON and PURPOSE interpretation (e. g. *t'es montée pourquoi?* 'why did you go up?'). Thus, the data seem to support the authors' claim that REASON-why, unlike PURPOSE-why, cannot occur in-situ.

With regard to (ii), we can first state that the high relative frequency of WH-object questions with WH-ESQ order in Figure 4 is not due to questions with D-linked WH-elements but to questions with monosyllabic *que* (recall that Modern French displays a complementary distribution in WH-object questions with *quoi* occurring in-situ and *que* in all other positions). The *est-ce que* question particle occurs mostly with *que*, which is in line with Elsig (2009) and Druetta (2002, 2003, 2008: 127). Although the absolute frequencies in Table 2 also point at a few cases of (simple) WH-adjuncts with *est-ce que*, their relative frequencies remain minimal. When a variable has variants with relative frequencies close to zero and at the same time other variants with much higher values, the near-zero-variants

might be instable phenomena, in which case a "small push" could bring them to extinction. We conclude that *est-ce que* is limited to simple, i. e. phonologically light, mostly monosyllabic WH-forms: The only clearly stable WH-construction occurring with *est-ce que* is the WH-object question with *que*. In the course of its grammaticalization since its appearance in the sixteenth century (Foulet 1921: 265), *est-ce que* has lost the meaning of the source construction in WH-questions (Druetta 2003): *Est-ce que* WH-questions are no longer emphatic. This source meaning only remains – to a minor extent – in present-day yes/no questions (Mosegaard-Hansen 2001: 471). In WH-questions, *est-ce que* has thus become a neutral, redundant interrogative particle (redundant because the interrogative feature is already expressed by the WH-element). The fact that *est-ce que* is clearly limited to *que* in modern spoken French suggests a further change: *Est-ce que* now primarily functions as a morpho-phonological host for the WH-clitic *que*. Interestingly, *que* differs from all other French WH-words in that it is not an independent word but a proclitic requiring a host (Poletto and Pollock 2004): It can either cliticize to a verb (*qu'as-tu dit* 'what have you said') – however, the WHVS order is, as we assume, not part of colloquial Modern French grammar. Alternatively, it can cliticize to *est-ce que* – which is thus the only remaining option for fronting *que* in the colloquial variety.

Finally, we notice that the mismatch between acceptability and frequency for the WHSV order shown in Figure 4 would be less pronounced if we took into account the relative frequency of *quand* questions (see Figure 5). The patterns of the very low relative frequency of *quand* and the low relative frequency of WH-objects are similar to the acceptability values of WHSV WH-adjunct and WH-object questions: These judgment values are rather low within the overall picture of all WH-variants. Furthermore, WH-adjunct questions have a slightly lower acceptability value than WH-object questions in Figure 4, i. e. the directionality of the acceptability-frequency mismatch between argument and adjunct questions disappears if we restrict ourselves to *quand* questions in the corpus.

5 Discussion

Having discussed the frequency of WH-ESQ questions in spontaneous speech, we will now compare their relative frequencies with their acceptability ratings. Please recall that the test sentences for the acceptability judgment test either contain the WH-object *qui* 'who(m)' as in (2a) or the WH-adjunct *quand* 'when' as in (2b). Although WH-ESQ questions with [+human] WH-objects or WH-adjuncts are extremely scarce or do not occur at all in usage (Table 2), they score high in acceptability (Figure 3). While the role of *est-ce que* is essentially limited to

being a clitization host for *que* in colloquial French, it remains a broadly available, optional interrogative particle in WH-questions in standard French. It is interesting to note – also as an anecdote on standard French prescriptivism – that *est-ce que* was only approved by the Académie Française in the 1930s – to be "disapproved" again in 1987 (Grevisse 1993: 605/606). Nowadays, the WH-ESQ question is neither considered elegant nor "popular" from a normative point of view. Rather, it can be described as neutral. In Section 4.3, this register neutrality has been represented as the coexistence of both a [+formal] and a [+colloquial] feature. That being said, how does the normative influence act on the acceptability judgments?

First, the fact that neutral WH-ESQ, like formal WHVS, scores highest in acceptability indicates that normative influence or bias on judgments does not seem to act as a *bonus* for the standard variety, but rather as a *malus*/cost for constructions that are [+colloquial] only. Interestingly, colloquial WH-in-situ is not among the constructions that scored highest, either.

The acceptability judgments on the highly formal complex inversion question – which is not discussed in detail in the present study because it does not belong to the envelope of variation – also support this assumption. WH-object questions with complex inversion such as (6b) received a mean acceptability score of 0.8, and WH-adjunct questions with complex inversion such as (6a) received a score of 0.94 – yet they hardly ever occur (see Table 1). A comparison with the other ratings in Figure 4 reveals that (6a) receives a relatively high score – irrespective of its particularly high level of formality. Importantly, normative influence does not explain any categorical difference in terms of acceptability vs. unacceptability, but it is one of the factors behind the systematic nuances within the range of acceptable constructions.

Second, we can observe a difference in the span of registers reflected by acceptability and frequency data. While frequency data from spontaneous speech (excluding highly formal speech contexts) provide insight into colloquial language, acceptability data reflect the entire range of registers available to a speaker. The results suggest that it is difficult for speakers to judge a variant that does not exist in register x as unacceptable, as long as it exists in register y. In other words, speakers seem to accept a construction if it belongs to any register of their language. However, based on the present results it is still unclear whether this phenomenon only occurs if register y is higher than register x, i. e. whether speakers are only unable to disregard those constructions with a higher stylistic value. The effect of register spanning in judgments is at least one part of the explanation as to why certain constructions hardly occur although they are rated as acceptable. It might also offer a diagnostic tool to distinguish between diglossia and bilingualism. Bilingualism would be characterized by a better

capacity to keep the languages apart when performing acceptability judgments. From this point of view, French speakers are diglossic rather than bilingual.

We can thus conclude that acceptability judgments cover a broad range of registers but are at the same time tinted by norm. They are tinted more or less, depending on the weight of such norms and the plainness of diglossia in a speech community. The results of the present study also suggest that the analogy often drawn between acceptability judgments and the laws of perception of psychophysical stimuli (Bard et al. 1996) could be a myth and too much of an idealization. Many linguists would agree that norms weigh rather more than less in France today, where the "ideology of the standard [...] is specially vigorous" (Gadet 2007: 27). Given that normative influences are a sociolinguistic phenomenon, a straightforward follow-up question for future research is whether or not this influence is subject to social variation; if so, this would mean that the relation between acceptability and frequency as such is subject to social variation. It has been shown that acceptability judgments can reflect systematic social differences (Adli 2013). In statistical terminology, this hypothesis would entail that the specific form of interaction between acceptability and frequency is socially dependent.

Ironically, acceptability judgments, which have long had a bad reputation in sociolinguistics, offer an interesting yet unexploited potential for sociolinguistic studies: They contribute to determining the envelope of variation in syntax (by taking into account acceptable but scarce constructions). Furthermore, they can help to identify normative influences. A promising path for both sociolinguistic and syntactic research is to work with both types of data, combining forces, and carefully laying out which aspects each type of data can and cannot reveal. Finally, the combination of acceptability and frequency can help to analyze syntactic change in progress more precisely: This approach can help to identify constructions that are no longer in use, but still exist in a non-colloquial variety of the speech community.

References

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

Adli, Aria (2006): French wh-in-situ Questions and Syntactic Optionality: Evidence from three data types. *Zeitschrift für Sprachwissenschaft* 25(2): 163–203.

Adli, Aria (2010a): Constraint Cumulativity and Gradience: Wh-Scrambling in Persian. *Lingua* 120(9): 2259–2294.

Adli, Aria (2010b): The Semantic Role of the Wh-Element and Subject Position in Spanish and Catalan. *STUF – Language Typology and Universals* 63(2): 103–117.

Adli, Aria (2011a): *Gradient Acceptability and Frequency Effects in Information Structure: a quantitative study on Spanish, Catalan, and Persian*. Habilitationsschrift, Universität Freiburg.

Adli, Aria (2011b): A Heuristic Mathematical Approach for Modeling Constraint Cumulativity: Contrastive Focus in Spanish and Catalan. *The Linguistic Review* 28(2): 111–173.
Adli, Aria (2011c): On the Relation between Acceptability and Frequency. In: Esther Rinke and Tanja Kupisch (eds.), *The development of grammar: language acquisition and diachronic change – In honour of Jürgen M. Meisel*, 383–404. Amsterdam/New York: John Benjamins.
Adli, Aria (2013): Syntactic Variation in French Wh-Questions: a quantitative study from the angle of Bourdieu's sociocultural theory. *Linguistics* 51(3): 473–515.
Auger, Julie (1994): *Pronominal Clitics in Quebec Colloquial French: a Morphological Analysis*. PhD dissertation, University of Pennsylvania.
Ayres-Bennett, Wendy (1994): Negative Evidence: Or Another Look at the Non-Use of Negative ne in Seventeenth-Century French. *French Studies* 48: 63–85.
Baayen, R. Harald, Cristina Burani and Robert Schreuder (1997): Effects of semantic markedness in the processing of regular nominal singulars and plurals in Italian. In: Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1996*, 13–33. Dordrecht: Kluwer.
Backus, Ad and Maria Mos (2011): Islands of (Im)Productivity in Corpus Data and Acceptability Judgments: Contrasting Two Potentiality Constructions in Dutch. In: Doris Schönefeld (ed.), *Converging Evidence: Methodological and Theoretical Issues for Linguistic Research*, 165–192. Amsterdam: John Benjamins.
Bader, Markus and Jana Häussler (2010): Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2): 273–330.
Bard, Ellen Gurman, Dan Robertson and Antonella Sorace (1996): Magnitude Estimation of Linguistic Acceptability. *Language* 72(1): 32–68.
Behnstedt, Peter (1973): *Viens-tu? Est-ce que tu viens? Tu viens ? Formen und Strukturen des direkten Fragesatzes im Französischen*. Tübingen: Narr.
Bybee, Joan L. and David Eddington (2006): A Usage-based Approach to Spanish Verbs of 'Becoming'. *Language* 82(2): 323–355.
Chomsky, Noam (1965): *Aspects of the Theory of Syntax*. Cambridge: MIT Press.
Coveney, Aidan B. (1996): *Variability in Spoken French. A Sociolinguistic Study of Interrogation and Negation*. Exeter: Elm Bank.
Coveney, Aidan and Laurie Dekhissi (2013): *Variation dans l'emploi des interrogatives partielles dans le cinéma de banlieue*. Paper presented at 'La syntaxe des interrogatives', Neuchâtel.
Crocker, Matthew and Frank Keller (2006): Probabilistic grammars as models of gradience. In: Gisbert Fanselow, Caroline Fery, Ralf Vogel and Matthias Schlesewsky (eds.), *Gradience in Grammar*, 227–245. Oxford: Oxford University Press.
Culbertson, Jennifer (2010): Convergent evidence for categorial change in French: From subject clitic to agreement marker. *Language* 86(1): 85–132.
Dekhissi, Laurie (in prep.): *Un nouveau français populaire?* PhD dissertation, University of Exeter.
Déprez, Viviane, Kristen Syrett and Shigeto Kawahara (2013): The interaction of syntax, prosody, and discourse in licensing French wh-in-situ questions. *Lingua* 124: 4–19.
Drijkoningen, Frank and Brigitte Kampers-Manhe (2008): On inversions and the interpretation of subjects in French. *Probus* 20(2): 147–209.
Druetta, Ruggero (2002): *Qu'est-ce tu fais*? État d'avancement de la grammaticalisation de *est-ce que*. Première partie. *Linguae* 2: 67–88.
Druetta, Ruggero (2003): *Qu'est-ce tu fais*? État d'avancement de la grammaticalisation de *est-ce que*. Deuxième partie. *Linguae* 1: 21–35.
Druetta, Ruggero (2008): *La question en français parlé : étude distributionnelle*. Torino: Trauben Edizioni.

Dufter, Andreas (2008): On explaining the rise of c'est-clefts in French. In: Ulrich Detges and Richard Waltereit (eds.), *The Paradox of Grammatical Change: Perspectives from Romance*, 31–56. Amsterdam: John Benjamins.

Elsig, Martin (2009): *Grammatical Variation across Space and Time – The French Interrogative System*. Amsterdam/Philadelphia: John Benjamins.

Escandell-Vidal, Victoria (2002): Echo-syntax and metarepresentations. *Lingua* 112: 871–900.

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/New York: Mouton de Gruyter.

Foster, Jennifer (2007): Real bad grammar: Realistic grammatical description with grammaticality. *Corpus Linguistics & Linguistic Theory* 3(1): 73–86.

Foulet, Lucien (1921): Comment ont évolué les formes de l'interrogation. *Romania* 47: 243–348.

Freyd, Max (1923): The graphic rating scale. *Journal of Educational Psychology* 14: 83–102.

Funke, Frederik (2010): *Internet-Based Measurement With Visual Analogue Scales: An Experimental Investigation*. PhD dissertation, Universität Tübingen.

Gadet, Françoise (2007): *La variation sociale en français (2ème édition)*. Paris: Ophrys.

Greenbaum, Sidney (1976): Syntactic Frequency and Acceptability. *Lingua* 40: 99–113.

Greenbaum, Sidney (1977): Judgments of Syntactic Acceptability and Frequency. *Studia Linguistica: Revue de Linguistique Generale et Comparee/Journal of General and Comparative Linguistics* 31: 83–105.

Greenberg, Joseph H. (1966): *Language universals: with special reference to feature hierarchies*. The Hague: Mouton.

Grevisse, Maurice (1993): *Le bon usage: grammaire française*. Paris: Duculot.

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.

Hamlaoui, Fatima (2011): On the role of phonology and discourse in Francilian French wh-questions. *Journal of Linguistics* 47(01): 129–162.

Haspelmath, Martin (2006): Against Markedness (and What to Replace It With). *Journal of Linguistics* 42(1): 25–70.

Kayne, Richard S. and Jean-Yves Pollock (1978): Stylistic inversion, successive cyclicity and move-NP in French. *Linguistic Inquiry* 9: 595–621.

Kempen, Gerard and Karin Harbusch (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/New York: Mouton de Gruyter.

Kempen, Gerard and Karin Harbusch (2008): Comparing Linguistic Judgments and Corpus Frequencies as Windows on Grammatical Competence: A Study of Argument Linearization in German Clauses. In: Anita Steube (ed.), *The Discourse Potential of Underspecified Structures*, 179–192. Berlin/New York: de Gruyter.

Labov, William (1982): Building on empirical foundations. In: Winfred P. Lehmann and Yakov Malkiel (eds.), *Perspectives on Historical Linguistics*, 17–92. Amsterdam/Philadelphia: John Benjamins.

Labov, William (1996): When Intuitions fail. In: Lisa McNair, Kora Singer, Lise M. Dolbrin and Micgelle M. Aucon (eds.), *Papers from the Parasession on Theory and Data in Linguistics. Chicago Linguistic Society 32*, 77–106. Chicago: Chicago Linguistic Society.

Lambrecht, Knud (2001): A framework for the analysis of cleft constructions. *Linguistics* 39(3): 463–516.
Manning, Christopher D. (2003): Probabilistic Syntax. In: Rens Bod, Jennifer Hay and Stefanie Jannedy (eds.), *Probabilistic Linguistics*, 289–341. Cambridge: MIT Press.
Meisel, Jürgen M., Martin Elsig and Matthias Bonnesen (2011): Delayed Acquisition of Grammar in First Language Development: Subject-Verb Inversion and Subject Clitics in French Interrogatives. *Linguistic Approaches to Bilingualism* 1(4): 347–390.
Mosegaard-Hansen, Maj-Britt (2001): Syntax in interaction: Form and function of yes/no interrogatives in spoken standard French. *Studies in Language* 25(3): 463–520.
Newmeyer, Frederick J. (2010): Accounting for rare typological features in formal syntax: three strategies and some general remarks. In: Jan Wohlgemuth and Michael Cysouw (eds.), *Rethinking Universals*, 195–221. Berlin/New York: Mouton de Gruyter.
Poletto, Cecilia and Jean-Yves Pollock (2004): On wh-clitics and wh-doubling in French and some North Eastern Italian dialects. *Probus* 16(2): 241–272.
Poplack, Shana (1989): The care and handling of a mega-corpus. In: Ralph Fasold and Deborah Schiffrin (eds.), *Language Change and Variation*, 411–451. Amsterdam: John Benjamins.
Prieto, Pilar and Gemma Rigau (2007): The Syntax-Prosody Interface: Catalan interrogative sentences headed by *que*. *Journal of Portuguese Linguistics* 6(2): 29–59.
Pullum, Geoffrey K. (2007): Ungrammaticality, rarity, and corpus use. *Corpus Linguistics & Linguistic Theory* 3(1): 33–47.
Reis, Marga (1991): Echo-w-Sätze und Echo-w-Fragen. In: Marga Reis and Inger Rosengren (eds.), *Fragesätze und Fragen. Referate anlässlich der Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, Saarbrücken 1990*, 49–76. Tübingen: Niemeyer.
Rigau, Gemma (2002): Els complements adjunts. In: Joan Solà, Maria-Rosa Lloret, Joan Mascaró and Manuel Pérez-Saldanya (eds.), *Gramàtica del català contemporani*, 2045–2110. Barcelona: Empúries.
Rijkhoff, Jan (2010): Rara and grammatical theory. In: Jan Wohlgemuth and Michael Cysouw (eds.), *Rethinking Universals*, 223–239. Berlin/New York: Mouton de Gruyter.
Sampson, Geoffrey R. (2007): Grammar without grammaticality. *Corpus Linguistics & Linguistic Theory* 3(1): 1–32.
Sobin, Nicholas (2010): Echo Questions in the Minimalist Program. *Linguistic Inquiry* 41(1): 131–148.
Stefanowitsch, Anatol (2008): Negative entrenchment: A usage-based approach to negative evidence. *Cognitive Linguistics* 19(3): 513–531.
Stepanov, Arthur and Wei-Tien Dylan Tsai (2008): Cartography and licensing of wh- adjuncts: a cross-linguistic perspective. *Natural Language & Linguistic Theory* 26(3).
Torrego, Esther (1984): On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15(1): 103–129.
Trubetzkoy, Nikolaus (1939): *Grundzüge der Phonologie.* Göttingen: Vandenhoeck & Ruprecht.
Zribi-Hertz, Anne (2010): Pour un modèle diglossique de description du français: quelques implications théoriques, didactiques et méthodologiques. *Journal of French Language Studies* FirstView: 1–26.
Zubizarreta, Maria Luisa and Jean-Roger Vergnaud (2005): Phrasal Stress, Focus, and Syntax. In: Martin Everaert and Henk van Riemsdijk (eds.), *The Syntax Companion, Vol. 3*, 522–568. Malden: Blackwell.

Appendix: Experimental material of the gradient acceptability judgment test

reference sentence: *Tous ont regardé qui ?*

	wh-adjunct	**wh-object**
WH-in-situ (1a)/(1b):	*Tu allumes le feu quand ?* *Tu nettoies la cuisine quand ?* *Tu enlèves ton pansement quand ?*	*Tu emmènes qui en vacances ?* *Tu reçois qui dans ton bureau ?* *Tu félicites qui à la cérémonie ?*
WH-ESQ (2a)/(2b)	*Quand est-ce que tu rends ton livre ?* *Quand est-ce que tu récupères ta voiture ?* *Quand est-ce que tu prends ton médicament ?*	*Qui est-ce que tu rejoins à la piscine ?* *Qui est-ce que tu amènes à la maison ?* *Qui est-ce que tu invites au cinéma ?*
WHSV (3a)/(3b)	*Quand tu finis ton projet ?* *Quand tu achètes le vélo ?* *Quand tu signes le contrat ?*	*Qui tu sers à table ?* *Qui tu accueilles chez toi ?* *Qui tu soutiens aux élections ?*
WHVSclit (4a)/(4b):	*Quand manges-tu le gâteau ?* *Quand ouvres-tu le cadeau ?* *Quand peints-tu la façade ?*	*Qui vois-tu cet après-midi ?* *Qui attends-tu chaque lundi ?* *Qui remplaces-tu demain ?*
SVWHO (5)	*Tu continues quand le repassage ?* *Tu jettes quand la poubelle ?* *Tu évalues quand les résultats ?*	
WH-cleft (8a)/(8b)	*Quand c'est que tu remplis le formulaire ?* *Quand c'est que tu écris ton livre ?* *Quand c'est que tu répares la moto ?*	*Qui c'est que tu entends dans ce hall ?* *Qui c'est que tu conduis à l'aéroport ?* *Qui c'est que tu déranges à la bibliothèque ?*
	items from the instruction phase *Qui ont-ils tous regardé ?* *Qui tous ont regardé ?* *Tous ont regardé qui ?* *Tous sont regardés qui ?* *Que sont-ils tous regardés qui ?*	**items from the training phase** *Qui c'est que tu accompagnes à la gare ?* *Que contrôle quel douanier à la frontière ?* *Tu fermes la porte quand ?* *Dis-moi, Jérémie a pas balayé quoi ?*