ARIA ADLI (Hamburg)

The semantic role of the *wh*-element and subject position in Spanish and Catalan

Abstract

This work presents experimental results on the position of the subject in *wh*-questions of the Spanish variety of Catalonia and of Catalan. Gradient acceptability judgments have been collected with a sample of native speakers, which were fairly balanced bilinguals. It is attempted to account for the striking similarity between the judgment patterns of the Spanish variety of Catalonia and Catalan, and to model at the same time the nuanced, but systematic differences. The similarities are explained by assuming the same set of rules or constraints, regarding the left-peripheral order of a shift topic and a focal element, and regarding the positional restrictions of subjects co-occurring with an argumental *wh*-element. However, the violation of a rule provokes slightly different degrees of markedness in Spanish and Catalan. This gradient phenomenon is captured by the idea that the violation of a rule comes with a specific cost, which can differ between languages, and that the degree of acceptability of a construction is a function of cumulative violation costs.

Keywords: postverbal subjects, Spanish, Catalan, markedness

1. Introduction

A particularity of today's linguistic situation in Catalonia is the fact that there are basically no Catalan monolinguals. Variation only consists in the degree of balance or language dominance. Contact with Spanish has become a constitutive element of Catalan. In set-theoretical terms, we have a set of Spanish monolinguals, an intersection of both Spanish and Catalan ($S \cap C$), but no set of Catalan monolinguals. Thus Catalan speakers are a proper subset of Spanish speakers: $C \subsetneq S$. Such a situation also raises the question of mutual influence and contact-induced change. Many phenomena of change are gradient (which does not exclude that at some point there might also be 'sudden jumps' in a system). Although the issue of change is not at the core of the present study, which is restricted to a snapshot of a present state, it nevertheless proposes a methodology that captures similarities and differences in a gradient manner, and thus allows formulating meaningful hypotheses with respect to possible effects of mutual language influence. Changes in the system, including contact-induced change, would first become manifest in fine changes of the violation cost of a rule. This approach can prove useful for typological studies, because it accounts for (i) macro-variational, categorical differences of grammaticality by assuming different sets of rules or constraints, (ii) micro-variational, gradient differences of markedness by assuming different violation costs between languages for the same rules, and (iii) the transitional, partly gradient process of grammatical change.

Already MEYER (1972) has pointed out that a number of Spanish sentences seem to require the subject to be in postverbal position. The same has been observed for Catalan

(BONET 1990). The following sentences show four pairs of *wh*-questions with pre- vs. post-verbal subject. (1a) to (4b) are Spanish, (5a) to (8b) are Catalan constructions. The grammaticality status of these sentences (e.g. starred vs. unstarred) is not indicated by purpose, because it is one of the goals of the gradient judgment test to determine their acceptability status, more precisely their gradient acceptability.

(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?'

(5a) *A qui aquestes aranyes li espanten?*
to whom these spiders DAT-CL frighten$_{3PL}$

(5b) *A qui li espanten aquestes aranyes*?
to whom DAT-CL frighten$_{3PL}$ these spiders
'Who fears these spiders?'

(6a) *Què la Roser va perdre abans d'ahir*?
what the Roser lost$_{PERIPHR.PRET}$ the day before yesterday

(6b) *Què va perdre la Roser abans d'ahir*?
what lost$_{PERIPHR.PRET}$ the Roser the day before yesterday
'What did Roser lose the day before yesterday?'

(7a) *Quan l' Elvira va imprimir aquest document?*
when the Elvira print$_{PERIPHR.PRET}$ this file

(7b) *Quan va imprimir l' Elvira aquest document?*
when print$_{PERIPHR.PRET}$ the Elvira this file
'When did Elvira print this file?'

(8a)   *Per què*   *l' Ernest*   *va abraçar*   *aquesta*   *noia?*
     why   the Ernest   hug$_{PERIPHR.PRET}$ this   girl

(8b)   *Per què*   *va abraçar*   *l' Ernest*   *aquesta*   *noia?*
     why   hug$_{PERIPHR.PRET}$ the Ernest   this   girl
     'Why did Ernest kiss this girl?'

The sentence pairs differ with respect to the type of the *wh*-element. The first two pairs of sentences of each language are *wh*-arguments, the last two pairs are *wh*-adjuncts. Note that the verb *gustar* ('like') in the Spanish sentences (1a) and (1b), as well as the verb *espantar* ('frighten') in the Catalan sentences (5a) and (5b) belong to the class of psychverbs (Belletti & Rizzi 1988). Following Masullo (1993) and Gutiérrez-Bravo (2006: 51), I assume that the oblique experiencer in these constructions occupies the same structural position as the subject in transitive constructions.

The debate in the literature has both a formal and a descriptive side: First, one has to try to account for the restrictions with preverbal subject on a formal level. Second, it is far from clear, which *wh*-elements can and which ones cannot co-occur with a preverbal subject. The diversity of opinions expressed for Spanish is probably, at least in part, due to dialectal variation between the authors. It might also be in part due to the fact that the debate also involves more or less marked constructions. It is not unusual that different authors have a different "level of tolerance" with such constructions, when it comes to a categorical choice in terms of ±acceptable.

2. Subject position and semantic role in *wh*-questions

An important element that distinguishes preverbal from postverbal subjects in Romance null subject languages such as Spanish, Catalan, or Italian, is their informationstructural status. There has been a long debate on the role of focus with respect to subject position. In particular, postverbal subjects have often been described as information focus, while preverbal subjects have been described as background – or when focused, then as contrastive focus (Contreras 1976; Cinque 1993; Zubizarreta 1998). According to Zubizarreta (1998: 76), prosodically prominent preverbal subjects can only carry contrastive focus (c-F), not information focus (i-F). Therefore, the subject in (9) can only be contrastive (nuclear or contrastive stress is represented by underlining). However, a construction with wide information focus as in (10) requires the subject to appear in a postverbal position. In other words, the position of the subject depends on the focus/background structure, specifically on the type of focus, whether it is contrastive or not.[1] According to Zubizarreta (1998: 127), the postverbal subject remains in the Spec of *v*P (López 2009: chap. 3.4.2 suggests the same for Catalan). She derives the VOS order in transitive constructions by overt verb movement to the Spec of T, driven by a strong V-feature in T, and subsequent adjunction of the entire VP to the left of *v*P, as shown in (11).

---

[1]   The story behind the relation between subject position and focus is more complex and also more controversial in Spanish. According to Zubizarreta (1998: 20–22, 1999: 4229, 4232), Ordóñez (2000: 29 ff.), Gutierrez-Bravo (2002: 51), Zagona (2002: 211), and Samek-Lodovici (2001: 347) narrow non-contrastive focus on a preverbal subject is illicit (i.e. *[S]$_{i-F}$VO). However, Toledo (1989: 226), García-Lecumberri (1995: 246, 341), and Hualde (2002: 106, 2003: 160) accept [S]$_{i-F}$VO constructions. A detailed discussion of this controversy is out of the scope of the present paper.

(9)   [Juan]_{c-F}   compró   una   lavadora.
       Juan   bought   a   washing machine

(10)   [Compró   una   lavadora   Juan]_{i-F}
        bought   a   washing machine   Juan

(11)   [$_{TP}$ compró T [$_{vP}$ [$_{VP}$una lavadora]$_i$ [$_{vP}$ Juan$_{[+F]}$ v ~~compró~~ t_i]]]

What makes *wh*-questions interesting in the scope of a discussion on information struc-
ture, is the fact that the (non-*wh*) subject does **not** carry the focus. Rather, the *wh*-element
carries the focus (CHOMSKY 1971; JACKENDOFF 1972) and the remainder of the sentence,
including the subject, belongs to the background (apart from broad focus questions such
as *what happened?*). In other words, the theoretical approaches, in which the correlation
between subject and focus is discussed, will probably not prove useful in the case of
wh-questions. Starting with TORREGO (1984), there has been a long discussion that has
mainly centred on the nature of the *wh*-element, in order to explain the restrictions of pre-
verbal subjects in Spanish *wh*-questions.

TORREGO (1984) drew the line along the ±argumental status of the *wh*-element. Accord-
ing to her, *wh*-arguments cannot appear with preverbal subjects, while *wh*-adjuncts can,
i.e. she assumes (1a) and (2a) to be ungrammatical, and (3a) and (4a) to be grammatical.
If we formulate hypotheses for Catalan based on her view for Spanish, then (5a) and (6a)
are expected to be ungrammatical and (7a) and (8a) to be grammatical.

However, a ±argumental distinction cannot explain all facts. SUÑER (1994) points out
that the term 'argumental' has to be refined in the sense of RIZZI (1990). *Wh*-phrases like
the Spanish measure phrase *cuánto* ('how much'), selected by their predicate but **not** theta-
marked, are then [–argumental]. Otherwise, one would not be able to explain the ungram-
maticality of (12) with a preverbal subject.

(12)   *Cuánto   esta   beba   pesará ?
        how much   this   baby girl   weigh$_{3SG.FUT}$
        'Who much would this baby girl weigh?'

Moreover, ZUBIZARRETA (2007) (quoted in BEAS 2007) points to a restriction on the type
of subject (and not just on the type of *wh*-element): Nonspecific indefinites cannot appear
preverbally in a *wh*-question.

(13)   *Cuándo   alguien   hizo   esta tarta?
        when   someone   did$_{3SG.PRET}$ this cake
        'When did someone make this cake?'

I build on the view that properties of the subject are relevant. More precisely, I will pro-
pose that a more detailed analysis of its topic properties contributes to the understanding
of the positional requirements in Spanish and Catalan *wh*-questions.

BAKOVIC (1998: 37), who works in an OT syntax framework suggests replacing the distinc-
tion in terms of ±argumental by a **scale of semantic roles** of *wh*-elements (see also LARSON
1988; MARGARET 1990; PESETSKY 1995). GUTIÉRREZ-BRAVO (2006: 75) proposes the hierarchy
in (15), which is supposed to be a compilation of the scales proposed by LARSON (1988:
382), MARGARET (1990: 16), and BAKOVIC (1998: 37). Dialectal variation is reflected in such a
hierarchy by placing the "cut-off-point" at different positions. However, the distinction be-
tween grammatical and ungrammatical constructions still remains binary. For example, the

argument/adjunct distinction assumed by TORREGO (1984) (or that one can assume for her dialect), would split in the hierarchy (14) *wh*-ARGUMENT, on the one hand, from *wh*-LOCA-TION, *wh*-MANNER, and *wh*-REASON, on the other.

(14)   ARGUMENT > LOCATION > MANNER > REASON

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

GUTIÉRREZ-BRAVO (2006) proposes a modification of standard OT, in which he works with the notion of markedness constraint. The candidates in the tableau ranked according to these constraints reflect an **order of markedness**.[2] He assumes a set of constraints that are violated, if an element with a specific semantic roles is in preverbal position. Note that GUTIÉRREZ-BRAVO (2006: 33) adopts ZUBIZARRETA's (1998: 122/123) reduced phrase structure for Spanish, according to which *wh*-elements, focus, and topic do not land in C but in T (the latter might also land in a TopP above T). I am not adopting this view, but rather work with a full-fledged C-system according to RIZZI (1997). This full C-system is even more expanded, because it further includes an IntP projection, as well as FRASCARELLI's (2007) different topic phrases (more on this later).

GUTIÉRREZ-BRAVO (2006) applies harmonic alignment and derives one constraint for each semantic role in (15). Therefore, the more the semantic role appears on the left of a semantic hierarchy scale like (14) or (15), the more marked is a construction with preverbal subject.

The scale in (16) is a reduced version of (15) and includes the relevant semantic roles for the *wh*-elements in sentences (1a) to (8b) (see Table 1 below).

(16)   EXPERIENCER > THEME > TIME > REASON

3. Methodology

The sentences (1a) to (8b) above are the test constructions of this study, i.e. simple *wh*-questions with an overt subject. The design has three independent variables. Variable A distinguishes four semantic roles of the *wh*-element, namely EXPERIENCER, THEME, TIME, REASON (henceforth '*wh*-role'). Variable B distinguishes preverbal and postverbal subjects. Finally variable C refers to the language, i.e. Spanish or Catalan. Each of the constructions (1a) to (8b) is presented in two lexical variants. Given three independent variables and two lexical variants every subject gave $4 \times 2 \times 2 \times 2 = 32$ different judgments. The arithmetic mean of the acceptability values of the two lexical variants is the dependent variable. There is no repetition of nouns and verbs in the experimental material, in order to avoid artefacts such as priming.

---

[2]   The notion of markedness constraint was first proposed by MÜLLER (1999). He essentially suggests that some competitions result in more than one optimal candidate. In such cases, a second competition takes place with a new set of constraint, namely markedness constraints. Note that GUTIÉRREZ-BRAVO (2006) does not work with a two-step procedure, but incorporates all constraints in a single competition.

A: semantic role of *wh*-element / B: subject position	a_1: EXPERIENCER		a_2: THEME		a_3: TIME		a_4: REASON	
	c_1: Span.	c_2: Cat.	c_1: Span.	c_2: Cat.	c_1: Span.	c_2: Cat.	c_1: Span.	c_2: Cat.
b_1: preverbal	(1a)	(5a)	(2a)	(6a)	(3a)	(7a)	(4a)	(8a)
b_2: postverbal	(1b)	(5b)	(2b)	(6b)	(3b)	(7b)	(4b)	(8b)

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

The sample consists of 54 native speakers of the Spanish variety of Catalonia (see Pay-rató 1985; Sabater 1991; various papers in Sinner & Wesch 2008) as well as of Catalan, and does not include linguists or students of linguistics. The sample is balanced between women and men, and between age groups (age range: 17 to 48; mean age: 27 and a half). The speakers are fairly balanced bilinguals. They usually grew up speaking Spanish with one parent (or member of the family) and Catalan with the other. They also continue to use both languages on a regular basis in daily life. One should bear in mind that bilingualism is the default situation in Catalonia. Catalan speakers are basically never monolinguals. There are Spanish monolinguals in Catalonia, but they have a specific background insofar they are migrants from other regions or countries. The sample plan reflects this reality, controlling as far as possible the variable 'balance'.[3]

The experimental method is a further development of the paper and pencilbased gradient grammaticality judgment test in Adli (2005). The present version of the method is computerbased. It has been coded in the programming language *php* and runs in a normal browser window. The computerbased version is easier to handle than drawing lines with ruler and pencil. Subjects read the sentence and express their judgment on a graphic rating scale. Fine nuances can be expressed in a very intuitive way. The experiment took place in a quiet room in front of a notebook computer.
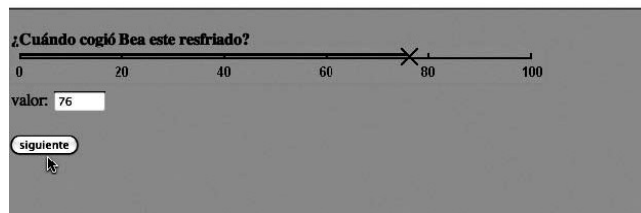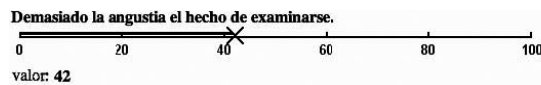

Figure 1: Gradient acceptability judgment test

---

[3]   3 out of 54 subjects had to be excluded from analysis. In two cases there were doubts to which extent their bilingualism could be called 'fairly balanced'. One person was excluded because she lacked sufficient collaboration during data collection.

Subjects draw a line on the computer screen by dragging a cross on a horizontal slider. The values run from 0 to 100 (the actual value is always displayed under the slider). The length of the line corresponds to the degree of acceptability, i.e. the higher the value the more acceptable the construction. At the beginning of the test, subjects judge a reference sentence, which serves as a scale anchor (i.e. they dispose, in addition to the two endpoints of the scale, of a concrete example to which they can always refer, when making up their mind).

The experiment began with a detailed instruction and training phase. Apart from learning to use the computer-based instrument, they are familiarized step by step with the concept of gradience and they learn not to take into consideration aspects that are irrelevant to syntax (e.g. opaque ideas on 'elegance', etc.). The entire procedure lasted in average 35 minutes (15 minutes for instruction and training, 20 minutes for the experimental phase).

4. Results

Figure 2 and Figure 3 show the arithmetic means of the acceptability values for the test sentences. The y-axis represents the degree of acceptability from 0 to 100, which corresponds to the length of the line traced in the judgment test in Figure 1. A first visual assessment of Figure 2 and Figure 3 reveals a striking similarity of the results between Spanish and Catalan, with some difference in the details that are going to be discussed.
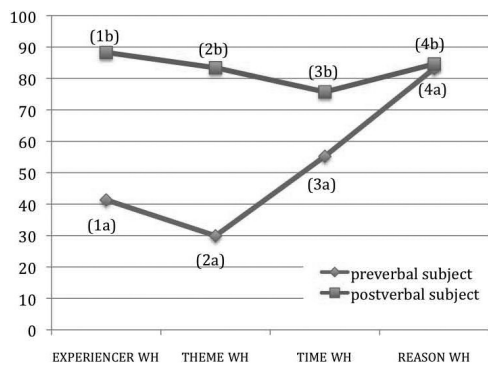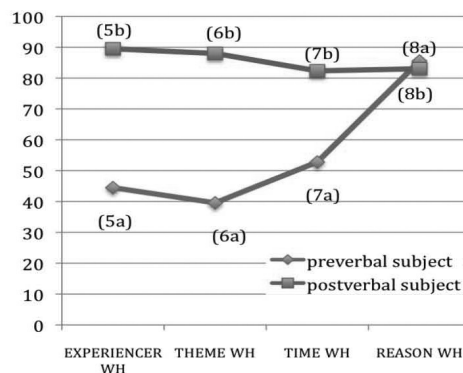

Figure 2: Gradient judgments for Spanish   Figure 3: Gradient judgments for Catalan

In the following, the results are statistically analyzed with a three-way analysis of variance (ANOVA) according to the design in Table 1. I will concentrate on the effects of variable B "subject position", which is the phenomenon under debate in this study. Main effect B is significant, showing an overall difference between constructions with preverbal and postverbal subject ($p < 0.000$). This effect is salient in both languages, i.e. it can be found independently in Spanish (simple main effect $B|c_1$: $p < 0.000$) and in Catalan (simple main effect $B|c_2$: $p < 0.000$).[4]

---

[4] (i) Preverbal vs. postverbal subject for all test sentences (main effect B): $F(1,50) = 382.1$, partial $\eta^2 = 0.884$, $p < 0.000$. (ii) Preverbal vs. postverbal subject in Spanish (simple main effect $B|c_1$): $F(1,50) = 271.6$, partial $\eta^2 = 0.845$, $p < 0.000$. (iii) Preverbal vs. postverbal subject in Catalan (simple main effect $B|c_2$): $F(1,50) = 315.8$, partial $\eta^2 = 0.863$, $p < 0.000$.

In a next step, the effect of subject position is analyzed separately for each *wh*-role; in other words, sentence pairs are compared. With regard to Spanish, we observe that all *wh*-roles show an effect of subject position, except for the sentence pair (4a) vs. (4b) with a REASON-*wh* ($p < 0.22$): (1a) vs. (1b) with an EXPERIENCER-*wh* reveals a significant difference ($p < 0.000$), as does (2a) vs. (2b) with a THEME-*wh* ($p < 0.000$), as well as (3a) vs. (3b) with a TIME-*wh* ($p < 0.000$).[5]

At first sight, the results for Catalan look nearly identical. All *wh*-roles show an effect of subject position, except for the sentence pair (8a) vs. (8b) with a REASON-*wh*: We find an effect for (5a) vs. (5b) with an EXPERIENCER-*wh* ($p < 0.000$), for (6a) vs. (6b) with a THEME-*wh* ($p < 0.000$), and for (7a) vs. (7b) with a TIME-*wh* ($p < 0.000$).[6] The difference between REASON-*wh* that shows no effect of subject, on the one hand, and EXPERIENCER-*wh*, THEME-*wh*, and TIME-*wh* that show an effect of subject, on the other, is statistically reflected in the interaction A x B between *wh*-role and subject position ($p < 0.000$).

With regard to Spanish and Catalan *wh*-questions with preverbal subject, the results show that the constructions (4a) and (8a) with a REASON-*wh* are fully unmarked, the constructions (1a), (2a), (5a) and (6a) with an EXPERIENCER-*wh* or a THEME-*wh* are highly marked, while the constructions (3a) and (7a) with a TIME-*wh* are also marked, but to a lesser degree. In terms of a ranking order, the results reflect the same hierarchy for Catalan and Spanish. EXPERIENCER-and THEME-*wh* are given the same rank, either because their results are fully identical (Spanish) or because the difference is very small (Catalan).

(17)   EXPERIENCER|THEME > TIME > REASON

However, the triple interaction effect $A \times B \times C$ ($p < 0.000$) between all three variables shows us that the patterns of Spanish and Catalan are, though strikingly similar, not identical. There are specific differences between the languages with respect to the relation between subject position and *wh*-role.[7] In order to evaluate these differences, we need to have a closer look at their **size**. This measure is expressed by the partial η^2 value, which expresses the amount of variance explained by an effect in the ANOVA model (COHEN 1973, 1988). Partial η^2 is a more precise indicator for the size of the difference than the difference between mean values (which would correspond to the vertical distance between two sentences in Figure 2 and Figure 3): Apart from the mean values, partial η^2 also takes into consideration information on the dispersion of the data, as well as on the sample size. It is broadly used in psychology (its systematic report is even required by the AMERICAN PSYCHOLOGICAL ASSOCIATION 2001: 25), but not yet established in linguistics (for one of the very few linguistic studies that use this measure, see KONDO-BROWN 2005). Partial η^2 values range from 0 to 1.

---

[5] The results of the simple main effect tests read as follows:
   (1a) vs. (1b) (simple main effect B|a_1c_1): $F(1,50) = 356.8$, partial $\eta^2 = 0.877$, $p < 0.000$.
   (2a) vs. (2b) (simple main effect B|a_2c_1): $F(1,50) = 366.0$, partial $\eta^2 = 0.880$, $p < 0.000$.
   (3a) vs. (3b) (simple main effect B|a_3c_1): $F(1,50) = 42.7$, partial $\eta^2 = 0.46$, $p < 0.000$.
   (4a) vs. (4b) (simple main effect B|a_4c_1): $F(1,50) = 0.6$, partial $\eta^2 = 0.011$, $p < 0.455$.
[6] (5a) vs. (5b) (simple main effect B|a_1c_2): $F(1,50) = 334.3$, partial $\eta^2 = 0.87$, $p < 0.000$.
   (6a) vs. (6b) (simple main effect B|a_2c_2): $F(1,50) = 243.8$, partial $\eta^2 = 0.83$, $p < 0.000$.
   (7a) vs. (7b) (simple main effect B|a_3c_2): $F(1,50) = 94.4$, partial $\eta^2 = 0.654$, $p < 0.000$.
   (8a) vs. (8b) (simple main effect B|a_4c_2): $F(1,50) = 1.5$, partial $\eta^2 = 0.03$, $p < 0.22$.
[7] Interaction $A \times B$: $F(3,150) = 162.7$, partial $\eta^2 = 0.765$, $p < 0.000$
   Interaction $A \times B \times C$: $F(3,150) = 6.5$, partial $\eta^2 = 0.116$, $p < 0.000$

Figure 4 compares the effect sizes for Spanish and Catalan. Each data point represents the difference between a sentence pair in terms of partial η^2. For example, the lower data point at the category TIME-*wh* shows the effect size of the difference between (3a) and (3b), and the upper data point the effect size of the difference between (7a) and (7b).

The effect sizes confirm that the ranking for both Spanish and Catalan can be described by a hierarchy as in (17), in particular if the order is meant to reflect a hierarchy of markedness as suggested by GUTIÉRREZ-BRAVO (2006). In both languages the largest effects occur with EXPERIENCER-and THEME-*wh*, an intermediate effect size is observed with TIME-*wh*, and no difference is identified with REASON-*wh*.

However, the differences between Spanish and Catalan reside mainly in one detail that a mere ranking order as in (17) cannot reflect: Constructions with a TIME-*wh* and preverbal subject are clearly more marked in Catalan than in Spanish. The effect size is much more pronounced in Catalan.[8]

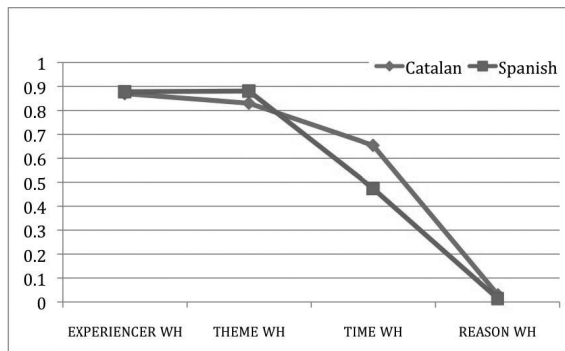Figure 4: Effect size (partial η^2) of pre- vs. postverbal subjects

5. Discussion

How can we explain the nuanced differences in markedness represented in Figure 4? We have seen that a constraint-based framework can cover degrees in acceptability, if the ranking of the candidates is correlated with the degree of markedness. However, is there independent evidence for the hierarchy in (17) and the pattern in Figure 4?

I suggest an explanation in terms of two principles, and suggest that their effects can cumulate: (i) A general markedness effect of preverbal subjects in *wh*-questions, which does not include constructions with REASON-*wh*. (ii) A difference between argument-and adjunct-*wh*-questions, in order to account for the high markedness of constructions with EXPERIENCER- or THEME-*wh*.

With regard to (i) the following question has to be asked: Why are *wh*-questions with a preverbal subject marked, and what can explain the exceptionality of REASON-*wh*? An analysis of the information structure will prove helpful for this issue. FRASCARELLI (2007)

---

[8] Furthermore, the difference between pre- and postverbal subjects for EXPERIENCER-*wh* equals the difference for THEME-*wh* in Spanish. This contrasts with Catalan, where the difference for THEME-*wh* is slightly less pronounced. However, since the effect sizes between EXPERIENCER- and THEME-*wh* in Catalan is fairly small, it does not seem to justify a difference in ranking in (17).

proposes a distinction between two types of topics (a similar distinction had already been used by Svenonius 2002; Karimi 2005). She calls the first one "aboutness shift topic", which newly proposes or reintroduces a topic in the discourse, and the second one "familiar topic", which refers to background information thereby ensuring topic continuity (a third type, contrastive topics are not relevant here, but see Büring 1999; Frascarelli & Hinter-hölzl 2007; Krifka 2007). Frascarelli (2007) shows that a distinction of these two topic types can account for the distribution of null subjects in Italian: Strong pronouns are about-ness shift topics, while null subjects are familiar topics referring to the closest aboutness shift topic for interpretation. I assume the same mechanism suggested for Italian to apply to Spanish and Catalan, which are also Romance null subject languages (more precisely, all three languages are consistent null subject languages in terms of the tripartite system of Holmberg et al. 2009). The important point is that all subjects in (1a) to (8b) are full or proper nouns – and not null pronouns. Their default topic interpretation is aboutness shift. Although it is not impossible that a familiar topic is realized by a full or proper noun in cer-tain contexts, the default realization for familiar topics is a null subject. Frascarelli (2007: 701) proposes a cartographic approach, in which the topic types correspond to different functional projections of the left periphery.

(18) $[_{\text{ForceP}} [_{\text{ShiftP}} [_{\text{GP}} [_{\text{ContrP}} [_{\text{FocP}} [_{\text{FamP}} [_{\text{FinP}}$

According to (18), a *wh*-element in FocP is located below an aboutness shift topic in ShiftP, but above a familiar topic in FamP. The Italian example (19) from Rizzi (2006), which is supposed to be an answer to the question 'what happened?', is reanalyzed by Fras-carelli (2007: 720) according to the structure below.

(19) $[_{\text{ShiftP}}$ un camion$_k$ $[_{\text{AgrSP}}$ ha $[_{vP} pro_k [_{\text{VP}}$ tamponato l' autobus per Roma $]]]]$
         a truck   have$_{3SG}$ bumped   the bus   for Rome
         'A truck has bumped into the bus for Rome'

She argues that the preverbal subject creates a topic/comment structure (i.e. a thetic reading is excluded), and that the subject *un camion* ('a truck') is typically interpreted as an aboutness shift topic.[9] Note that Frascarelli (2007) works with a nonmovement account and assumes the topic in the Spec of ShiftP to be coreferent with a null pronoun in subject position. With regard to a new topic appearing in postverbal position, she assumes that it is coreferent with a silent copy in the Spec of ShiftP.

Building on this account I assume that subjects in Spanish and Catalan that are moved out of *v*P into a preverbal position target the appropriate topic projection in the left periphery. In case the subject remains in *v*P, as is assumed for postverbal subjects, the cor-rect interpretation is ensured by merging a silent copy in the topic projection in the left periphery.

The problem is that the subjects in (1a) to (8b), which are full or proper nouns, would have to appear above the *wh*-element, in order to be interpreted as an aboutness shift topic. Independent evidence comes from results of spontaneous speech data, taken from the Spanish treebank of the sgs database, as reported in Adli (in press b). The results

---

9   Frascarelli (2007) shows that preverbal subjects in Italian can sit in A-positions – the same has also been assumed for Spanish (Suñer 2003) – and thus serve as antecedent to referential null subjects. This view is at odds with the 'pronominal Agreement' analysis in Alexiadou & Anagnostopoulou (1998), and in Barbosa (2000).

of that study show that among 384 *wh*-questions with an overt (non-*wh*) subject, 61.7 %
are realized with a postverbal subject (WH V S), only 0.5 % with a preverbal subject
following the *wh*-element (WH S V), but 37.8 % with the subject **preceding** the *wh*-element
(S WH V). In other words, an overt subject, which is not postverbal, appears nearly ex-
clusively above the *wh*-element, as in (20) (see also MEYER 1972; FONTANA 1994; ZUBI-
ZARRETA 1998)

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

Why are constructions with REASON-*wh* and preverbal subject not marked? We can ex-
plain this effect, if we do not assume REASON-*wh*-elements in FocP (where I assume all
other left-peripherical *wh*-elements), but in a position above ShiftP. STEPANOV & TSAI
(2008), adopting and slightly modifying RIZZI's (1990, 1997, 1999) original proposal, place
REASON-why-elements (which cross-linguistically show a special behavior among *wh*-
elements) in IntP. IntP is supposed to be a very high projection in CP located below
ForceP and above FocP. Integrating IntP in (18), we obtain (21), which now can explain
why a REASON-*wh*-element can be followed by a preverbal, shift-topical subject.

We can assume that the test sentences with preverbal subject, except for those with
REASON-*wh*, violate a principle or constraint that can be formulated as in (22). This principle
excludes focal elements followed by an aboutness shift topic in the C-system in languages
with a left periphery as in (21).

(21)   [ForceP [IntP [ShiftP [GP [ContrP [FocP [FamP [FinP

(22)   *[CP FOCUS SHIFTTOP]

The following data taken from STEPANOV & TSAI (2008) show that REASON-why-elements
indeed have to be placed very high within the left periphery: They have scope over the
focus operator instantiated by clefting.

(23)   Why was it Adam who ate the apple?

(24)   *When was it Adam who ate the apple?

Furthermore, the following Italian examples from RIZZI (1999) illustrate that *why*-ele-
ments, unlike other *wh*-elements, can co-occur with contrastive focus:

(25)   *Che   cosa   *[a Gianni]*$_{c-F}$ *hanno*   *detto*   *(non*   *a*   *Piero)*?
        what   thing   to Gianni   have$_{3SG}$ said   (not   to   Piero)
        'What have they said to Gianni, (not to Piero)?'

(26)   *Perché*   *[questo]*$_{c-F}$ *avremmo*   *dovuto*   *dirgli,*   *non*   *qualcos'*   *altro*?
        why   this   have$_{1PL}$ should   said-CL$_{3SG}$ not   something   else
        'Why should we have told him this, not something else?'

Thus, the C-structure in (21) explains, why *wh*-elements, except for REASON-*wh*, cannot
co-occur with preverbal subjects. However, it does not explain, why the markedness of
EXPERIENCER-*wh* and THEME-*wh* is more pronounced than the markedness of TIME-*wh*. What
distinguishes the one from the other is their ±argumental status. The results suggest that
wh-elements that are +argumental and receive a *θ*-role (SUÑER 1994), have a higher degree

of markedness than other *wh*-elements. Following SUÑER (1994), this effect is due to a mini-mality effect on the [+argumental] feature: The licensing relation between the inflected verb and the fronted *wh*-element cannot be established due to the intervening subject which also bears [+argumental]. Thus, the test sentences with EXPERIENCER-*wh* and THEME-*wh* violate a principle or constraint that we can state as in (27).[10]

(27) $*[WH+_{arg} S+_{arg} V+_{arg}]$

The different nuances in markedness can be accounted for by assuming that the viola-tion of a rule has a specific, quantifiable **cost** and that the costs of violations of several rules cumulate (see USZKOREIT 1987; KELLER 2000; ADLI in press a, on the idea of the cumulation of numerical violation costs). The higher the cost, the higher the degree of markedness.

In the following, I will illustrate a simplified model of cumulative constraint violation costs. Violation costs can be represented as the difference between the marked preverbal and the unmarked postverbal variant, i.e. the higher the violation cost, the higher the partial η^2 value as shown in Figure 4. In this model numerical violation costs are expressed in terms of partial η^2.[11] The idea is that (i) (1a), (2a), (3a), (5a), (6a), and (7a) violate prin-ciple (22), and (ii) (1a), (2a), (5a), and (6a) violate, **in addition**, principle (27). (4a) and (8a) violate neither principle (22) nor principle (27). The partial η^2 values (precisely indicated in footnotes 5 and 6 and shown in Figure 4) reveal that violation of (22) has a cost of 0.460 in Spanish and of 0.654 in Catalan. (27) has a violation cost of 0.4185 in Spanish and of 0.196 in Catalan.[12]

---

[10] A general distinction between *wh*-arguments and *wh*-adjuncts cannot explain the data presented by INCLÁN-NICHOL (1997). She admits *wh*-arguments with preverbal subjects, as long as they are specific (see the D-linked and partitive examples in (i) and (ii)), or as long as they have an assertion or nega-tive polarity marker. However, it is not clear to me, whether these constructions should be considered as fully unmarked (on a par with constructions with REASON-*wh*), or whether they are just less marked than constructions with other *wh*-arguments. In order to propose a satisfying answer to this issue, the constructions discussed by INCLÁN-NICHOL (1997) would need to be included in a gradient accept-ability judgment test in future research. The model could then be eventually refined.

(i) *¿A   cuál   estudiante   Lupe   llamó   el   fin   de semana*?
     To   which   student   Lupe   call$_{3SG.PRET}$ the   end   of week
     'Which student did Lupe call this weekend?'

(ii) *¿A   quién   de los presentes   Lupe   saludó   en   la fiesta*?
      to   who   of the present   Lupe   greet$_{3SG.PRET}$ in   the party
      'Who of the present people did Lupe greet at the party?'

[11] See ADLI (2009) for a more complex quantitative model of constraint cumulativity, in which violation costs are expressed by the amount by which the acceptability is lowered from a given optimum.

[12] The calculation for the Spanish data reads as follows: Step 1: The partial η^2 of (3a) vs. (3b) is equal to the cost of (22):

$cost_{Sp}(22) = 0.460$

Step 2: The mean value of the partial η^2 of (1a) vs. (1b) and of the partial η^2 of (2a) vs. (2b) is equal to the cost of (22) plus the cost of (27):

$(0.877 + 0.880)/2 = cost_{Sp}(22) + cost_{Sp}(27)$

$\leftrightarrow cost_{Sp}(27) = (0.877 + 0.880)/2 - cost_{Sp}(22) = 0.8785 - 0.460 = 0.4185$

The calculation for the Catalan data reads as follows: Step 1: The partial η^2 of (7a) vs. (7b) is equal to the cost of (22):

$cost_{Cat}(22) = 0.654$

Step 2: The mean value of the partial η^2 of (5a) vs. (5b) and of the partial η^2 of (6a) vs. (6b) is equal to the cost of (22) plus the cost of (27):

$(0.870 + 0.830)/2 = cost_{Cat}(22) + cost_{Cat}(27)$

$\leftrightarrow cost_{Cat}(27) = (0.870 + 0.830)/2 - cost_{Cat}(22) = 0.850 - 0.654 = 0.196$

6. Conclusion

In sum, the striking similarity of the judgment patterns of Spanish variety of Catalonia and Catalan is reflected in the fact that the same grammatical principles are at work. The difference between the two languages lies in the fact that the principle, which translates the argument/adjunct asymmetry in *wh*-questions, i.e. (27), is less prominent in Catalan than in Spanish. The degrees of markedness of the different *wh*-questions with preverbal subjects (excluding REASON-*wh*) is more similar in Catalan than in Spanish, because the Catalan pattern goes mainly back to one general principle, i.e. (22), that prohibits preverbal subjects in *wh*-questions, irrespective of their argumental status. (22) is more prominent in Catalan than in the Spanish variety of Catalonia.

A model, which translates degrees of markedness in violation costs could also prove useful in modelling language change. It can account for subtle and continuous shifts of markedness, which presumably occur before a categorical change of the system takes place, i.e. before the set of relevant principles/constraints itself is modified.

Abbreviations

CL	clitic	PL	plural
DAT	dative	PRET	preterit
FUT	future	SG	singular
PERIPHR.PRET			

References

ADLI, ARIA (2005): Gradedness and consistency in grammaticality judgments, in: KEPSER, STEPHAN & REIS, MARGA (eds.), *Linguistic evidence: empirical, theoretical, and computational perspectives.* Berlin: Mouton de Gruyter, 7–25.

ADLI, ARIA (in press a): Constraint cumulativity and gradience: wh-scrambling in Persian, in: *Lingua*.

ADLI, ARIA (in press b): On the relation between acceptability and frequency, in: RINKE, ESTHER & KUPISCH, TANJA (eds.), *The development of grammar: language acquisition and diachronic change – Volume in honour of Jürgen M. Meisel.* Amsterdam, New York: Benjamins.

ADLI, ARIA (2009): A heuristic mathematical approach for modeling constraint cumulativity. *Submitted*.

ALEXIADOU, ARTEMIS & ANAGNOSTOPOULOU, ELENA (1998): Parametrizing AGR: word order, V-movement and EPP checking, in: *Natural Language and Linguistic Theory* 16, 491–539.

AMERICAN PSYCHOLOGICAL ASSOCIATION (2001): *Publication manual of the American Psychological Association (5th ed.)*. Washington: APA.

BAKOVIC, ERIC (1998): Optimality and inversion in Spanish, in: BARBOSA, PILAR; FOX, DANNY; HAGSTROM, PAUL; MCGINNIS, MARTHA & PESETSKY, DAVID (eds.), *Is the best good enough?* Cambridge: MIT Press, 35–58.

BARBOSA, PILAR (2000): Clitics: a window into the null subject property, in: COSTA, JOÃO (ed.), *Essays in portuguese comparative syntax.* Oxford: Oxford University Press, 31–94.

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

BELLETTI, ADRIANA & RIZZI, LUIGI (1988): Psych-verbs and θ-theory, in: *Natural Language and Linguistic Theory* 6(3), 291–352.

BONET, EULÀLIA (1990): Subjects in Catalan, in: *MIT Working Papers in Linguistics* 13, 1–26.

BÜRING, DANIEL (1999): Focus, in: BOSCH, PETER & VAN DER SANDT, ROB (eds.), *Linguistic cognitive and computational perspectives.* Cambridge: Cambridge University Press, 142–165.

CHOMSKY, NOAM (1971): Deep structure, surface structure and semantic interpretation, in: STEINBERG, DANNY D. & JAKOBOVITS, LEON A. (eds.), *Semantics: an interdisciplinary reader in philosophy, linguistics and psychology.* Cambridge: Cambridge University Press.

CINQUE, GUGLIELMO (1993): A null theory of phrase and compound stress, in: *Linguistic Inquiry* 24, 239–298.

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

COHEN, JACOB (1988): *Statistical power analysis for the behavioral sciences (2nd edition).* Hillsdale: Erlbaum.

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

FONTANA, JOSEP (1994): A residual A-Bar position in Spanish, in: DUNCAN, ERIN; FARKAS, DONKA & SPAELTI, PHILIP (eds.), *Proceedings of WCCFL 12.* Stanford: CSLI, 233–250.

FRASCARELLI, MARA (2007): Subjects, topics and the interpretation of referential *pro*: an interface approach to the linking of (null) pronouns, in: *Natural Language and Linguistic Theory* 25(4), 691–734.

FRASCARELLI, MARA & HINTERHÖLZL, ROLAND (2007): Types of topics in German and Italian, in: WINKLER, SUSANNE & SCHWABE, KERSTIN (eds.), *On information structure, meaning and form.* Amsterdam, Philadelphia: Benjamins, 87–116.

GARCÍA-LECUMBERRI, MARÍA LUISA (1995): *Intonational signaling of information structure in English and Spanish: a comparative study.* Bilbao: Servicio Editorial de la Universidad del País Vasco.

GUTIÉRREZ-BRAVO, RODRIGO (2002): Focus, word order variation and intonation in Spanish and English, in: WILTSHIRE, CAROLINE R. & CAMPS, JOAQUIM (eds.), *Romance phonology and variation: Selected papers from the 30th Linguistic Symposium on Romance Languages, Gainesville, Florida, February 2000.* Amsterdam: Benjamins, 39–53.

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.

HOLMBERG, ANDERS; NAYUDU, AARTI & SHEEHAN, MICHELLE (2009): Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi, in: *Studia Linguistica (Special Issue: Partial Pro-drop)* 63(1), 59–97.

HUALDE, JOSÉ IGNACIO (2002): Intonation in Spanish and the other Ibero-Romance languages: overview and status quaestionis, in: WILTSHIRE, CAROLINE R. & CAMPS, JOAQUIM (eds.), *Romance phonology and variation.* Amsterdam: Benjamins, 101–115.

HUALDE, JOSÉ IGNACIO (2003): El modelo métrico y autosegmental, in: PRIETO, PILAR (ed.), *Teorías de la entonación.* Barcelona: Ariel, 155–184.

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.

KARIMI, SIMIN (2005): *A minimalist approach to scrambling: evidence from Persian.* Berlin: Mouton de Gruyter.

KELLER, FRANK (2000): Gradience in grammar. Experimental and computational aspects of degrees of grammaticality. PhD dissertation, University of Edinburgh.

KONDO-BROWN, KIMI (2005): Differences in language skills: heritage language learnersubgroups and foreign language learners, in: *The Modern Language Journal* 89(4), 563–581.

KRIFKA, MANFRED (2007): Basic notions of information structure, in: FÉRY, CAROLINE; FANSELO, GISBERT & KRIFKA, MANFRED (eds.), *Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS) 6.* Potsdam: Universitätsverlag Potsdam, 13–56.

LARSON, RICHARD K. (1988): On the double object construction, in: *Linguistic Inquiry* 19, 335–391.

LÓPEZ, LUÍS (2009): *A derivational syntax for information structure.* Oxford: Oxford University Press.

MARGARET, SPEAS (1990): *Phrase structure in natural language.* Dordrecht: Kluwer.

MASULLO, PASCUAL J. (1993): Two types of quirky subjects: Spanish versus Icelandic, in: *Proceedings of NELS 23.* Amherst: University of Massachusetts, 303–317.

MEYER, PAULA (1972): Some observations on constituent order in Spanish, in: CASAGRANDE, JEAN & SACIUK, BOHDAN (eds.), *Generative studies in Romance languages.* Rowley: Newbury House Publishers, 184–195.

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

ORDÓÑEZ, FRANCISCO (2000): *The clausal structure of Spanish: a comparative study.* New York: Garland.

PAYRATÓ, LLUÍS (1985): *La interferència lingüística. Comentaris i exemples català – castellà.* Barcelona: Montserrat (Curial edicions catalanes).

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

Rizzi, Luigi (1990): *Relativized minimality*. Cambridge: MIT Press.

Rizzi, Luigi (1997): The fine structure of the left periphery, in: Haegeman, Liliane (ed.), *Elements of grammar.* Dordrecht: Kluwer, 281–337.

Rizzi, Luigi (1999): On the position "Int(errogative)" in the left periphery of the clause. Ms. Siena: University of Siena.

Rizzi, Luigi (2006): On the form of chains: Criterial positions and ECP effects, in: Cheng, Lisa Lai-Shen & Corver, Norbert (eds.), *Wh movement: moving on.* Cambridge: MIT Press, 97–133.

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

Samek-Lodovici, Vieri (2001): Crosslinguistic typologies in optimality theory, in: Legendre, Géraldine; Grimshaw, Jane & Vikner, Sten (eds.), *Optimality-theoretic syntax.* Cambridge, Massachusetts: MIT Press, 315–353.

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

Stepanov, Arthur & Tsai, Wei-Tien Dylan (2008): Cartography and licensing of wh-adjuncts: a cross-linguistic perspective, in: *Natural Language & Linguistic Theory* 26(3), 589–638.

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

Suñer, Margarita (2003): The lexical preverbal subject in a romance null subject language. Where are thou?, in: Núñez-Cedeño, Rafael; López, Luis & Cameron, Richard (eds.), *A romance perspective on language knowledge and use.* Amsterdam, Philadelphia: John Benjamins, 341–357.

Svenonius, Peter (2002): Subject positions and the placement of adverbials, in: Svenonius, Peter (ed.), *Subjects, expletives, and the EPP.* Oxford, New York: Oxford University Press, 201–242.

Toledo, Guillermo Andrés (1989): Señales prosódicas del foco, in: *Revista Argentina de Lingüística* 5, 205–230.

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

Uszkoreit, Hans (1987): *Word order and constituent structure in German.* Stanford, CA: CSLI Publications.

Zagona, Karen (2002): *The syntax of Spanish.* Cambridge, Oxford: Cambridge University Press.

Zubizarreta, Maria Luisa (1998): *Prosody, focus and word order.* Cambridge: MIT Press.

Zubizarreta, Maria Luisa (1999): Tema y foco, in: Bosque, Iganacio & Demonte, Violeta (eds.), *Gramática descriptiva de la lengua española.* Madrid: Espasa-Calpe, 4216–4244.

Zubizarreta, Maria Luisa (2001): The constraint on preverbal subjects in Romance interrogatives: a Minimality effect, in: Hulk, Aafke & Pollock, Jean-Yves (eds.), *On Romance inversion.* Oxford: Oxford University Press.

Zubizarreta, Maria Luisa (2007): Aspects of the left periphery in Spanish. Paper presented at 'Linguistic Symposium of Romance Languages'.

Aria Adli
Universität Hamburg
Institut für Romanistik
von-Melle-Park 6
D - 20146 Hamburg
GERMANY
aria.adli@uni-hamburg.de