Gradedness and Consistency in Grammaticality Judgments

Aria Adli

SFB 441 – University of Tübingen, Germany

`aria.adli@uni-tuebingen.de`

The methodological issue of the unreliability of certain introspective data circulating in syntactic literature has already been mentioned by several authors (e.g. Schütze, 1996; Adli, 2003). One particularly problematic phenomenon is that questionable judgments are sometimes quoted in theoretical studies without prior critical empirical verification contributing thus to the formation of "myths" in the literature. I shall argue in this paper that one case in point is the widely-quoted *que* → *qui* 'rule' in French. It will be demonstrated that this rule, which has been introduced into the literature solely on the basis of un-controlled introspective data, is not confirmed by experimental studies in which a controlled process of data-collection is applied to a whole sample of test persons and which essentially make use of a *graded* concept of grammaticality. Furthermore, it will be shown that the use of quantitative data on graded grammaticality does not only allow the interpretation of the (mean) grammaticality value, but also the consideration of another measure, namely the (*intra*-individual) *judgment consistency* observed for a given construction. This newly introduced measure has the advantage to complement the information about the exact grammaticality value with the information on the difficulty to give stable judgments allowing a more complete evaluation of the grammatical quality of a structure.

1   The importance of graded grammaticality judgments: a case study on *que-qui* in French

The *que* → *qui* rule essentially states that an ECP violation can be avoided in French if *qui* is used instead of the usual complementizer *que* in sentences where a *wh*-phrase has been extracted from subject position (see Perlmutter, 1971; Kayne, 1977). This rule rests on the 'premise' that there is a clear difference in grammaticality between (2a) and (2b) (all four sentences are taken from Hulk & Pollock, 2001).

(1a)   Quel   livre crois-tu **que**   les filles vont acheter.

which book think-you COMP$_{que}$ the girls will   buy

(1b) * Quel   livre crois-tu **qui**   les filles vont acheter.

which book think-you COMP$_{qui}$ the girls will   buy

(2a) * Quelles filles crois-tu **que**   vont acheter ce   livre-là.

which   girls think-you COMP$_{que}$ will   buy   that book-there

(2b)   Quelles filles crois-tu **qui**   vont acheter ce   livre-là.

which   girls think-you COMP$_{qui}$ will   buy   that book-there

The *que → qui* rule has been an often used argument in syntactic theorizing. Chomsky (1977) compares it with free deletion in COMP in English, Rizzi (1990; 1997) supports his assumptions concerning the agreement process in the COMP system with this rule, to give only two prominent examples.

The aim of this paper was to test this assumption in an experimentally controlled process of data-collection using a *graded* concept of grammaticality. In order to measure graded grammaticality judgments an instrument based on the principle of *graphic rating* (cf. Guilford, 1954: 270) has been developed (cf. Adli, 2003 for details). Part of the design is an extensive instruction and training phase. Given that reliability can generally be improved by the use of several items, each syntactic structure was presented in 4 lexical variants. Indeed, the reliability of the measurements turned out to be sufficiently high (Cronbach's α = 0,85). 78 French native speakers participated in the experiment. Validity was ensured by means of a special index (called *violation of trivial judgments*) reflecting the capability of the subject to give graded grammaticality judgments. By means of this criterion those persons who were apparently not able to perform this task could be identified and excluded; the data of 65 persons could be utilized for the statistical analyses.

Given that the measure of graded grammaticality does not reflect the dichotomous distinction between well-formed and ill-formed sentences and given that such an information is still - for theory-internal reasons - important, felicitous as well as infelicitous constructions were included in the test design in order to make available comparative scale points for the interpretation process. Therefore, the experiment did not only cover subject-object-interrogatives with long extraction over *que* and/or *qui*. The clearly felicitous constructions (3a) and (3b) with a PP-parenthetical "d'après vous" and the sentences (4a) and (4b) with the expression "croyez-vous" were also included – however, their syntax will not be discussed here (but see Adli, 2003).

(3a)   Quel appache, **d'après vous**, méconnaît les obstacles   de l'hiver?
which Appache, according you, ignores   the difficulties of the winter

(3b)   Quel animal,   **d'après vous**, rôtissent les esquimaux de l'igloo?
which animal,   according you, grill   the Eskimos   of the Igloo

(4a) (?) Quel architecte, **croyez-vous**, conçoit les demeures du président?
which architect think you designs the residences of the president

(4b) (?) Quel argent, **croyez-vous**, investissent les organisateurs du bal?
which money think you invest the organisers of the ball

(5a) ?? Quel ingénieur, **pensez-vous, qui** conçoit la fusée de l'Aérospatiale?
which engineer think you qui$_{COMP}$ designs the rocket of Aérospatiale

(5b) * Quel idiot, **pensez-vous, que** perd les clefs de la maison?
which idiot think you que$_{COMP}$ looses the keys of the house

(5c) ?? Quel appel, **pensez-vous, que** reçoivent les policiers du quartier?
which call think you que$_{COMP}$ receive the police officers of the district

The data was analysed with a repeated measures ANOVA. I did not only take into consideration information about the significance level, but also about the effect size of the differences (in terms of partial η^2, cf. Cohen, 1973). In order to take into account the whole details of the results a complete set of orthogonal simple effects (cf. Bortz, 1999: 254) was tested as regards the subject interrogatives, contrasting (i) (3a) vs. (4a), (ii) (5a) vs. (5b), as well as (iii) (3a) and (4a) vs. (5a) vs. (5b).

The results show a partial η^2 of 0.183 ($p<0.000$) for contrast (i), a value of 0.149 ($p<0.001$) for contrast (ii), but an amount as high as 0.875 ($p<0.000$) for contrast (iii). Though there is a significant difference between the *qui*-form (5a) and the *que*-form (5b), it is a matter of fine-grained differences *within* the range of ungrammatical constructions.

The effect sizes show that the different subject-initial constructions divide into two clearly separated groups with an eye-catching decrease in grammaticality between them. The form with *qui* thus cannot be considered as the licensed counterpart of the form with *que*. The *que* → *qui* rule emerges as a myth, and it must consequently be eliminated from the discussion.
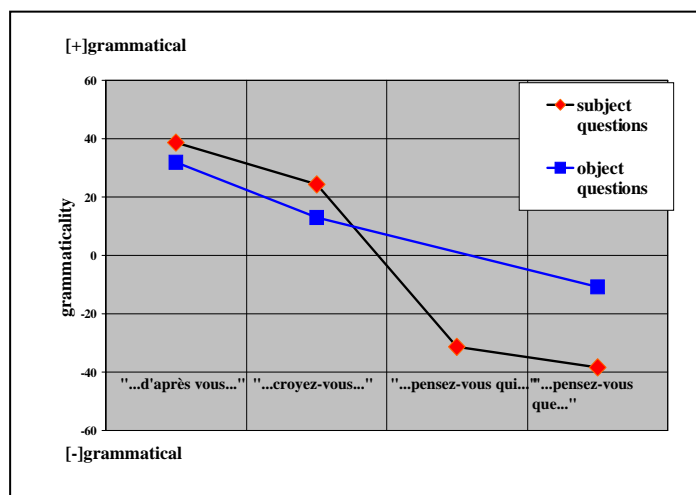
As to the question why the ungrammaticality of the *qui*-form is less sharp, I suggest rather psycholinguistic factors to be responsible: The use of *qui* instead of *que* evokes the structure of subject relative clauses (i.e. the nominative of *qui* has a sort of resumptive character) which alleviates the repair mechanisms.

2 Graded grammaticality and the measure of judgment consistency

It is not surprising that the measure of judgment consistency has been so far ignored in syntactic research: Its calculation generally requires a metrical (i.e. graded) grammaticality scale. This measure indicates the intra-individual degree of agreement between the judgments of the 4 lexical variants for each construction. To this end, I conducted reliability analyses using the average-measure intraclass correlation coefficient (ICC) of the absolute agreement type (cf. McGraw & Wong, 1996). The results given in the figure below show that (i) consistency of grammaticality judgments is not a stable factor but depends on the respective construction, (ii) in terms of our example both subject extractions, with *qui* and with *que*, have comparable consistency values, and (iii) it is more difficult to give consistent judgments to object- than to subject-interrogatives.
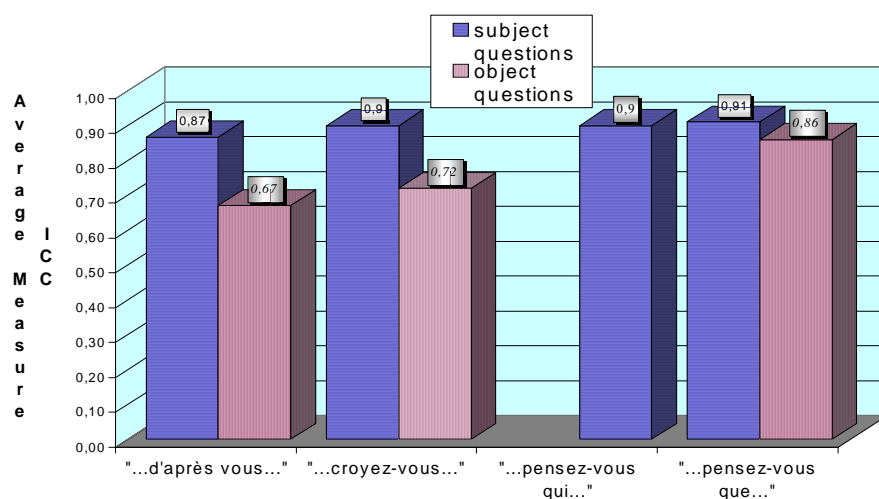
Consistency in the judgment of object interrogatives improves with increasing suboptimality as reveals a comparison of both figures (note that in French long object extraction as in (5c) is fairly marginal): Consistency is much higher for (5c) than for (3b) or (4b), i.e. there is an *interaction* between the degree of suboptimality and the sentence initial element. Hence, the analysis of judgment consistency provides another piece of empirical evidence in terms of the discussion about the syntax and the processing of subject- vs. object- initial interrogatives in French.[1]

The controlled measurement of a graded concept of grammaticality does, on the one hand, allow to obtain a detailed picture of grammaticality contrasts. Syntactic discus-

---

[1] The judgment consistencies are in line with the judgment values themselves which also reveal a difference between subject- and object-questions ((3a) vs. (3b): $p < 0.034$; (4a) vs. (4b): $p < 0.000$; (5b) vs. (5c): $p < 0.000$). Furthermore, I assume a correlation between the greater difficulty to give consistent judgments to object-initial questions and the higher cognitive cost, reported for many languages, to parse sentences with fronted objects.

sions can be placed on a more solid empirical base reducing the risk of myth production. On the other hand, the same measurements can be analysed from another point of view, namely with respect to judgment consistency which is a new and complementary source of information, worth to be hence considered in grammar research.

References

Adli, Aria (2003). *Grammatische Variation und Sozialstruktur: eine empirische Untersuchung anhand von Interrogativstrukturen des Französischen*. PhD dissertation, University of Tübingen.

Bortz, Jürgen (1999) *Statistik für Sozialwissenschaftler*. Berlin, New York: Springer.

Chomsky, Noam (1977). On wh-movement. In: Culicover, Peter, Thomas Wasow & Adrian Akmajian (eds.). *Formal syntax*, pp. 71-132. New York: Academic Press.

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

Guilford, Joy P. (1954). *Psychometric methods*. New York: MacGraw-Hill.

Hulk, Aafke & Jean-Yves Pollock (2001). Subject positions in Romance and the theory of Universal Grammar. In: Hulk, Aafke & Jean-Yves Pollock (eds.). *Inversion in Romance*, pp. 3-20. Oxford: Oxford University Press.

Kayne, Richard S. (1977). French relative que. In: Luján, Marta & Fritz Hensey (eds.). *Current Studies in Romance Linguistics*. Washington D.C.: Georgetown University Press.

McGraw, Kenneth O. & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1): 30-46.

Perlmutter, David (1971). *Deep and Surface Structure Constraints in Syntax*. New York: Holt, Rinehart and Winston.

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

Rizzi, Luigi (1997). The fine structure of the left periphery. In: Haegeman, Liliane (ed.). *Elements of Grammar*, pp. 281-337. Dordrecht: Kluwer.

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