

Experimental Evidence for the Syntax of Phrasal Comparatives in Polish

Roumyana Pancheva and Barbara Tomaszewicz*

1 An Island Effect in Phrasal Comparatives

Our starting point is an empirical observation and its analysis, found in Pancheva (2009):

- (1) a. In *wh*-fronting languages, if the *more*-phrase of a comparative is an external argument, a phrasal comparative is degraded.
- b. The gradient unacceptability of phrasal comparatives is due to a subject-island violation, conditioned by constraints on *wh*-movement and the small-clause syntax of the complement of *than*.

We first review the data behind the generalization and the formal model that has been proposed as its explanation. In section 2 we discuss the gradient nature of the effect and the resulting need for eliciting acceptability judgments in controlled experimental conditions allowing statistical analysis. In section 3 and 4 we discuss two acceptability-rating studies on Polish testing the validity of the empirical observation in (1a) and the theoretical explanation in (1b). The results support (1). We note the implications of our findings for theories of subject islands and conclude with a summary.

1.1 The Empirical Generalization

We illustrate (1a) with examples from Polish. The counterpart of *than* in phrasal comparatives is the preposition *od* ‘from’. The DP following *od* (‘the remnant’) is case-marked genitive by *od* (see (2a) and (3a)). In clausal comparatives the counterpart of *than* is *niż*. It does not license case on DP remnants; thus we do not analyze *niż* as a preposition. Its sister is clearly a clause, which often is only partially elided. In (2b) and (3b) the *niż*-clause has been maximally elided to include only a single remnant. In (2b) the remnant *Anna* is marked nominative, as it is the subject of the elided finite *niż*-clause. In (3b), the remnant *Słowacja* is case-marked accusative, as it is the direct object of the elided verb *visit* in the *niż*-clause. In general, DP-remnants in clausal comparatives have case features determined by their grammatical role in the *niż*-clause. See Juzwa (2006) on ellipsis in *niż*-comparatives, and Pancheva (2009) for further discussion of comparatives in Polish.

- (2) a. Marek zwiedził *więcej* *miejsc* *od* Anny.
Marek visited more places from Anna_{GEN}
‘Marek visited more places than Anna.’
- b. Marek zwiedził *więcej* *miejsc* *niż* Anna.
Marek visited more places than Anna_{NOM}
‘Marek visited more places than Anna did.’
- (3) a. ^{??/*}*Więcej uczniów* zwiedziło Czechy *od* Słowacji.
more students visited Czech R. from Slovakia_{GEN}
‘More students visited the Czech Republic than Slovakia.’
- b. *Więcej uczniów* zwiedziło Czechy *niż* Słowację.
more students visited Czech R. than Slovakia_{ACC}
‘More students visited the Czech Republic than visited Slovakia.’

*We are grateful to Renata Barzycka, Dorota Klimek-Jankowska, Krzysztof Migdalski and Ewa Tomaszewicz, and to the students of the Philological School of Higher Education (WSF) and the Institute of English Studies (IFA), Wrocław University, without whom we could not have conducted our experiments. Thanks to the audiences of the 34th Penn Linguistics Colloquium, the Slavic Syntax and Semantics Workshop at the 33rd GLOW Colloquium, Wrocław, Poland, and the GIST 1 Workshop on Antilocality and Snowballing Movement, Belgium, and to our many colleagues, too numerous to list here, who have discussed the formal aspects of this work with us. Thanks to Elsi Kaiser for discussion of the data-analysis aspects. We gratefully acknowledge support by a fellowship from the Andrew Mellon Foundation and by an Advancing Scholarship in the Humanities and Social Sciences grant from USC, awarded to Roumyana Pancheva.

As the examples above illustrate, when the *more*-phrase is a direct object, the phrasal and clausal comparative are both acceptable. The same is true when the *more*-phrase is an adverbial, predicative or attributive adjective, an indirect object, or an unaccusative subject (not illustrated here). When the *more*-phrase is an external argument subject, the phrasal comparative is degraded.

Bulgarian, Serbian/Croatian, Slovenian, Greek and Hungarian are like Polish. All have overt *wh*-movement. In conformity with (1a), when the *more*-phrase is an external argument, the phrasal comparative is noticeably degraded relative to its clausal counterpart. When the *more*-phrase is a direct or indirect object, an internal argument subject, an adverbial or an adjectival phrase, both clausal and phrasal comparatives are acceptable.¹ Non-*wh*-fronting languages like Turkish, Hindi, Japanese and Korean, on the other hand, do not exhibit the contrast in (2)-(3).

1.2 Background on the Syntax of Clausal and Phrasal Comparatives

We adopt a common analysis of clausal comparatives. A partial LF is in (4a): *than* has a finite CP complement, where a *wh*-degree operator binds a degree variable in a position corresponding to that of *more* in the matrix. At PF, parts of the clause are elided: the phrase containing the degree variable is elided obligatorily, and any larger containing phrases may be elided optionally. In (4b) the *vP* is elided, as shown by the shaded format.² In English the *wh*-degree operator is null; in other languages (e.g., Bulgarian) it is overt (*kolkoto* ‘how-many’/ ‘how-much’).

- (4) He visited more places than she did.
- | | | | |
|-------------|---|---|------|
| a. ... than | [_{CP} <i>wh-many</i> ₁ she ₂ did | [_{vP} t ₂ visit d₁-places]] | (LF) |
| b. ... than | [_{CP} \emptyset she ₂ did | [_{vP} t ₂ visit d₁-places]] | (PF) |

There are two approaches to phrasal comparatives: the ‘direct analysis’, which posits that *than* has a DP complement, and the ‘reduction analysis’, which takes the complement of *than* to be a clause (see Bhatt and Takahashi (2007) for recent discussion). We adopt the analysis in Pancheva (2009), a variant of the ‘reduction analysis’, as the only account that can capture naturally the generalization in (1).³ Under this analysis, the structure of the phrasal counterpart of (4) is as in (5). *Than* has a small clause complement, whose subject – the remnant – it exceptionally case-marks. We represent the small clause as a focus projection, to highlight the similar information status of remnants in clausal and phrasal comparatives (see footnote 2), but the exact nature of this position is not relevant here.⁴ A *wh*-degree operator moves from a position parallel to that of *more* in the matrix. However, unlike in clausal comparatives, this operator is not attracted by a complementizer. The movement is not feature-driven, but happens purely for the creation of a degree predicate (see Heim and Kratzer 1998 on linking variable abstraction to movement). The movement is very local, to the first node of type *t*, here the *vP*, as necessary for interpretation. The LF resulting from the operations just described is in (5a). At PF the *vP* is elided, as indicated by shading in (5b).⁵

- (5) He visited more places than her.
- | | | |
|-------------|---|------|
| a. ... than | [_{FP} she ₂ [_{vP} <i>wh-many</i> ₁ [_{vP} t ₂ visit d₁-places]]]] | (LF) |
| b. ... than | [_{FP} she ₂ [_{vP} \emptyset [_{vP} t ₂ visit d₁-places]]]] | (PF) |

To summarize, in both phrasal and clausal comparatives the complement to *than* is a clause interpreted as a degree predicate at LF. Consequently, both types of comparatives can employ the same degree quantifier *more*, of which the degree-predicate denoting *than*-clause is an argument

¹English too behaves in conformity with (1a), but because both its phrasal and clausal comparatives are introduced by *than*, it is harder to demonstrate this. See Pancheva (2009), section 4.5, for an argument that comparatives like *More students visited the Czech Republic than Slovakia* must be clausal.

²English may also delete the TP, given that *He visited more places than she* is acceptable for some. The remnant moves out of the TP, possibly to a focus position. See (6)-(7) for such a representation for Polish.

³We do not offer here a critique of the alternative analyses; see Pancheva (2009) for discussion.

⁴Depending on how case is licensed, the remnant may further move to a specifier of the preposition, as in Merchant (2009); note that for him this is an ECM-movement out of a finite, not a small clause.

⁵Languages with overt *wh*-degree operators do not pronounce them in phrasal comparatives. We assume with Pancheva (2009) that the larger *vP* has to be elided, and do not discuss here why this should be so.

(further assuming, as usual, that *than* is vacuous). In both types of comparatives the gradable property in the *than*-clause is syntactically represented and has to meet the conditions on the licensing of ellipsis. On the other hand, the two types of comparatives differ in key aspects, most notably in the locality constraints imposed on *wh*-movement by the different clausal structures.

1.3 Formal Analysis

The Polish *niż*-clauses in (2b) and (3b) have the structure in (6) and (7) (given in English glosses for ease of presentation). The *wh*-degree operator is null, as in English. It moves to Spec, CP from a position parallel to that of *more*. The remnant moves out of the TP, possibly to a focus position. The TP is elided at PF. (6) and (7) differ in the position where the degree variable is left: inside the direct object or inside the subject; notably this difference has no effect on acceptability.

- (6) ... *niż* [CP *wh-many*₁ [FP Anna₂ [TP t₂ [vP t₂ visit d₁-places]]]]
 (7) ... *niż* [CP *wh-many*₁ [FP Slovakia₂ [TP d₁-students₃ [vP t₃ visit t₂]]]]

The Polish *od*-clauses in (2a) and (3a) have the structure in (8) and (9), according to the ‘small clause analysis’ that we are adopting. The preposition *od* has a small-clause complement; *od* ECMs the remnant, which acts as the subject of the small clause. *Wh*-movement to vP, the first node of type *t*, is responsible for the interpretation of the complement of *od* as a degree predicate. Assuming that *od* (like *niż* and *than*) has no semantic contribution, the *od*-PP denotes a degree predicate, just like the *niż*-clause, and is an appropriate argument to *more*. The vP is elided at PF.

- (8) ... *od* [FP Anna₂ [vP *wh-many*₁ [vP t₂ visit d₁-places]]]
 (9) ^{??/*} ... *od* [FP Slovakia₂ [vP *wh-many*₁ [vP d₁-students visit t₂]]]]

Just like (6) and (7), (8) and (9) differ in the position of the degree variable: whether it is inside the direct object or the vP-subject. However, unlike in the case of clausal comparatives, here this difference in the position of the degree variable has a marked effect on acceptability.

The analysis of these facts outlined in (1b) has two key ingredients – an anti-locality constraint on movement and a prohibition against sub-extraction from subjects – both generally applicable in syntax. In phrasal comparatives *wh*-movement targets vP. Movement of the whole *wh*-phrase in Spec, vP (e.g., *wh-many students* in (9)) is precluded by the anti-locality constraint. Movement of just the *wh*-degree part is precluded as a subject-island violation. This explains the unacceptability of phrasal comparatives such as (9)/(3a) where the *more*-phrase is an external argument subject. Importantly, the violation is not triggered by *more* in the matrix clause but by the *wh*-degree operator in the *than*-clause, extracted from a position parallel to that of *more*.

No problem arises in (8)/(2a) because the *wh*-degree operator moves from an object position. Movement from inside VP to vP is not too-local. There is no problem in clausal comparatives either. As (6) and (7) show, *wh*-movement proceeds to Spec, CP, avoiding an anti-locality violation.

2 Gradient Acceptability

The unacceptability of (3a) is not categorical. Speakers of Polish, when consulted informally, would characterize sentences of this type as not very good but not fully unacceptable either. There is a fair amount of variability not only between but also within speakers. Complicating matters, the sentences in (2a) and (3b) were also sometimes reported as less than perfect, though still better than (3a). The indicated judgments in (2)-(3) are therefore an idealization.

Given this variability in speakers’ responses, it would be difficult to claim that the ‘small clause theory’ is empirically superior to its alternatives. It could be the case that phrasal comparatives in Polish are always less acceptable than their clausal counterparts for some reason, perhaps because of language change, or perhaps because of competition, and that some yet unknown difficulty with subject comparatives in particular makes the relative unacceptability of (3b) particularly salient. Clearly, in order to make sense of the data, we need to elicit judgments in controlled conditions allowing subsequent quantitative analysis. This is why we conducted two acceptability-rating studies following standard experimental procedures and testing a large number of speakers.

3 Experiment 1

Our first experiment addresses the question of whether native speakers of Polish find phrasal comparatives with *more* in subject position degraded, relative to their clausal counterparts, as well as relative to both phrasal and clausal comparatives with *more* in object position. In other words, we seek to find statistically significant experimental evidence for the pattern of judgments obtained informally and represented in an idealized fashion in (2)-(3).

3.1 Method

We conducted an off-line acceptability-rating study, comparing phrasal and clausal comparatives with *more* in subject and object position. We manipulated the type of *than* (*od* vs. *niż*) and the position of *more* (subject vs. object) in a 2 × 2 repeated-measures design.

Experimental items consisted of sets of four sentences each, e.g., (10). ‘Object *od*’ is shorthand for ‘phrasal comparative with *more* in object position’; the other conditions follow the same naming convention. Note that the position of *more* is not of interest per se but only as a convenient overt indicator of the position of the non-overt *wh*-degree operator in the *than*-clause.

(10) object *od* – object *niż*

- a. Jak dotąd Justyna przeczytała *więcej* obowiązkowych lektur *od* Iwony.
 b. Jak dotąd Justyna przeczytała *więcej* obowiązkowych lektur *niż* Iwona.
 as till-now Justina read more obligatory readings than Ivona
 ‘So far, Justina has read more of the obligatory readings than Ivona.’

subject *od* – subject *niż*

- c. Jak dotąd *więcej* *moich* kolegów przeczytało Trylogię *od* Lalki.
 d. Jak dotąd *więcej* *moich* kolegów przeczytało Trylogię *niż* Lalkę.
 as till-now more my friends read Trilogy than Lalka
 ‘So far, more of my friends have read the Trilogy than the novel Lalka.’

Within each item, the phrasal and clausal conditions were matched pair-wise and were identical, except for *than* and the case of the remnant. Within each item, the subject and object conditions were closely matched: they used the same transitive verb, in the perfective aspect, and were similar in length, on average 9.8 vs. 9.4 words per sentence, respectively. Structurally, however, they had to differ – the remnant had to be the subject of the *than*-clause in the object conditions, and the direct object in the subject conditions, so that the *wh*-degree operator could originate in the object or subject position, respectively, and move independently of the remnant. To see the point more clearly, consider an alternative: could the subject and object conditions have looked like (11), where both remnants are subjects, or like (12), where the remnants are both objects, instead?

- (11) a. *More students* read books than professors.
 b. Students read *more books* than professors.
 (12) a. *More students* read books than magazines.
 b. Students read *more books* than magazines.

The (a) and (b) sentences appear very well matched, but the non-overt syntax of the remnant and the *wh*-degree operator are very different. In (11a) and (12b) the *wh*-degree operator is initially contained in the remnant, in the other sentences it is not. Thus, we cannot separate the movement of the *wh*-degree operator – our phenomenon of interest – from the movement of the remnant.

- (13) a. *More students* read books than [*wh-many*₁ [*d*₁-professors₂ [_{VP} t₂ read books]]]]
 b. Students read *more books* than [*wh-many*₁ [professors₂ [_{VP} t₂ read *d*₁-books]]]]
 (14) a. *More students* read books than [*wh-many*₁ [magazines₂ [*d*₁-students₃ [_{VP} t₃ read t₂]]]]
 b. Students read *more books* than [*wh-many*₁ [*d*₁-magazines₂ [_{VP} students read t₂]]]]

Our design avoids this problem (see e.g., (6) vs. (7)), but at the cost of having the subject and object conditions describe different events and involve different lexical items, apart from the verb.

We prepared 24 experimental items, based on 24 different transitive verbs, and distributed them into 4 lists, such that each list contained only one condition of each item. We added 48 fillers, the same in each list, consisting of 16 grammatical sentences, 16 ungrammatical sentences (multiple *wh*-questions with some *wh*-words not fronted, sub-extractions across finite clauses, doubling of complementizers), and 16 sentences judged to be of intermediate acceptability (subject sub-extractions within a clause, movement of d-linked *wh*-words from finite clauses). The order of the 72 sentences in each list was pseudo-randomized to ensure that no two experimental items appeared next to one another. A 1-7 rating scale was added under each sentence.

Participants were recruited among college students in Wrocław, Poland. Fifty-six volunteers took part in the experiment. The participants were instructed to rate the sentences according to how natural they sound, with 1 the lowest and 7 the highest rating. Before the experiment began, participants were provided with three examples: a sentence that would probably be rated 6 or 7 by a Polish native speaker, another sentence that would probably be rated anywhere between 3 and 5, and a third sentences that would probably receive a rating of 1 or 2. Participants were told to trust their intuitions and to not spend too much time making a decision about a rating.

3.2 Predictions and Results

We predicted that the ‘subject *od*’ condition would receive the lowest rating of all conditions. We also expected an interaction between the factors in a 2×2 repeated-measures ANOVA, as the level ‘subject’ of the factor ‘position of *more*’ should have a greater negative effect on phrasal than on clausal comparatives. This is indeed what we found, along with some less expected results.

Two participants were excluded from analysis for reversing the scale, as revealed by their performance on the filler items. The scores of the remaining 54 participants were normalized, to standardize the way the 1-7 scale was used. Each rating was converted to a z-score, based on each participant’s mean and standard deviation over the 24 experimental items.⁶ Analyses are reported on the basis of the normalized scores; results on raw scores are similar throughout.

3.2.1 Group Analyses

Descriptive statistics are reported in Table 1.⁷ A two-way repeated-measures ANOVA yields a main effect of type of *than*, with *niż* conditions getting higher ratings than *od* conditions, both in analysis by participant ($F_1(1,53) = 37.77, p < .001$) and by item ($F_2(1,23) = 43.44, p < .001$). There is a main effect of position of *more*, with ‘object’ conditions rated higher than ‘subject’ conditions, by participant ($F_1(1,53) = 38.33, p < .001$) and by item ($F_2(1,23) = 11.38, p = .003$). There is an interaction approaching significance by participant ($F_1(1,53) = 3.88, p = .054$) but not by item ($F_2(1,23) = 1.68, p = .208$). However, see below for results with trimmed outlier scores.

condition	subject <i>od</i>	subject <i>niż</i>	object <i>od</i>	object <i>niż</i>
mean	M = -.48	M = .16	M = -.07	M = .38
std. error	SE = .07 / .10	SE = .06 / .07	SE = .05 / .14	SE = .05 / .06

Table 1: Condition means and standard errors (by participant and by item), 54 participants

Pair-wise comparisons reveal that the ‘subject *od*’ condition is rated significantly lower than each of the other three conditions: ‘subject *od*’ vs. ‘subject *niż*’ $t_1(53) = -5.72, p < .001, t_2(23) = -5.11, p < .001$; ‘subject *od*’ vs. ‘object *od*’ $t_1(53) = -5.26, p < .001$ and $t_2(23) = -2.94, p = .007$; ‘subject *od*’ vs. ‘object *niż*’ $t_1(53) = -7.74, p < .001$ and $t_2(23) = -8.88, p < .001$. Here and elsewhere, paired t-test results are two-tailed and Bonferroni-corrected significance level is .008.

Outlier identification reveals that in the ‘subject *niż*’ condition the scores of 3 participants fall more than 1.5 IQRs below the 25th percentile.⁸ Trimming the 3 scores to the value of 1.5 IQRs

⁶ Z-score = (original score – mean)/st.dev.

⁷ The raw means are 4.35 (‘subject *od*’), 5.60 (‘subject *niż*’), 5.04 (‘object *od*’), 5.96 (‘object *niż*’).

⁸ IQR, the interquartile range, is the difference between the 75th and the 25th percentiles, i.e., the range in which the middle 50% of the data falls. Values in excess of 1.5 IQRs above the 75th percentile or below the 25th percentile are commonly considered outliers.

below the 25th percentile results in a significant interaction ($F_1(1,53) = 4.09, p = .048$). (The condition statistics change to $M = .17, SE = .05$). Two items are outliers in the ‘object *niż*’ condition, the same 2 items plus another are outliers in the ‘object *od*’ condition, and another 2 items are outliers in the ‘subject *niż*’ condition. In each case the scores are lower than 1.5 IQRs below the 25th percentile. Trimming the 7 scores to the value of 1.5 IQRs below the 25th percentile results in a significant interaction ($F_2(1,23) = 5.07, p = .034$). (The condition statistics change to ‘subject *niż*’ $M = .18, SE = .07$; ‘object *od*’ $M = .03, SE = .09$; and ‘object *niż*’ $M = .40, SE = .06$).⁹

3.2.2 Analyses by Subgroups

Syntactic constraints hold for individual speakers. If our measure was sensitive enough, we would expect each participant to follow the predicted pattern of the ‘subject *od*’ condition being rated the lowest. Off-line acceptability rating studies, however, are imprecise, as they filter the results of the implicit grammatical constraints through a layer of explicit decision-making. They are also more prone to errors than on-line measures. Therefore, we need to allow for some deviation from the predicted pattern when examining individual participants’ performance.

The ‘subject *od*’ condition was rated the lowest by each of 38 out of 54 participants. Importantly, for these participants we also obtain a significant interaction in a two-way repeated-measures ANOVA, both by subject ($F_1(1,37) = 20.70, p < .001$) and by item ($F_2(1,23) = 8.52, p = .008$). The two main effects are also significant: type of *than* ($F_1(1,37) = 108.58, p < .001$ and $F_2(1,23) = 80.95, p < .001$); position of *more* ($F_1(1,37) = 97.21, p < .001$ and $F_2(1,23) = 23.50, p < .001$). Pair-wise analyses confirm that the ‘subject *od*’ condition is rated significantly lower than each of the other three conditions: ‘subject *od*’ vs. ‘subject *niż*’ $t_1(37) = -13.01, p < .001$ and $t_2(23) = -7.46, p < .001$; ‘subject *od*’ vs. ‘object *od*’ $t_1(37) = -9.75, p < .001$ and $t_2(23) = -4.91, p < .001$; ‘subject *od*’ vs. ‘object *niż*’ $t_1(37) = -13.37, p < .001$ and $t_2(23) = -10.89, p < .001$.

Three additional participants deviated from the expected pattern only because they rated ‘object *od*’ sentences just slightly less than ‘subject *od*’ sentences (the difference in mean raw scores is $\leq .5$). We assume that their performance is not different than the group of 38 discussed above. When we perform analyses on the larger group of 41, all predictions are met. The ANOVA yields a significant interaction by participant ($F_1(1,40) = 19.15, p < .001$) and by item ($F_2(1,23) = 7.24, p = .013$). The main effects are also significant: type of *than* ($F_1(1,40) = 120.14, p < .001$; $F_2(1,23) = 80.73, p < .001$) and position of *more* ($F_1(1,40) = 64.59, p < .001$; $F_2(1,23) = 18.21, p < .001$). Pair-wise t-tests confirm that the ‘subject *od*’ condition receives a significantly lower rating than the other three conditions: ‘subject *od*’ vs. ‘subject *niż*’ $t_1(40) = -13.41, p < .001$ and $t_2(23) = -7.37, p < .001$; ‘subject *od*’ vs. ‘object *od*’ $t_1(40) = -8.20, p < .001$ and $t_2(23) = -4.39, p < .001$; ‘subject *od*’ vs. ‘object *niż*’ $t_1(40) = -13.19, p < .001$ and $t_2(23) = -10.85, p < .001$.

Of the remaining 13 participants, 7 rated both the ‘object *niż*’ and the ‘subject *niż*’ conditions as low or lower than the ‘subject *od*’ condition, and 6 did so for just the ‘subject *niż*’ condition. Nine of these 13 participants also rated ‘object *niż*’ sentences as low or lower than ‘object *od*’ sentences, i.e., their different pattern of performance could be seen completely independently of the ‘subject *od*’ condition – the condition of interest here. Mindful that we cannot draw firm conclusions about individual participants, we speculate that some of these speakers may have a different dialect with respect to clausal comparatives. Some Polish speakers accept an overt *wh*-word in clausal comparatives, *jak*. Possibly these participants strongly prefer *niż jak* or even just *jak* in clausal comparatives, and this is why they rated our *niż* sentences so low.

3.3 Discussion

As predicted, the ‘subject *od*’ condition received the lowest rating, confirmed by statistical tests. However, in light of the main effects observed in the ANOVA, it is important to ask whether the lowest rating is not simply an additive result of the two factors. The main effects indicate that overall phrasal comparatives are less acceptable than clausal comparatives, and that subject comparatives are less acceptable than object comparatives. These two factors could in principle impact

⁹ The trimmed outliers are 1.4% of the scores in the analysis by participants and 7.3% of the scores in the analysis by items.

independently the acceptability of ‘subject *od*’ sentences, without a role for a subject-island violation. Crucially, we obtained a significant interaction between the factors (in the analysis with trimmed outliers), indicating that the lowest mean for ‘subject *od*’ is not purely cumulative. We attribute this interaction to the presence of a subject-island effect. We further confirmed that the pattern of ‘subject *od*’ sentences receiving the lowest rating obtained individually for a large subset of our participants. The remaining group may have a different dialect for *niż* comparatives.

To summarize, Experiment 1 reveals three notable facts. First, phrasal comparatives in general are less acceptable (main effect of ‘type of *than*’). While we did not expect this finding, it does not contradict the formal analysis. Phrasal comparatives in Czech are no longer productive; possibly Polish is changing in a similar way. Second, comparatives with *more* subjects are degraded relative to comparatives with *more* objects (main effect of ‘position of *more*’). This is likely due to a greater difficulty of extracting from a subject than from an object position, observed in the case of both QR of the *more*-phrase in the matrix and of *wh*-movement of the degree *wh*-phrase in the embedded clause. While this finding too was not expected, it also does not contradict the formal analysis. Finally, the significant interaction between the two factors suggests that the lowest mean for the ‘subject *od*’ condition is not a linear additive effect of two independent factors. The ‘small clause analysis’ predicts this result, and thus receives support from the experimental findings. The alternative theories predict a lack of interaction since, in the absence of a subject-island violation, ‘subject *od*’ comparatives should only show the linear additive effect of the two main factors: *od* being less acceptable than *niż* and subject comparatives being less acceptable than object ones.

4 Experiment 2

Our second experiment addresses the theoretical explanation for the degraded acceptability of subject phrasal comparatives. The analysis in Pancheva (2009) attributes the effect to a subject-island violation. We thus would expect the subject effect in phrasal comparatives in Polish to show a similar pattern of unacceptability as sub-extraction from subjects in degree questions.

4.1 Method

Experiment 2 was an off-line acceptability-rating study of degree dependencies involving subjects in comparatives and questions. We manipulated the type of construction (comparative vs. degree question) and the type of extraction (full vs. sub-extraction) in a 2×2 repeated-measures design. By assumption, clausal comparatives involve full extraction of a *wh-many* nominal phrase from the subject position in the *niż*-clause, and phrasal comparatives involve sub-extraction of a *wh-many* degree operator from the *vP* subject in the *od*-clause.

Experimental items consisted of sets of four sentences, see (15).

(15) subject *od* – subject *niż*

- | | | | | | | | | |
|----|-----------|--------|-------------|--------|--------|----------|------------|---------------|
| a. | Wczoraj | więcej | sprzątaczek | umyło | klatkę | schodową | <i>od</i> | windy. |
| b. | Wczoraj | więcej | sprzątaczek | umyło | klatkę | schodową | <i>niż</i> | windę. |
| | yesterday | more | cleaners | washed | case | stair | | than elevator |
- ‘Yesterday more cleaners washed the staircase than the elevator.’

subject question *sub-extraction* – subject question *full extraction*

- | | | | | | | |
|----|------------|-------------|-------------|--------|--------|-----------|
| c. | <i>Ile</i> | wczoraj | sprzątaczek | umyło | klatkę | schodową? |
| | how-many | yesterday | cleaners | washed | case | stairs |
| d. | <i>Ile</i> | sprzątaczek | wczoraj | umyło | klatkę | schodową? |
| | how-many | cleaners | yesterday | washed | case | stairs |
- ‘How many cleaners washed the staircase yesterday?’

In creating our questionnaires, we followed the same procedure as in Experiment 1. We had 24 experimental items, based on 24 different transitive verbs in the perfective aspect (9 verbs were the same as in Experiment 1, though their arguments were modified to fit better with the question conditions). We had 48 fillers evenly divided among 3 levels of acceptability. The experimental items were distributed over 4 lists following a Latin square, and the same fillers were added to each list. The order of the 72 sentences in each list was pseudo-randomized so no two experimen-

tal items would appear next to one another. A 1-7 rating scale was printed under each sentence. Participants were given the same instructions and practice sentences as in Experiment 1.

Seventy-two volunteers from Wrocław, Poland, different than the participants in Experiment 1, took part in Experiment 2.

4.2 Predictions and Results

We predicted that the ‘subject *od*’ and the ‘subject question sub-extraction’ conditions would receive lower ratings, as both involve a subject-island violation. We thus expected a significant main effect of ‘type of extraction’ in a 2×2 repeated-measures ANOVA. We further predicted an interaction between the two factors, with ‘subject *od*’ rated lower than ‘subject question sub-extraction’, as the effect of sub-extraction would add to the independently observed effect of *od* comparatives being rated lower than *niż* comparatives, regardless of the position of *more*.

One participant was excluded from analysis for reversing the scale, as determined from the performance on fillers. Scores were normalized, based on each participant’s ratings over the 24 experimental items.¹⁰ Analyses are reported on the z-scores.

4.2.1 Group Analyses

Descriptive statistics are shown in Table 2.¹¹ A two-way repeated-measures ANOVA yields a main effect of type of construction, with questions getting higher ratings than comparatives, both by participant ($F_1(1,70) = 14.17, p = .001$) and by item ($F_2(1,23) = 12.52, p = .002$). More importantly for us, there is also a main effect of type of extraction, with the ‘full extraction’ conditions rated higher than the ‘sub-extraction’ conditions, by participant ($F_1(1,70) = 210.02, p < .001$) and by item ($F_2(1,23) = 168.30, p < .001$). Finally, there is an interaction, significant both by participant ($F_1(1,70) = 10.50, p = .002$) and by item ($F_2(1,23) = 14.72, p = .001$).

condition	subject <i>od</i>	subject <i>niż</i>	question sub-extract.	question full extract.
mean	M = $-.62$	M = $.35$	M = $-.16$	M = $.42$
std. error	SE = $.06 / .12$	SE = $.04 / .08$	SE = $.06 / .08$	SE = $.05 / .07$

Table 2: Condition means and standard errors (by participant and by item)

Pair-wise comparisons show that participants gave significantly lower ratings to the ‘subject *od*’ condition than to each of the other conditions: ‘subject *od*’ vs. ‘subject *niż*’ $t_1(71) = -12.70, p < .001$ and $t_2(23) = -13.13, p < .001$; ‘subject *od*’ vs. ‘subject question sub-extraction’ $t_1(70) = -4.17, p < .001$ and $t_2(23) = -4.27, p < .001$; ‘subject *od*’ vs. ‘subject question full extraction’ $t_1(70) = -11.64, p < .001$ and $t_2(23) = -9.95, p < .001$. ‘Subject question sub-extraction’ was rated significantly lower than ‘subject question full extraction’: $t_1(70) = -6.91, p < .001$ and $t_2(23) = -7.10, p < .001$. The two full extraction conditions, ‘subject question full extraction’ and ‘subject *niż*’, did not differ from one another: $t_1(70) = 1.03, p = .307$ and $t_2(23) = 1.06, p = .301$. Here and elsewhere, paired t-tests results are two-tailed and Bonferroni-corrected significance level is .008.

There was one participant outlier in the ‘question full extraction’ condition and one item outlier in the ‘question sub-extraction’ condition. Trimming the two scores to 1.5 IQRs below the 25th percentile preserves the significant main effects (type of construction: $F_1(1,70) = 14.26, p < .001$, $F_2(1,23) = 14.28, p = .001$; type of extraction: $F_1(1,70) = 211.61, p < .001$, $F_2(1,23) = 192.88, p < .001$) and interaction ($F_1(1,70) = 10.47, p = .002$, $F_2(1,23) = 17.20, p < .001$).¹²

Notably the ‘subject *od*’ and ‘subject question sub-extraction’ conditions elicited a wide range of responses, as the histograms below illustrate. Clearly, some participants do not find subject sub-extraction questions degraded, while for others the effect of the subject island violation is sizeable.

¹⁰ Two participants each skipped an item, so their z-scores and averages are calculated over 23 scores.

¹¹ Results are similar when analyses are performed on raw scores. The condition means are 4.05 (‘subject *od*’), 5.72 (‘subject *niż*’), 4.83 (‘subject question sub-extraction’), 5.88 (‘subject question full extraction’).

¹² The trimming affects 0.35% of the scores in the analysis by participants and 1.04% of the scores in the analysis by items.

The fact that the range of ratings for phrasal comparatives is similarly large, lends support to the claim that subject sub-extraction is at play here too, along with other factors.

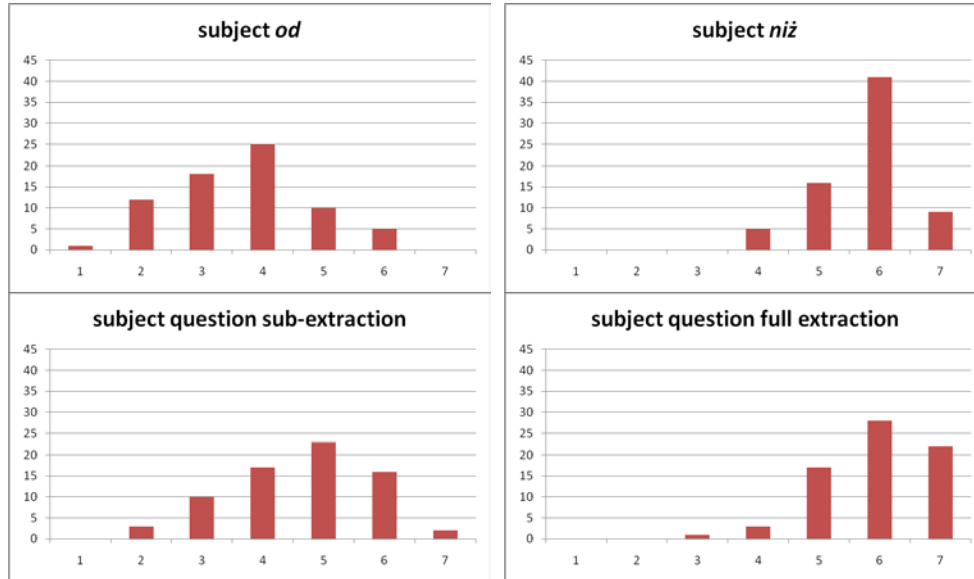


Figure 1: Histograms of the four conditions calculated on z-scores

4.2.2 A Note on Individual Participants' Performance

Without attributing too much importance to specific numbers, for the reasons discussed earlier, it is nevertheless instructive to have a sense of the participants' individual performance. All but 3 participants rated the *od* sentences lower than the *niž* sentences. The 'question sub-extraction' condition was rated lower than the 'question full extraction' condition by all but 15 participants. The 'subject *od*' condition was rated as low or lower than the 'question sub-extraction' condition by 48 out of 71 participants (for 7 more the unexpected difference in raw mean scores was $\leq .5$).

4.3 Discussion

We found a main effect of 'type of extraction', as predicted. We attribute this effect to a subject-island violation. This is an uncontroversial claim in the case of questions. In the case of comparatives, the subject-island effect is predicted by the 'small clause theory'. However, given the main effect of 'type of *than*' obtained in Experiment 1, we can only conclude that the results are consistent with the 'small clause theory', not that they disprove the alternative accounts. It could be that the only reason 'subject *od*' comparatives are rated lower than 'subject *niž*' comparatives is that *od* comparatives in general are less acceptable than *niž* comparatives (Experiment 1). If this effect is stronger than the effect of sub-extraction in questions, an interaction would result, without there being a subject-island effect in subject *od* comparatives. Thus, the interaction in Experiment 2, unlike the one in Experiment 1, has to be interpreted with caution.¹³

The individual participants' performance is largely as expected. The 'small clause analysis' predicts that each participant should rate 'subject *od*' sentences lower than 'subject *niž*' sentences, as indeed mostly happens. The alternative analyses would not necessarily make this prediction, but they are compatible with the finding. Another observation is that most participants show evidence for a subject-island effect in questions, but interestingly not all. This finding is notable in itself because it speaks to the broader issue of the status of subject islands, but it is also of relevance for the acceptability of 'subject *od*' sentences. If indeed there are speakers who do not show evidence of a subject-island effect, they wouldn't find 'subject *od*' sentences degraded, beyond the inde-

¹³ We designed and ran Experiment 2 before analyzing the results of Experiment 1, and did not anticipate the main effect of type of *than* in Experiment 1.

pendent main effect of ‘type of *than*’ observed in Experiment 1. Finally, some participants rated the ‘subject *od*’ condition higher than the ‘subject question sub-extraction’ condition. Possibly, the overt island violation in questions made the unacceptability more salient for these speakers, though it is hard to draw firm conclusions. We are reminded that acceptability ratings are an imperfect measure of grammaticality, and that their main value is in allowing group analyses.

Our findings are of relevance for recent debates on whether or not *vP*-subjects are islands (Nunes and Uriagereka 2000, Gallego and Uriagereka 2007, Stepanov 2007, Chomsky 2008, a.o.). According to some analyses *vP* subjects prohibit sub-extraction for structural reasons; according to others, only raised subjects are islands and this is so because of a freezing effect. Given our proposal and our results concerning ‘subject *od*’ comparatives, *vP*-subjects are islands, as reflected in lower acceptability. Experimental findings by Jurka (2009) are in line with our results. He finds that extraction from *vP*-subjects in German is degraded, rated on average 3.55 on a 1-7 scale, vs. an average of 6.17 for extraction out of *VP*-internal objects. Additionally, he documents a wide range of variability among speakers in how acceptable they find sub-extraction from both *vP*-subjects and *TP*-subjects. We find a similar variability in the ratings of comparatives and questions involving sub-extraction from subjects.

5 Conclusions

Experiment 1 confirms that when the *more*-phrase is an external argument, the phrasal comparative in Polish is degraded. The ‘small clause analysis’ predicts a significant interaction between the factors ‘type of *than*’ and ‘position of *more*’ while the alternative analyses predict a lack of interaction. The results lend support for the ‘small clause analysis’ and against the alternatives. Experiment 2 provides some evidence that the unacceptability of phrasal comparatives is due to a subject-island violation. The ‘small clause analysis’ predicts a significant interaction between the factors ‘type of construction’ and ‘type of extraction’, while the alternative theories do not necessarily predict such an interaction, though they are compatible with it. The results conform to the predictions of the ‘small clause analysis’ but do not provide evidence against the alternatives. Experiment 2 further reveals that violations of subject islands elicit a wide range of acceptability ratings, in line with previous experimental findings. The subject-island effect is gradient.

References

- Bhatt, Rajesh and Shoichi Takahashi. 2007. Direct comparisons: resurrecting the direct analysis of phrasal comparatives. *Proceedings of SALT 17*.
- Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, ed. R. Freidin, C. P. Otero, and M. L. Zubizarreta, 133-166. MIT Press.
- Gallego, Ángel and Juan Uriagereka. 2007. Sub-extraction from subjects: a phase theory account. In *Romance Linguistics 2006*, ed. J. Camacho, N. Flores-Ferrán, L. Sánchez, V. Déprez and M. J. Cabrera, 149-162. John Benjamins.
- Heim, Irene and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Blackwell.
- Jurka, Johannes. 2009. Gradient acceptability and subject islands in German. Ms. University of Maryland.
- Juzwa, Urszula. 2006. The Syntax of Ellipsis in English and Polish: A Comparative View. Doctoral Dissertation, Adam Mickiewicz University.
- Merchant, Jason. 2009. Phrasal and clausal comparatives in Greek and the abstractness of syntax. *Journal of Greek Linguistics* 9: 134–164.
- Nunes, Jairo and Juan Uriagereka. 2000. Cyclicity and extraction domains. *Syntax* 3: 20–43.
- Pancheva, Roumyana. 2009. More students attended FASL than CONSOLE. Paper presented at Formal Approaches to Slavic Linguistics 18, Cornell University.
- Stepanov, Arthur. 2007. The end of CED? Minimalism and extraction domains. *Syntax* 10 (1), 80-126.

Department of Linguistics
 University of Southern California
 Los Angeles, CA 90089–1693
 pancheva@usc.edu
 btomasze@usc.edu