Looking Beyond Information Structure: Evidence from Magnitude Estimation Test Experiments for Weight Effects on Slovak Word Order

Jakob Horsch and Martina Ivanová

Abstract: The principle of end-weight (Quirk et al. 1985) posits that language users prefer to place short ("light") before long ("heavy") constituents because this is easier to cognitively process (Hawkins 1994). Weight effects on constituent order have been discussed for well over a century in languages like English and German (cf. Behaghel 1909, 1930). In Slavic languages, however, they have received little attention so far (cf. Kizach 2012: 251).

Rather, the focus has been on information structure (Short 2002: 494). However, Goldberg's Tenet #5 predicts that general cognitive restraints such as weight effects apply across languages (Goldberg 2003: 219). This article presents the results of a pilot study that investigates a phenomenon known as Heavy NP Shift, in which the $_{VP}$[NP PP] pattern changes to $_{VP}$[PP NP] when the NP is heavier than the PP.

Employing the Magnitude Estimation Test method (Hoffmann 2013), grammaticality acceptability ratings from 39 L1 Slovak speakers were elicited. The results show that Slovak is susceptible to weight effects, such that placing short before long constituents is always preferred. Moreover, the results provide evidence for the existence of a $_{VP}$[V NP PP] pattern in Slovak that has been identified as "basic" for English (Hawkins 1994: 20). This supports Hawkins's (2004) Performance-Grammar Correspondence Hypothesis, which posits that grammaticalized patterns in analytic languages are the result of performance preference and therefore preferred in synthetic languages.

1. Introduction*

The effect of constituent weight (i.e., constituent length and complexity) on word order has been observed and discussed for well over a century

* We would like to thank Monika Hornáček Banášová at the University of Ss. Cyril and Methodius in Trnava, Slovakia, and Peter Barrer at Comenius University in Bratislava, Slovakia, for their invaluable assistance in recruiting subjects and providing facilities for carrying out the Magnitude Estimation Test experiments. We express our gratitude to Thomas Hoffmann, who has provided valuable advice regarding the Magnitude Estimation Test method and statistical analysis of the data.

This paper was written in the context of the research project 09I03-03-V04-00713 "Weight Effects in English and Slovak" at the Ľudovít Štúr Institute of Linguistics of the Slovak Academy of Sciences. The project is generously funded by the Government of the Slovak Republic's Recovery and Resilience program (scholarship for excellent researchers R2-R4), funded in the context of the European Union's NextGenerationEU package.

(cf. Behaghel 1909, 1930). The "principle of end-weight" (Quirk et al. 1985) states that placing light before heavy constituents is preferred due to processing efficiency effects (Hawkins 1994), where "light" means short/less complex, and "heavy" means long/more complex.[1] This light-before-heavy principle can be explained as being rooted in locality-based approaches, which predict a processing advantage when the distance between major syntactic heads is minimized. For example, in the context of his theory of Early Immediate Constituents (EIC), Hawkins (1994) has argued that syntactic heads being placed closer to each other result in locality effects (cf. also Gibson's (1998, 2000) syntactic prediction locality theory). This has led to $_{VP}$[NP PP] becoming the "basic order of the English VP" (Hawkins 1994: 20) because PPs tend to be heavier (i.e., longer and more complex) than NPs. To illustrate this, consider examples (1) and (2) (from Hawkins 1994: 20):

(1) John $_{VP}$[gave $_{NP}$[a book] $_{PP}$[to Mary]]

(2) John $_{VP}$[gave $_{PP}$[to Mary] $_{NP}$[a book]]

According to EIC, English speakers will prefer (1), because here the heads of the three syntactic constituents (V, NP, PP) in the VP are adjacent, which is the optimal situation from a locality perspective for ease of processing. Note that this follows a cognitive approach that does not assume an "unmarked" word order due to semantic roles or case.

Now, when the NP is heavier than the PP, a phenomenon called Heavy NP Shift[2] (HNPS) occurs: The NP is "shifted" so that V and DO are separated by a prepositional phrase (Mains et al. 2015), thereby increasing processing efficiency as in (3) (example from Hawkins 1994: 57):

(3) I $_{VP}$[gave $_{PP}$[to Mary] $_{NP}$[the valuable book that was extremely difficult to find]]

It is noteworthy that HNPS does not occur categorically. Studies have shown that it is most likely when the weight difference between a light PP

[1] In the following text, we operationalize constituent weight as the number of orthographic words in a constituent, a simple and tried-and-tested method (cf., e.g., Szmrecsányi 2004: 1037) based on the observation that the "number of words, number of nodes, and number of phrasal nodes […] are so highly correlated that it is impossible to choose among them on empirical grounds" (Wasow and Arnold 2003: 121). Accordingly, in the following text, "light" is to be understood as fewer, and "heavy" as more words.

[2] The term "shift" originates in analyses based on derivational models of syntax (Mains et al. 2015).

and a heavy NP is at least four words (Hawkins 1994; cf. also Wasow 1997; Stallings and MacDonald 2011). In fact, even when there are considerable differences in weight, HNPS does not always occur (cf. Wasow 2002). In this context, previous studies (e.g., Konieczny 2000; Konieczny and Döring 2003) have offered an explanation in anti-locality effects, which are explained through expectation-based accounts. In contrast to locality theories which are based on memory limitations, expectation-based approaches posit different key cognitive constraints as determining syntactic complexity (cf. Gibson and Wu 2013). A prominent example of an expectation-based approach is the so-called surprisal theory, according to which the difficulty of processing an input in the context in which it appears decreases as the input's conditional probability increases (Hale 2001; Levy 2008; Smith and Levy 2008, 2013). To estimate surprisal, different psychometric variables have been used recently in the theoretical studies, including acceptability judgements (if surprisal correlates with perception, one should expect responses that are surprising in context to obtain lower scores, i.e., a negative relationship between surprisal and score) (e.g., Lau et al. 2017; Wallbridge et al. 2022). Thus, expectation-based theories can also provide explanations for longer-before-shorter orders (cf. Balling and Kizach 2017).

Furthermore, ceiling effects have been observed, to the effect that "increasing NP length appears to bring HNPS only to a par with the canonical [i.e., basic] constituent order, such that HNPS is never significantly preferred over the canonical [i.e., basic] word order" (Medeiros et al. 2021: 437). In other words, speakers are highly reluctant to deviate from $_{VP}$[NP PP], underscoring the fact that it is "basic" and has been "grammaticalized" in English (Hawkins 1994: 20).

What makes HNPS particularly interesting is that it allows investigating cognitive effects on word order with minimal "interference" from other factors because it is a "rare case of phrase ordering that does not affect grammatical role assignment" (Stallings et al. 1998: 395). That is, the direct object NP stays a direct object NP even when shifted;[3] this is not the case with other syntactic alternations such as the ditransitive/to-dative alternation. In fact, HNPS is a perfect candidate for investigating the effects of processing efficiency on constituent order because, unlike other syntactic alternations,[4] it is triggered exclusively by constituent length/complexity. Neither is it stylistically

[3] In generative approaches, HNPS is argued to belong to the category of wh-movement, i.e., it is assumed that A-movement is responsible for the shift (Haegeman and Guéron 1999: 221–24), because the moved NP is no longer in the canonical direct object position and has not been assigned a different case by its new position, which would be the case in A-movement (cf. also Haegeman 1994: 418–21).

[4] HNPS has also been discussed in the context of patterns other than $_{VP}$[V NP PP], e.g., the dative alternation. However, all of these syntactic alternations appear to

motivated nor have animacy effects been found (Stallings et al. 1998: 410). Moreover, Kizach (2014) has noted that speakers "care more about complexity than about givenness" because "the potential benefit of heeding complexity (having short-before-long wherever possible) is […] much larger". Similarly, Šimík (2021: 3) finds that "[i]nformation structure and syntax […] are related only very loosely" in Czech, explicitly refuting the notion that "syntax, which generates [word order] alternations is tightly related to information structure, which motivates them".

The claim that weight effects prevail over givenness when it comes to word ordering is particularly relevant for research on the Slavic languages. While there is a large body of research on the principle of end-weight and HNPS in languages such as English and German (e.g., Behaghel 1909: 139, 1930: 85; Quirk et al. 1972; Quirk et al. 1985; Hawkins 1994, 2004, 2014; Wasow 1997; Arnold et al. 2000; Stallings et al. 1998; Stallings and MacDonald 2011; Mains et al. 2015; Melnick 2017; Medeiros et al. 2021), in Slavic languages little attention has been paid to weight effects so far. For example, Kizach (2012: 251) has noted that "whereas newness has repeatedly been argued as being a relevant factor for Russian, little or no attention has been paid to weight". Rather, the focus has been on information structure, i.e., givenness.[5] For example, Short notes that it is consistently claimed in grammars of Slovak that "pragmatic constituent order takes precedence over syntax" (Short 2002: 494), and that "[m]odern Slovak sources decline to refer to any unmarked order of constituents in terms of basic word order" (Short 2002: 566).[6] In other

be semantically motivated. For the purposes of the present investigation, the term "heavy NP shift" entails none of these other syntactic alternations.

[5] We are grateful to an anonymous reviewer for pointing out that "[w]hile the term 'information structure' was introduced […] by Halliday (1967), by and large in the sense of givenness, its meaning nowadays is broader, encompassing also topic-comment and background-focus structuring (see, e.g., Lambrecht 1994 […] or Krifka 2008)".

[6] Note that Šimík et al. (2014) investigated the impact of givenness on the position of the direct object (DO) in Czech with respect to three clause-mate constituents—subject (S), verb (V), and VP-modifying PP (X). In two controlled acceptability judgment experiments, two main observations were established: (i) for objects in all-new contexts and (ii) for given objects. In contexts with given objects, the order S-V-X-DO (condition 1) was significantly less acceptable than all other orders; the orders S-V-DO-X, S-DO-V-X, and DO-S-V-X (conditions 2, 3, and 4, respectively) showed no significant difference (conditions 2 and 3 violated the GIVEN–NEW principle and condition 4 did not violate it). In this study, only two variables were used in the design of the experiment: referentiality of the object and position of the object. However, in one of the examples presented in Šimík et al.'s (2014) Experiment 1 (in all-new contexts), object NP and VP-modifying PP differed with respect to weight (light/heavy). The results of the study showed that there was no significant difference

words, neither has processing efficiency been discussed as a factor affecting word order in Slovak, nor has there been any mention of a resulting "basic" word order such as the one that Hawkins (1994) determined for English (see above).[7] This is evident from our review of numerous grammars of the Slovak language (Pauliny et al. 1963; Orlovský 1971; Mistrík 1983, 1988, 2003; Pauliny 1981, 1997; Pavlovič 2012; Ivanová 2016) as well as more than six decades' worth of issues of two of the most authoritative linguistic journals on the Slovak language, *Jazykovedný časopis* (https://www.juls.savba.sk/ediela/jc/) and *Slovenská reč* (https://www.juls.savba.sk/ediela/sr/).

However, it is unlikely that weight effects should not play a role in word ordering in Slovak. As a synthetic language, its word order is certainly more flexible than English, but it should nevertheless follow the principle of end-weight and exhibit Heavy NP shift. This is predicted by Goldberg's Tenet #5, which posits that "[c]ross linguistic generalizations are explained by appeal to general cognitive constraints" (Goldberg 2003: 219). Moreover, Hawkins's Performance-Grammar Correspondence Hypothesis (PGCH) states:

> When the grammar of one language is more restrictive and eliminates one or more structural options that are permitted by the grammar of another, the restriction will be in accordance with performance preferences. The preferred structure will be retained and "fixed" as a grammatical convention, the dispreferred structures will be removed. (Hawkins 2004: 5)

In other words, PGCH predicts that the fixed word orders of analytic languages are the result of performance preference due to processing efficiency, and that these fixed word orders will mirror word orders that are preferred in languages with more flexible word orders, i.e., synthetic languages. This

between the word orders *heavy non-referential object NP — light PP* and *light PP — heavy non-referential object NP* (on a Likert-type rating scale with nine points, the first word order achieved a mean rating of 7.79 and the second word order a mean rating of 7.6). Unfortunately, Šimík et al. (2014) did not investigate any weight distinction of the investigated object NPs and PPs in Experiment 2 for given objects. We therefore suggest that the results of their experiment (which is robustly corroborated by the data), i.e., that the scrambling of the given object is just as acceptable as keeping it *in situ* as long as it is not placed clause-finally, should be further tested for weight effects.

[7] As noted by an anonymous reviewer, "for the closely related Czech, it has been taken for granted that there is something like basic word order, see, e.g., Sgall et al. (1980) et seq., where basic order is called 'systémové' (systemic in English texts) and where it is one of the core principles determining the order within the new (not 'contextually bound') area of the utterance". Furthermore, the reviewer has pointed out that Siewierska and Uhlířová (1998) in their survey "also don't seem to mention explicitly a basic order (which they discuss primarily based on Czech and Polish)".

has been demonstrated in corpus-based studies on the Slovak comparative correlative (CC), a biclausal construction (e.g., [*Čím menej rečí tu bude,*]$_{C1}$ [*tým skôr zaspím*]$_{C2}$ 'The less talking there is here, the sooner I will fall asleep'), where Horsch (2020, 2021) showed that there is a statistically significant preference for constituent orders in both subclauses C1 and C2 that mirror the grammaticalized order in English.

These observations have led us to the following **research questions**:

i. As in English, is there a preference for placing light before heavy constituents in Slovak?

ii. Has this led to a basic, grammaticalized order such as the one that has been determined for English?

Our **hypotheses** are as follows:

i. Goldberg's Tenet #5 (Goldberg 2003: 219) maintains that general cognitive restraints such as processing efficiency will result in cross-linguistic generalizations. Based on research on English that has adduced ample evidence in support of the principle of end-weight, and thus, a preference for light constituents to precede heavy constituents, we predict that the same holds true for Slovak.

ii. Hawkins's (2004) PGCH predicts that the fixed word orders of analytic languages such as English will be mirrored in preferred word orders in more synthetic languages such as Slovak. We therefore predict that Slovak also has a constituent order that can be described as "basic". Given the findings for English, we predict that this is the $_{VP}$[NP PP] order.[8]

To answer the research questions outlined above, we decided to carry out experiments using the Magnitude Estimation Test method to elicit grammaticality acceptability judgments from L1 Slovak speakers. The method is discussed in Section 2; the results are presented in Section 3. Section 4 offers

[8] Šimík and Wierzba (2017), using gradient acceptability judgments and multiple regression, showed that, for Czech, Slovak, and Polish, postverbal objects were equally acceptable irrespective of whether the objects were preceded or followed by an indirect object or an adverbial, i.e., the orders $_{VP}$[object NP PP] and $_{VP}$[PP object NP] showed no significant difference in acceptability. Their results display no significant effects on the relative ordering of object NP and PP. However, the interaction between weight order and phrase order was not tested in this study.

a discussion of these results in light of current research, and in Section 5 we offer a conclusion, including suggestions for further research.

2. Methodology: Magnitude Estimation Test (MET)

The MET method (Bard et al. 1996; Cowart 1997; Hoffmann 2013) features grammaticality acceptability judgments,[9] i.e., it requires subjects to "judge stimuli relatively to a reference item" (Hoffmann 2013: 99) instead of restricting them to a scale with absolute values, e.g., the Likert scale. The method is based on the assumption that humans are better at making relative judgments than absolute judgments (Hoffmann 2013: 99).[10] As Cowart has pointed out, the ratings thus obtained translate to "proportional relation among the numbers assigned to different stimuli", which "should reflect the proportions of the stimuli themselves; thus a sentence that is judged twice as good as another should get a number twice as high as that assigned to the other sentence" (Cowart 1997: 73–74).

In practice, MET questionnaires present a set of sentences (test items) to subjects, who rate them with a number that they are basically free to choose. However, they are asked to provide this number "proportional to a constant

[9] Note that such experiments measure *degrees of acceptability* (Hoffmann 2013: 104; who also refers to Featherston 2007a, b; Fanselow 2007; and Newmeyer 2007) and not grammaticality in the Chomskian sense, i.e., whether a sentence belongs to the language system, and is therefore "grammatical", or does not, making it "ungrammatical" (as discussed in Hoffmann 2013: 104). That is to say, while the experiments that are part of the present investigation were carefully designed and implemented following best practices in an attempt to "get as close to grammaticality as possible" (Hoffmann 2013: 104), their results do not permit any definitive claims about the grammaticality (or ungrammaticality, for that matter) of the investigated structures in the Chomskian sense.

[10] As pointed out by an anonymous reviewer, MET as a method has recently also been critically scrutinized (e.g., Weskott and Fanselow 2011; Sprouse 2011; Fukuda et al. 2012; Verhagen et al. 2020). However, as noted by Hoffmann (p.c. in August 2024), MET seems to yield reliable results akin to other methods. Verhagen et al., for example, arrive at the conclusion that "ratings expressed on a Magnitude Estimation scale did not differ systematically from the ratings expressed on a Likert scale" (Verhagen et al. 2020: 61–62), and similarly, Fukuda et al. (2012: 335) found "overwhelming consistency in their results" from Likert scale and MET tests. On a final note, the main point of criticism seems to be that the method is "more difficult for experimenters to implement and requires more mathematical sophistication on the part of experiment participants" (Fukuda et al. 2012: 336).

However, in the authors' own experience with conducting MET experiments for the present study as well as a previous one (Horsch 2023a, b), subjects did not report any notable challenges in understanding and applying the method. Moreover, preparations for the experiments and their implementation did not cause any considerable difficulties.

reference sentence" (Hoffmann 2013: 99) that they have assigned a number of their choosing at the beginning of the experiments. Thus "subjects do not have to rate stimuli on a scale […] which might artificially limit their choices" but rather "decide on their own scale and make as many fine-grained choices as they deem necessary" (Hoffmann 2013: 103). Our questionnaires also featured grammatical and ungrammatical fillers, which later provided baselines against which test items can be compared. This makes it possible to assess test items not just in relation to each other, but also in relation to grammatically correct and incorrect items.

To answer the research questions outlined in §2, we decided to test two variables with two levels each, WEIGHT ORDER (levels: LIGHT-HEAVY, HEAVY-LIGHT) and PHRASE ORDER (levels: NP-PP, PP-NP). We wanted to test all possible factor combinations (conditions):

- LIGHT NP — HEAVY PP
- HEAVY NP — LIGHT PP
- LIGHT PP — HEAVY NP
- HEAVY PP — LIGHT NP

Following Cowart (1997: 49–50), the aim was to present to every participant all four of these conditions but use different lexical material each time to ensure that "the informant is never able to confront a sentence in quite the same way twice" (Cowart 1997: 50). Because of the low number of conditions (four), we decided to create[11] eight lexicalizations to test each condition twice. This also allowed us to install a safeguard against lexical effects since two lexicalizations have been shown to be able to neutralize lexical effects of any items other than the investigated variables (cf. Schütze and Sprouse 2013). Each lexicalization featured a VP consisting of a V, an NP and a PP, the latter of which featured the prepositions *do* '(in)to', *na* 'on(to)', *pod* 'under', *z* 'out of', 'from', *medzi* 'between', and *od* 'from' (4a–h):[12]

[11] Note that due to time constraints, we created experimental sentences from scratch and did not use authentic sentences; it is acknowledged that some experimental studies featuring acceptability ratings use authentic or shortened authentic examples (e.g., Divjak 2017).

[12] In the valency patterns captured in valency dictionaries, object actants and obligatory adverbials are captured by the scheme S–V–O–PP with object actants preceding adverbials. However, this is merely a convention which does not exclude the possible scrambling of the object actant in the sentence structure. The possible influence of the valency status on their ordering within sentence structure has been investigated in several studies on word order for Czech. For example, word order involving object actants and adverbial adjuncts was investigated by Rysová (2011), who demonstrated that valency status of adverbials (obligatory versus non-obligatory) may have an in-

(4) **Lexicalizations**[13]

a. Eva dala (ten drahý) kabát (, čo jej manžel kúpil vo Viedni,) do (profesionálnej) čistiarne (, ktorú jej odporučila suseda,)

'Eve brought the (that expensive) coat (that her husband had bought her in Vienna) to the (professional) dry cleaner (that her neighbor had recommended)'

b. Sestrička odoslala (rôzne) vzorky (, ktoré deň predtým odobrala pani Novákovej,) na (rozsiahlu) kontrolu (, ktorú si vyžiadal jej neurológ,)

'The nurse sent (various) samples (that she had taken from Ms. Nováková the day before) to an (extensive) examination (that her neurologist had asked for)'

c. Hostia položili (svoj veľký) kufor (, ktorý s námahou vynášali po schodoch,) pod (biely) stôl (, čo stojí pri okne v kuchyni,)

'The guests put the (their big) suitcase (that they had laboriously carried up the stairs) under the (that white) table (that is next to the window in the kitchen)'

fluence on word order (with non-obligatory adverbials preceding patients and obligatory adverbials following patients). However, the investigation was focused mostly on manner adverbials and lightness or heaviness of valency actants, and adjuncts were not tested as possible variables affecting word order. In another study, Rysová and Mírovský (2014) focused on testing valency as a word order factor in Czech by studying the results of a series of corpus queries and the decision trees algorithm. The main aim of the study was to test the hypothesis that obligatory adverbials (in terms of valency) follow the non-obligatory (i.e., optional) ones in the surface word order. For the 10-fold cross-validation in the decision trees experiments, various sets of features were tested, including form and length of one of the governed nodes and its subtree (e.g., length in words, length in characters, verbal modality, dependent clause). This study showed that valency (obligatoriness of sentence members) is not a strong factor influencing surface word order in Czech. However, it was also shown that there are several features in their feature selection that influence word order of contextually non-bound free modifiers of a verb in Czech, one of them being the length of verb modifiers (Rysová and Mírovský 2014: 979). In the experiment presented in our study, both obligatory and non-obligatory adverbial PPs were used in the lexicalizations. Hence, further investigations on the possible influence of valency status on the preferential word order is necessary to satisfactorily settle the issue.

[13] Note that the order of NP and PP of these lexicalizations (4a–h) was varied in the actual questionnaire, such that *all possible combinations* (i.e., NP-PP *and* PP-NP) were tested. They are shown here with NP preceding PP simply for ease of exposition.

(4) d. Marek vybral (ten vlnený) sveter (, čo má červené a modré pruhy,) z (mokrej) tašky (, ktorú si celú oblial kávou,)

'Mark took the (that woolen) sweater (that had red and blue stripes) from the (large) bag (all over which he had poured coffee)'

e. Učiteľka vpustila (malých) školákov (, ktorí sa práve vrátili z prestávky,) do (školskej) jedálne (, ktorá príjemne voňala od palaciniek,)

'The teacher let the (little) schoolkids (who just came back from break) into the (school) canteen (that smelled pleasantly like pancakes)'

f. Snúbenci pozvali (tú milú) tetu (, ktorá už dlhé roky žije v zahraničí,) na (letnú) svadbu (, ktorá sa mala konať v júli,)

'The bridal pair invited the (that nice) aunt (who had lived many years abroad) to [their] (summer) wedding (that was going to take place in July)'

g. Janka vložila (svoju novú) knihu (, ktorú jej daroval strýko z Rakúska,) medzi (tie) ostatné (, ktoré už mala odložené na poličke,)

'Joanne put the (her new) book (that her uncle from Austria had given her) between (those) others (that she had already put on the shelf)'

h. Katka dostala (ten veľký) darček (, ktorý bol zabalený v zlatom papieri,) od (najlepších) kamarátok (, ktoré pozvala na svoje narodeniny,)

'Kate got a (that big) present (that was wrapped in golden paper) from her (best) friends (that she had invited to her birthday)'

These eight lexicalizations were then used to create all four factor combinations twice, resulting in 8 × 4 = 32 tokens, which were then divided into four material sets of eight tokens each using the Latin squares method (Hoffmann 2011: 29, cf. also Keller 2000: 60; Keller and Alexopoulou 2005: 1121). The aim was to present every factor combination to every participant twice, but without using the same lexicalization more than once in every material set. Thus, material set 1 appeared as follows (5a–h):

(5) **Test items, material set 1**[14]

a. Eva dala kabát do profesionálnej čistiarne, ktorú jej odporučila suseda.
(LIGHT NP — HEAVY PP)

'Eve brought the coat to the professional dry cleaner that her neighbor had recommended.'

b. Sestrička odoslala rôzne vzorky, ktoré deň predtým odobrala pani Novákovej, na kontrolu.
(HEAVY NP — LIGHT PP)

'The nurse sent various samples, which she had taken from Ms. Nováková the day before, to an examination.'

c. Hostia položili pod stôl svoj veľký kufor, ktorý s námahou vynášali po schodoch.
(LIGHT PP — HEAVY NP)

'The guests put under the table their large suitcase, which they struggled to carry up the stairs.'

d. Marek vybral z mokrej tašky, ktorú si celú oblial kávou, sveter.
(HEAVY PP — LIGHT NP)

'Mark took from the large bag, over which he had poured coffee, the sweater.'

e. Učiteľka vpustila školákov do školskej jedálne, ktorá príjemne voňala od palaciniek.
(LIGHT NP — HEAVY PP)

'The teacher let the schoolkids into the school canteen, which pleasantly smelled of pancakes.'

f. Snúbenci pozvali tú milú tetu, ktorá už dlhé roky žije v zahraničí, na svadbu.
(HEAVY NP — LIGHT PP)

'The bridal pair invited that nice aunt, who had lived many years abroad, to the wedding.'

[14] Translations, which were not part of the actual experiment, have been added. The translations may partially deviate from canonical English word order to better match the original and partly to serve as word-for-word glosses as well.

(5) g. Janka vložila medzi ostatné svoju novú knihu, ktorú jej daroval strýko z Rakúska.
(LIGHT PP — HEAVY NP)

'Jana put between those others her new book, which her uncle from Austria had given her.'

h. Katka dostala od najlepších kamarátok, ktoré pozvala na svoje narodeniny, darček.
(HEAVY PP — LIGHT NP)

'Kate got from her best friends, whom she had invited to her birthday party, a present.'

Following Hoffmann (2013: 108), as many material sets as conditions (i.e., four) were created using the Latin squares method. Sixteen filler items were then created to achieve a filler:stimuli ratio of 2:1, following Cowart's suggestion that "[i]n general, there should be at least twice as many filler sentences as there are [test items]" (Cowart 1997: 92). Half of the filler items (i.e., eight) were grammatical and half were ungrammatical. The latter featured two types of violations,[15] subject-verb agreement (e.g., 6a) and word order (6b):

(6) **Examples of ungrammatical filler items**

a. **Rodičia** sa ma **spýtalo**, či chcem ísť s nimi na dovolenku alebo zostať doma.

'(My) parents asked me if I wanted to join them for a holiday trip or rather stay home.'

b. Adam **ponáhľal sa**, lebo musel stihnúť vlak do Viedne na pracovný pohovor.

'Adam was in a hurry because he had to catch a train to Vienna for a job interview.'

The 16 fillers were added to the eight test items, resulting in a set of 24 tokens. Four material sets, each consisting of 24 tokens, were created such that each material set featured the four possible factor combinations with different lexicalizations and in different places in the questionnaire. The tokens were

[15] Subject-verb agreement violations featured subjects with plural markers and past-tense verbs with singular neuter markers, e.g., in (6a) the subject is *rodičia* 'parents', i.e., plural and masculine animate, whereas the past-tense verb (*sa*) *spýtalo* has a singular neuter marker (*-o*) instead of the plural masculine animate marker (*-i*).

Word order violations featured clitics in third position when they should have been in second, e.g., in (6b) the reflexive pronoun *sa* is in third position when it should be in second position (i.e., following the subject *Adam*).

randomized[16] using blocking, "a controlled randomization procedure" where "each experimental sentence [is assigned] to a block so that only one sentence of each type appears in each block" (Cowart 1997: 94).

The questionnaires, which were printed on ten paper pages, also featured two training sessions to familiarize subjects with the MET method. The questionnaires were filled out under "carefully constructed experimental settings" (Hoffmann 2013: 99) to make sure that the data obtained was valid, objective, and reliable. To this end, they were personally administered by one of the authors at Comenius University in Bratislava and the University of Ss. Cyril and Methodius in Trnava. All subjects were recruited with the help of colleagues, who also organized suitable rooms for carrying out the METs so a comparable setting could be ensured for all subjects. In line with general ethics guidelines, all questionnaires were fully anonymous, only asking subjects to provide metadata such as their age, gender, country, and mother tongue(s), but never their name. Before filling out the questionnaires, subjects were informed about the time frame (about 45 minutes) and the three phases of the experiment (a non-linguistic training phase, a linguistic training phase, and the final experiment). They were also informed that there were no risks for them and that they would remain completely anonymous, and they were informed about their rights, including their right to withdrawal from participation without providing a reason at any time and their right to omit or refuse to answer or respond to any question asked of them. Finally, they were provided with one of the authors' email addresses in case of questions about the study. All of this information was printed on an information sheet that subjects were allowed to take home.

3. Results

A total of 39 L1 Slovak speakers (30 f, 7 m, 2 non-binary) were recruited to fill out the MET questionnaires: 29 students from three different groups from Comenius University in Bratislava (COM) and 10 from the University of Ss. Cyril and Methodius (UCM) in Trnava. The 39 subjects were all university students with an age range of 20–26.

Table 1 on the following page provides an overview of how many subjects from each group filled out which material sets (MS). It shows that at least nine subjects filled out each material set.

[16] The following restrictions were implemented when randomizing the items: The first and last three items were fillers, the first item was a grammatical filler, and test items were never adjacent to each other, i.e., always separated by a filler item.

Table 1. Overview of subject groups and material sets

	MS 1	MS 2	MS 3	MS 4	Totals
COM 1	3	3	3	2	11
COM 2	2	2	1	1	6
COM 3	2	3	3	4	12
UCM	3	3	2	2	10
Totals	10	11	9	9	39

All questionnaires were scanned and the ratings entered into a spreadsheet,[17] where *z*-scores[18] were calculated for further analysis. The *z*-score mean of one particular LEXICALIZATION (0.419) turned out to deviate considerably from the other seven (range: −0.235 to 0.213), so it was excluded from further analysis. This was deemed unproblematic, since every condition was tested with two LEXICALIZATIONS per material set.

Mixed-effects modeling was employed to test for significant random and fixed effects. To this end, the `lmer()` function from the *lme4* package (Bates et al. 2015, 2020) for *R* for Windows was employed. Specifically, the `step()` function from the *lmerTest* package for *R* (Kuznetsova et al. 2015, 2020) was used. This is a function that "automates the step-down approach" used in stepwise regression (Kuznetsova et al. 2015: 8) using the `lmer()` function from *lme4*. Table 2 on the following page provides an overview of the tested fixed and random effects and their levels.

[17] All results, as well as scans of the questionnaires, have been made available on an OSF repository that is accessible via the following link: https://osf.io/kvrwh/.

[18] It is acknowledged that an anonymous reviewer has noted that "Z-score transformation relies on data being normally distributed, which they rarely are in judgment data". However, as noted by Cooksey, "transforming raw scores into *z*-scores does not change the shape of a distribution or rank ordering of individuals within that distribution" (Cooksey 2020: 126). In fact, all it does is make the data comparable so that it can be fed into the mixed-effects models. Furthermore, as an inspection of the distribution of our data revealed, the reviewer's criticism is only partially applicable, anyway: As the histogram in the Appendix shows, while the distribution cannot be described as prototypically Gaussian, neither is it completely random (as noted by Hoffmann, p.c. in August 2024).

Table 2. Overview of fixed and random effects, with levels

Fixed effects	Random effects
WEIGHT ORDER	SUBJECT (001, 002, 003, … 039)
(LIGHT_HEAVY, HEAVY_LIGHT)	GENDER (MALE, FEMALE, NON-BINARY)
PHRASE ORDER (NP_PP, PP_NP)	GROUP (COM1, COM2, COM3, UCM)
	LEXICALIZATION[19] (A, B, C, D, F, G, H)
	MATERIAL SET (MS1, MS2, MS3, MS4)

The full model subjected to stepwise regression using *lmerTest* was thus as follows:

Z_SCORE ~ WEIGHT_ORDER * PHRASE_ORDER + (1|MS)
+ (1|SUBJECT) + (1|GENDER) + (1|LEXICALIZATION) + (1|GROUP)

Stepwise regression showed that none of the random effects were significant. This suggests that the experiment design and implementation were successful in minimizing the influence of extra-linguistic factors and testing the desired linguistic variables only. Table 3 shows the results and the order of elimination of random effects from the model.

Table 3. Likelihood ratio tests for random effects and their order of elimination

	$\chi 2$	Elimination no.	*p*-value
GENDER	0.00	1	1.000[20] n.s.
SUBJECT	0.10	2	0.751 n.s.
MS	0.23	3	0.633 n.s.
GROUP	0.37	4	0.541 n.s.
LEXICALIZATION	0.63	5	0.427 n.s.

[19] Lexicalization E was excluded from analysis because it deviated considerably from the others (see above).

[20] The model estimated a standard deviation of 0 for the random effect GENDER, which resulted in *lme4* returning the message `boundary (singular fit)` and a *p*-value of

Table 4. F-tests for the fixed effects, and their order of elimination

	F	Elimination no.	*p*-value
WEIGHT ORDER: PHRASE ORDER	0.142	1	0.706 n.s.
WEIGHT ORDER	35.70	kept	< 0.001 ***
PHRASE ORDER	5.90	kept	0.016 *

As Table 4 shows, both fixed effects turned out to be significant. WEIGHT ORDER was very significant ($p < 0.001$***); PHRASE ORDER was significant ($p < 0.05$*). The interaction WEIGHT ORDER:PHRASE ORDER was not significant ($p = 0.70645$ n.s.). Thus, the final model determined by stepwise regression using *lmerTest* was as follows:

```
z_score ~ weight_order + phrase_order
```

In other words, both linguistic variables significantly influenced subjects' ratings, with WEIGHT ORDER having a more profound impact than PHRASE ORDER. The implications are discussed in more detail in §4 below.

Figure 1 on the following page plots the *z*-score means of each of the four test items for a visual inspection of the results. Table 5 on the following page provides the *z*-score means as well as standard deviation (SD) and standard error (SE) values. The grammatical filler *z*-score mean was 0.60 (SD 0.72; SE 0.04); for the ungrammatical fillers, this value was −0.62 (SD 0.95; SE 0.05). These values are plotted as "baselines" against which test items can be judged.

As expected (cf. hypotheses (i) and (ii)), the best-rated condition was LIGHT NP — HEAVY PP (*z*-score mean 0.428), i.e., those test items where a light constituent preceded a heavy constituent and where an NP preceded a PP. Conversely, HEAVY PP — LIGHT NP was rated worst (*z*-score mean −0.333). Taking a closer look at WEIGHT ORDER, the plot shows that LIGHT before HEAVY was always preferred by the subjects, regardless of PHRASE ORDER. Recall that this was a very significant effect in the mixed-effects model ($p < 0.001$), indicating that subjects were particularly sensitive to WEIGHT ORDER. Moreover, as Table 5

1.000. This is an indication that the slope "could not be estimated" by the model (Winter 2020: 259)—that is, "the level of 'between-group' variabilityis not sufficient to warrant incorporating random effects in the model" (Bates 2010: 10). In other words, GENDER had a negligible influence on the outcome of the experiment and was therefore removed from the model before proceeding with stepwise regression (as suggested by Bates 2010: 10; cf. also Pasch et al. 2013 and Barr et al. 2013).

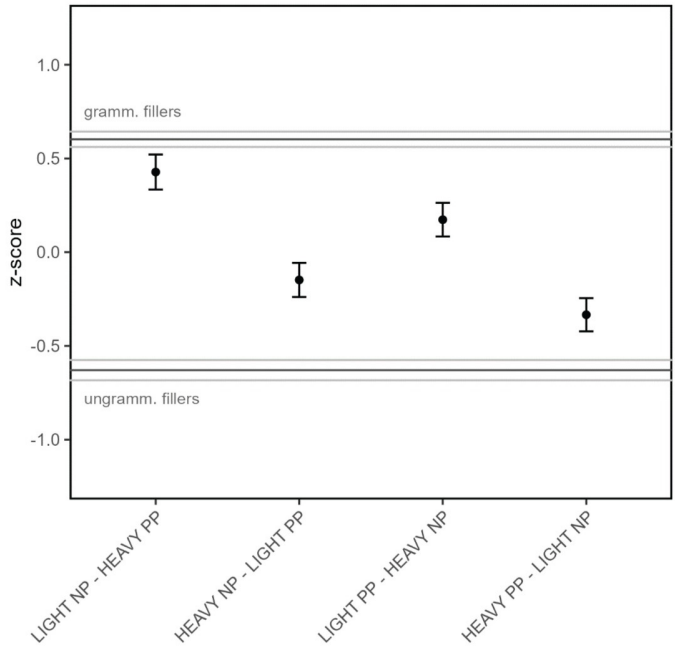

Figure 1. *Z*-score means of WEIGHT ORDER : PHRASE ORDER (error bars indicate standard errors)

Table 5. *Z*-score means of WEIGHT ORDER : PHRASE ORDER

WEIGHT ORDER	PHRASE ORDER	*z*-score	SD	SE
LIGHT_HEAVY	NP_PP	0.428	0.770	0.093
HEAVY_LIGHT	NP_PP	− 0.148	0.756	0.091
LIGHT_HEAVY	PP_NP	0.173	0.748	0.090
HEAVY_LIGHT	PP_NP	− 0.333	0.724	0.088

shows, only LIGHT before HEAVY had positive *z*-scores, whereas HEAVY before LIGHT received negative *z*-scores. This confirms hypothesis (i) that weight effects affect constituent order in Slovak. Concerning PHRASE ORDER, the plot shows that NP before PP was always preferred, confirming hypothesis (ii), i.e.,

the existence of a "basic" $_{VP}$[NP PP] pattern that has been identified in English. However, note that this effect was less significant ($p < 0.05$*) than WEIGHT ORDER, indicating that subjects were less sensitive to it.

In summary, mixed-effects modeling and the plots have confirmed both hypotheses outlined in §1, that is, LIGHT before HEAVY constituents are always preferred. The same applies to NP before PP. Also, the results show that WEIGHT ORDER plays a considerably more important role than PHRASE ORDER in Slovak.

4. Discussion

The data presented here are the result of the first-ever investigation of weight effects in Slovak using an experimental approach, and the first using the MET method to investigate weight effects in any language. We have been able to answer both research questions, demonstrating that weight effects do affect Slovak constituent order and that there is evidence for the existence of a "basic" $_{VP}$[NP PP] pattern. These findings are in line with claims made in the literature upon which our hypotheses were based. Firstly, regarding hypothesis (i), Goldberg's Tenet #5 posits that "[c]ross-linguistic generalizations are explained by appeal to general cognitive constraints" (Goldberg 2003: 219). Weight effects are the result of processing efficiency (Hawkins 1994), which is a general cognitive constraint in this sense. Therefore, it is unsurprising that they affect word order in Slovak in the same way as in English, that is, producing a preference for placing light before heavy constituents. This effect was shown to be very significant ($p < 0.001$***) in the data that we collected from L1 Slovak speakers.

Secondly, regarding hypothesis (ii), Hawkins's (2004) PGCH predicts that grammaticalized patterns in analytic languages are the result of performance preference and will mirror those that are preferred in synthetic languages. Our results confirm this prediction, as evidenced by the preference for the NP-PP phrase order:[21] All other things being equal (i.e., weight order), subjects always preferred NPs preceding PPs. This effect was significant ($p < 0.05$*), indicating that phrase order plays a less important role than weight order. An

[21] We are grateful to an anonymous reviewer who pointed out the experiment in Šimík and Wierzba (2017), which "involved the same type of comparison; the authors found a numerical, but not statistically significant, preference for the NP-PP order in Slovak, Czech, and Polish". As the anonymous reviewer observed, it should be noted "that Šimík [and] Wierzba's experiment had more power (many more lexicalizations and also more participants)" and that, "[g]iven that the phrase order effect is not as large (as compared to the weight order effect), it's good to be slightly more cautious about the result" of the present study. Furthermore, "it cannot be ruled out that the NP-PP preference reduces to [sic] the weight effect. After all, PPs have at least two words (and nodes, in most syntactic frameworks), while NPs can easily have just one".

interaction of weight order and phrase order could not be detected, suggesting that they are independent of each other.

5. Conclusion

The results presented here are part of a pilot study. As such, they are to be considered a first, rather rudimentary, and therefore anything but conclusive attempt at determining whether weight effects play a role at all in Slovak word order. However, our first results do indicate that this is the case. Moreover, we have found evidence for the existence of a "basic" NP-PP phrase order in the Slovak VP. Both of these findings address a significant gap in research on the effects on word order beyond information structure. This applies to Slovak in particular and the Slavic languages in general, where the principle of end-weight and processing efficiency have received little to no attention so far; cf., e.g., Kizach's remarks on Russian (Kizach 2012: 251). Given our findings, however, we believe that weight effects and basic phrasal patterns cannot be ignored in research on word order in Slavic languages. The fact that both weight order and phrase order turned out to be statistically significant effects in our study has serious implications for approaches that disregard weight effects or consider them secondary to information structure. Certainly, future editions of Slovak grammars, which have so far all but ignored weight effects, should at the very least briefly mention the principle of end-weight. They should also discuss the idea of basic or grammaticalized phrasal patterns that arise as a result of performance preference (cf. Hawkins's 2004 PGCH), addressing a serious gap in research that Fried (2017: 241) has described as follows: "syntactic patterning […] has been generally left just about untouched in Slavic linguistics".

As a pilot study, our data is anything but comprehensive. For example, we tested only four out of eight possible conditions (and not, e.g., LIGHT NP — LIGHT PP or HEAVY PP — LIGHT PP). From a methodological point of view, we acknowledge that the experiment design leaves room for improvement. For example, authentic sentences from, e.g., corpora can be used to create test items (see fn 7). We also did not take into consideration so-called "wrap-up" effects[22] that may affect acceptability ratings of sentences. Furthermore, the phrases we used as test items varied in length (measured as number of words), if only slightly; verbs with different valency were used, and some of the relative clauses also differed with respect to perspective or subject maintenance

[22] That is, participants' decisions about sentence acceptability/grammaticality may be different for words toward the end of sentences compared to the preceding words because their decisions become more fixed at later points. For this reason, it has been suggested that test items be constructed in such a way that they are followed by a coordinated main clause to reduce such wrap-up effects (e.g., Balling and Kizach 2017).

and restrictiveness. All of these factors might have skewed the participants' acceptability ratings, such that in future, more extensive investigations will be necessary to examine the relevance of these factors for acceptability ratings. Due to time constraints, however, none of these issues could be addressed in this pilot study.

Nevertheless our—admittedly very preliminary—results do demonstrate that there is a need for more investigations on factors influencing word order in Slovak (and Slavic languages, for that matter) beyond information structure. Such studies should be based on more data and more types of data, and they should take into account more linguistic variables. For example, corpus data can be a valuable addition to the experimental data presented. Syntactic patterns consisting of a verb followed by an NP and a PP in either order can be extracted from corpora that have been annotated using Universal Dependencies, and frequencies of each order can be determined. Following Kizach's (2012: 270) findings that there was a "correlation between the most efficient orders and the most frequent orders" in his corpus study using data from the Russian National Corpus, it is plausible that Slovak corpus data will show similar effects. That is, a higher frequency of corpus tokens with light before heavy constituents, as well as a higher frequency of the $_{VP}$[NP PP] pattern over $_{VP}$[PP NP]. Corpus data can also help shed more light on the effects of weight difference between NP and PP on the likelihood of HNPS occurring (see discussion in §1), and ceiling effects in this regard (cf. Medeiros et al. 2021).

Moreover, previous research on English has shown that the argument structure properties of particular verbs may play a role. Stallings et al. (1998) claim in the context of their "verb disposition hypothesis" that verbs that can take sentential complements in addition to NPs, such as *find* and *reveal*, are more likely to be involved in HNPS than those that can only take NPs, such as *transfer* and *delay* (Stallings et al. 1998: 397). Such effects can also be investigated using corpus data, by looking at the likelihood of HNPS occurring with particular verbs. Also, as Kizach (2012: 270) has noted, "an interesting topic for future research would be to investigate how givenness (theme-rheme) and weight interact". For example, future investigations could incorporate the variable INFORMATION STRUCTURE, with the levels TOPICAL and FOCUSED, depending on the left context of a corpus token or the contents of a short text that subjects must read before filling out an MET questionnaire. This way, it can be determined which factor is more important when it comes to word order, weight effects or information structure.

Finally, extending research on word order in general and HNPS in particular to other languages appears warranted in light of our findings, as our results have confirmed the cross-linguistic predictions made by Goldberg's Tenet # 5 (Goldberg 2003: 219) and Hawkins's (2004) PGCH.

References

Arnold, Jennifer, Anthony Losongeo, and Ryan Ginstrom. (2000) "Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering". *Language* 76(1): 28–55.

Balling, Laura W., and Johannes Kizach. (2017) "Effects of surprisal and locality on Danish sentence processing: An eye-tracking investigation". *Journal of psycholinguistic research* 46(5): 1119–36.

Bard, Ellen G., Dan Robertson, and Antonella Sorace. (1996) "Magnitude estimation of linguistic acceptability". *Language* 72(1): 32–68.

Barr, Dale J., Roger Levy, Christopher Scheepers, and Harry J. Tily. (2013) "Random effects structure for confirmatory hypothesis testing: Keep it maximal". *Journal of memory and language* 68(3): 255–78.

Bates, Douglas. (2010) *lme4: Mixed-effects modeling with R*. New York: Springer.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. (2015) "Fitting linear mixed-effects models using lme4". *Journal of statistical software* 67(1): 1–48.

————. (2020) Reference manual for Package "lme4". Available at: http://cran.r-project.org/web/packages/lme4/lme4.pdf.

Behaghel, Otto. (1909) "Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern". *Indogermanische Forschungen* 25: 110–42.

————. (1930) "Von deutscher Wortstellung". *Zeitschrift für Deutschkunde*. Vol. 44 of *Zeitschrift für den deutschen Unterricht*. Leipzig: B. G. Teubner, 82–89.

Cooksey, Ray W. (2020) "Descriptive statistics for summarising data". *Illustrating statistical procedures: Finding meaning in quantitative data*. Singapore: Springer, 61–139.

Cowart, Wayne. (1997) *Experimental syntax: Applying objective methods to sentence judgements*. Thousand Oaks, CA: Sage.

Divjak, Dagmar. (2017) "The role of lexical frequency in the acceptability of syntactic variants: Evidence from *that*-clauses in Polish". *Cognitive science* 41(2): 354–82.

Fanselow, Gisbert. (2007) "Carrots—perfect as vegetables, but please not as a main dish". *Theoretical linguistics* 33(3): 353–67.

Featherston, Sam. (2007a) "Data in generative grammar: The stick and the carrot". *Theoretical linguistics* 33(3): 269–318.

————. (2007b) "Reply". *Theoretical linguistics* 33(3): 319–33.

Fried, Mirjam. (2017) "Construction grammar in the service of Slavic linguistics, and vice versa". *Journal of Slavic linguistics* 25(2): 241–76.

Fukuda, Shin, Grant Goodall, Dan Michel, and Henry Beecher. (2012) "Is magnitude estimation worth the trouble?" Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz, and Alex Trueman, eds. *Proceedings of the 29th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, 328–36.

Gibson, Edward. (1998) "Linguistic complexity: Locality of syntactic dependencies". *Cognition* 68(1): 1–76.

————. (2000) "The dependency locality theory: A distance-based theory of linguistic complexity". Alec P. Marantz, Yasushi Miyashita, and Wayne O'Neil, eds. *Image, language, brain: Papers from the first mind articulation project symposium*. Cambridge, MA: MIT Press, 94–126.

Gibson, Edward, and Iris H. H. Wu. (2013) "Processing Chinese relative clauses in context". *Language and cognitive processes* 28(1–2): 125–55.

Goldberg, Adele E. (2003) "Constructions: A new theoretical approach to language". *TRENDS in cognitive sciences* 7(5): 219–24.

Haegeman, Liliane. (1994) *Introduction to government & binding theory*. 2nd ed. Oxford: Blackwell.

Haegeman, Liliane, and Jacqueline Guéron. (1999) *English grammar—a generative perspective*. Oxford: Blackwell.

Hale, John T. (2001) "A probabilistic Earley parser as a psycholinguistic model". *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, vol. 2. Pittsburgh, PA: Association for Computational Linguistics, 159–66.

Hawkins, John A. (1994) *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press. [Cambridge Studies in Linguistics 73.]

————. (2004) *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

————. (2014) *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.

Hoffmann, Thomas. (2011) *Preposition placement in English: A usage-based approach*. Cambridge, UK: Cambridge University Press. [Studies in English Language.]

————. (2013) "Obtaining introspective acceptability judgements". Manfred Krug and Julia Schlüter, eds. *Research methods in language variation and change*. Cambridge, UK: Cambridge University Press, 99–118.

Horsch, Jakob. (2020) "Vidíš tým lepšie, čím bližšie sa pozeráš: Slovak comparative correlative CC' constructions from a construction grammar perspective". *Jazykovedný Časopis* 71(1): 25–40.

————. (2021) "Slovak comparative correlatives: A usage-based construction grammar account". *Constructions and frames* 13(2): 193–229.

Horsch, Jakob. (2023a) *The comparative correlative construction in World Englishes: A usage-based construction grammar approach*. PhD thesis, University of Eichstätt-Ingolstadt, Eichstätt, Germany. Available at: https://opus4.kobv.de/opus4-ku-eichstaett/frontdoor/index/index/docId/833.

————. (2023b) "Corroborating corpus data with elicited introspection data: A case study". *Jazykovedný Časopis* 74(2): 60–69.

Ivanová, Martina. (2016) *Syntax slovenského jazyka*. 2nd ed. Prešov, Slovakia: University of Prešov.

Keller, Frank. (2000) *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. PhD thesis, University of Edinburgh.

Keller, Frank, and Theodora Alexopoulou. (2005) "A crosslinguistic, experimental study of resumptive pronouns and that-trace effects". Bruno G. Bara, Lawrence Barsalou, and Monica Bucciarelli, eds. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. London: Psychology Press, 1120–25. Available at : https://escholarship.org/uc/item/6xg3x3hx.

Kizach, Johannes. (2012) "Evidence for weight effects in Russian". *Russian linguistics* 36(3): 251–70.

————. (2014) "Complexity, givenness and g-maze". Abstract of talk at the LinG/RTG2636 Colloquium, University of Göttingen. Available at: https://www.uni-goettingen.de/en/487714.html. Last accessed 25 October 2023.

Konieczny, Lars. (2000) "Locality and parsing complexity". *Journal of psycholinguistic research* 29(6): 627–45.

Konieczny, Lars, and Philipp Döring. (2003) "Anticipation of clause-final heads: Evidence from eye-tracking and SRNs". Peter P. Slezak, ed. *Proceedings of the 4th International Conference on Cognitive Science*, 13–17 July 2003. Sydney: University of New South Wales, 330–35.

Krifka, Manfred. (2008) "Basic notions of information structure". *Acta linguistica Hungarica* 55(3–4): 243–76.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. (2020) lmerTest: Tests for random and fixed effects for linear mixed effects models (lmer objects of lme4 package). R package version 3.1-3. Available at: https://CRAN.R-project.org/package=lmerTest.

Kuznetsova, Alexandra, Rune H. B. Christensen, Cecile Bavay, and Per B. Brockhoff. (2015) "Automated mixed ANOVA modeling of sensory and consumer data". *Food quality and preference* 40: 31–38.

Lambrecht, Knud. (1994) *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge, UK: Cambridge University Press. [Cambridge Studies in Linguistics 71.]

Lau, Jey Han, Alexander Clark, and Shalom Lappin. (2017) "Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge". *Cognitive science*, 41(5): 1202–41.

Levy, Roger. (2008) "Expectation-based syntactic comprehension". *Cognition* 106: 1126–77.

Mains, Paul, Kevin McGowan, and David J. Medeiros. (2015) "Gradient acceptability by length in heavy NP shift". Presented at the 89th Annual Meeting of the Linguistic Society of America, Portland, OR.

Medeiros, David, Paul Mains, and Kevin B. McGowan. (2021) "Ceiling effects on weight in heavy NP shift". *Linguistic inquiry* 52(2): 426–40.

Melnick, Robin Jeffrey. (2017) *Consistency in variation: On the provenance of end-weight*. PhD thesis, Stanford University.

Mistrík, Jozef. (1983) "Moderná slovenčina. 2. vyd". Bratislava: Slovenské pedagogické nakladateľstvo.

————. (1988) *A grammar of contemporary Slovak*. Bratislava: Slovenské pedagogické nakladateľstvo.

————. (2003) *Gramatika slovenčiny*. Bratislava: Slovenské pedagogické nakladateľstvo.

Newmeyer, Frederick J. (2007) "Commentary on Sam Featherston, 'Data in generative grammar: The stick and the carrot'". *Theoretical linguistics* 33(3): 395–99.

Orlovský, Jozef. (1971) *Slovenská syntax*. Bratislava: Obzor.

Pasch, Bret, Benjamin M. Bolker, and Steven M. Phelps. (2013) "Interspecific dominance via vocal interactions mediates altitudinal zonation in neotropical singing mice". *The American naturalist* 182(5): E161–E173.

Pauliny, Eugen. (1997) *Krátka gramatika slovenská*. Bratislava: Národné literárne centrum—Dom slovenskej literatúry.

————. (1981) *Slovenská gramatika (Opis jazykového systému)*. Bratislava: Slovenské pedagogické nakladateľstvo.

Pauliny, Eugen, Jozef Ružička, and Jozef Štolc. (1963) *Slovenská gramatika*. Bratislava: Slovenské pedagogické nakladateľstvo.

Pavlovič, Jozef. (2012) *Syntax slovenského jazyka I*. Available at: https://pdfweb.truni.sk/download?e-skripta/syntax-1.pdf.

Quirk, Randolph, Sidney Greenbaum, and Geoffrey Leech. (1972) *A grammar of contemporary English*. London: Longman.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. (1985) *A comprehensive grammar of the English language*. New York: Longman.

Rysová, Kateřina. (2011) "Ke slovosledu v konstrukcích s obligatorními participanty (se zaměřením na slovosled patientu a způsobového slovesného doplnění)". Vladimír Petkevič and Alexandr Rosen, eds. *Korpusová lingvistika Praha 2011—Gramatika a značkování korpusů*. Prague: Nakladatelství Lidové noviny, 62–69.

Rysová, Kateřina, and Jiří Mírovský. (2014) "Valency and word order in Czech—A corpus probe". Nicoletta Calzolari, Khaled Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, eds. *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*. Reykjavík, Iceland: European Language Resources Association, 975–80.

Schütze, Carson T., and Jon Sprouse. (2013) "Judgment data". Robert J. Podesva and Devyani Sharma, eds. *Research methods in linguistics*. Cambridge, UK: Cambridge University Press, 27–50.

Sgall, Peter, Eva Hajičová, and Eva Buráňová. (1980) *Aktuální členění věty v češtině*. Prague: Academia.

Short, David. (2002) "Slovak". Bernard Comrie and Greville G. Corbett, eds. *The Slavonic languages*. London: Routledge, 533–92. [Routledge Language Family Descriptions.]

Siewierska, Anna, and Ludmila Uhlířová. (1998) "An overview of word order in Slavic languages". Anna Siewierska, ed. *Constituent order in the languages of Europe*, vol. 1. Berlin, New York: De Gruyter Mouton, 105–50. [Empirical Approaches to Language Typology.]

Smith, Nathaniel J., and Roger Levy. (2008) "Optimal processing times in reading: A formal model and empirical investigation". *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* 30(30): 595–600.

————. (2013) "The effect of word predictability on reading time is logarithmic". *Cognition* 128(3): 302–19.

Sprouse, John. (2011) "A test of the cognitive assumptions of magnitude estimation: Commutativity does not hold for acceptability judgments". *Language* 87(2): 274–88.

Stallings, Lynne M., and Maryellen C. MacDonald. (2011) "It's not just the 'heavy NP': Relative phrase length modulates the production of heavy-NP shift". *Journal of psycholinguistic research* 40(3): 177–87.

Stallings, Lynne M., Maryellen C. MacDonald, and Padraig G. O'Seaghdha. (1998) "Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift". *Journal of memory and language* 39(3): 392–417.

Szmrecsányi, Benedikt. (2004) "On operationalizing syntactic complexity". Gérard Purnelle, Cédrick Fairon, and Anne Dister, eds. *Le poids des mots. Proceedings of the 7th International Conference on Textual Data Statistical Analysis*. Louvain-la-Neuve, Belgium: Université Catholique de Belgique (UCLouvain), 1032–39.

Šimík, Radek. (2021) *Information structure and its relation to syntax, semantics, and reference*. Unpublished manuscript. Available at: https://lingbuzz.net/005822.

Šimík, Radek, Marta Wierzba, and Beste Kamali. (2014) "Givenness and the position of the direct object in the Czech clause". Cassandra Chapman, Olena Kit, and Ivona Kucerová, eds. *Proceedings of Formal Approaches to Slavic Linguistics (FASL) 22: The McMaster Meeting 2013*. Ann Arbor, MI: Michigan Slavic Publications, 302–18.

Šimík, Radek, and Marta Wierzba. (2017) "Expression of information structure in West Slavic: Modeling the impact of prosodic and word-order factors". *Language* 93(3): 671–709.

Verhagen, Véronique, Maria Mos, Joost Schilperoord, and Ad Backus. (2020) "Variation is information: Analyses of variation across items, participants, time, and methods in metalinguistic judgment data". *Linguistics* 58(1): 37–81. DOI 10.1515/ling-2018-0036.

Wallbridge, Sarenne, Peter Bell, and Catherine Lai. (2022) "Investigating perception of spoken dialogue acceptability through surprisal". *Interspeech 2022: The 23rd Annual Conference of the International Speech Communication Association*. Red Hook, NY: Curran Associates, 4506–10.

Wasow, Thomas. (1997) "Remarks on grammatical weight". *Language variation and change* 9(1): 81–105.

————. (2002) *Postverbal behavior*. Stanford, CA: Center for the Study of Language and Information.

Wasow, Thomas, and Jennifer Arnold. (2003) "Post-verbal constituent ordering in English". *Determinants of grammatical variation in English*. Berlin: De Gruyter, 119–54.

Weskott, Thomas, and Gisbert Fanselow. (2011) "On the informativity of different measures of linguistic acceptability". *Language* 87(2): 249–73.

Winter, Bodo. (2020) *Statistics for linguists: An introduction using R*. New York/London: Routledge.

Appendix

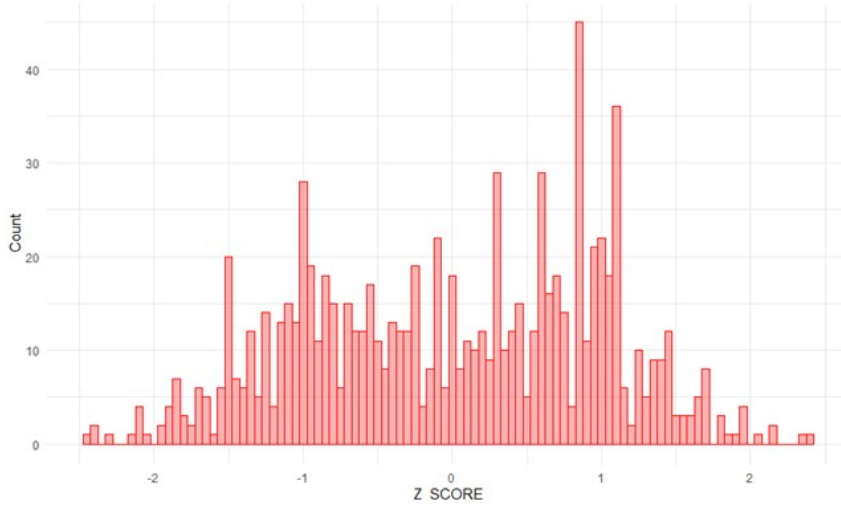

Figure 2. Distribution of *z*-scores in the MET data set

Jakob Horsch
Slovak National Corpus Department
Ľudovít Štúr Institute of Linguistics of the Slovak Academy of Sciences
Bratislava, Slovakia
jakob.horsch@korpus.sk

Martina Ivanová
Institute of Slovak and Media Studies
Faculty of Arts
University of Prešov
Prešov, Slovakia
and
Ľudovít Štúr Institute of Linguistics of the Slovak Academy of Sciences
Bratislava, Slovakia
martina.ivanova@unipo.sk