

Contextual triggers of the Hungarian pre-verbal focus structure

A guided production study

Tamás Káldi^{1,2}, Levente Madarász¹ & Anna Babarczy^{1,2}

¹Research Institute for Linguistics, Hungarian Academy of Sciences

²Department of Cognitive Sciences, Budapest University of Technology and Economics

The study uses a novel experimental method to investigate contextual factors claimed in the theoretical literature to license the use of Hungarian pre-verbal focus. These factors are: (i) identification, (ii) contrast, (iii) availability of a set on which the focus operates and (iv) whether this set is explicit or implicit. We tested the effects of these factors using online surveys in which respondents read short texts describing a context and saw a cloud of randomly arranged words. The experimental task was to create sentences that naturally fit the context by clicking the words in the cloud. Results show that narrow identification and contrast reliably predict the use of pre-verbal focus as does the availability of a set regardless of explicitness.

Keywords: Hungarian focus, word order, identification, contrast, contextual factors, production experiment, information packaging

1. Introduction

1.1 Hungarian: A free word-order language

Hungarian is a free word order language; almost any order of constituents will make up a grammatical structure. Consider the (incomplete) list of word order variations in (1): any one of the examples is a grammatical Hungarian sentence.

- (1) a. Meg-harapta¹ a kutya a postást.
 VM²-bite-PST.3SG the dog-NOM-SG the postman-SG-ACC
 ‘The dog bit the postman.’
- b. Meg-harapta a postást a kutya.
 c. A kutya meg-harapta a postást.
 d. A postást meg-harapta a kutya.
 e. A kutya a postást meg-harapta.
 f. A postást a kutya meg-harapta.
 g. A postást a kutya harapta meg.
 h. A kutya a postást harapta meg.
 i. A postást harapta meg a kutya.
 j. A kutya harapta meg a postást.

As it is a generally valid assumption that nothing is without a cause, a question regarding the observation in (1) promptly presents itself: what causes speakers of Hungarian to opt for one of the above word order variations in one particular instance, and why are they using another one in another instance? This is the question that the current study attempts to at least partially answer with respect to two sentence types in (1): what factors contribute to the use of (1c), which will be referred to by the well-established term “neutral” sentence (see e.g. Kálmán 1985), and (1h), which we shall label “pre-verbal focus” sentence. We will argue that two of the most crucial factors are identification and contrast; and we will propose that both of these can be captured in terms of the availability of sets, whether explicit, implicit or evoked in the context. Before elaborating on these factors, first we provide a short introduction to the basic notions of sentential focus and related aspects in the following sections.

1.2 Accounts of word order: Discourse configurationality

The question formulated above was first raised and meticulously studied by Sámuel Brassai (1852, 1860) who divided the sentence into two parts: the *inchoativum* and the *bulk* (*zöm*). In Brassai’s account, the former contains elements that “practically lay a basis for the meaning of the sentence in the listener’s mind, i.e., they are calling attention, and pointing forward, connecting the mental activity of

1. According to the Hungarian orthographic conventions, the verbal modifier (VM) and the verb are written as one word. In the current study hyphenated forms are used for illustrative purposes.

2. The abbreviation VM will be used for Verbal Modifier.

the listener with that of the speaker”³ (1860: 351). As far as the *bulk* is concerned, its function is “the communication of an action or a circumstance of an action that the speaker supposes to be unknown to the listener” (1860: 72). Although in certain cases the *bulk* may contain old information, it can be argued that it is the more informative part of the sentence. For this reason, as Brassai also points out, the *inchoativum* is optional, while the *bulk* is obligatory. In another study, Brassai mentions that sentences should not be studied only in isolation, since they serve the role of building blocks in texts. That is, “speech never serves to express merely single thoughts, but sequences of them, and just as every word, every form has a certain word order value, every sentence, every structure of thought has a given sentence order value” (Brassai 1885: 30-31).

The passages from Brassai’s works cited above clearly outline a framework in which the use of the different word order variations exemplified in (1) can be accounted for. There are two closely related considerations: one of these is the information status of a particular element (given or new, emphatic or backgrounded), while the second is text coherence. Although at first sight it may seem redundant to differentiate the two aspects, in our opinion, the distinction is justifiable. If a text is to be coherent, then the information in the sentences it is comprised of must be structured with respect to their status, as required by the text. However, different word order variations of a sentence can appear in the same text depending on what is emphasized without compromising coherence. Also, isolated sentences can express the information status of their constituents via particular word order variations. Therefore, it is not only coherence that determines word order, but also, and perhaps more importantly, the information status of the elements. For this reason, Brassai’s findings bear special relevance to our investigation, since it is his work that first meticulously studied the interrelations between context, the psychological state of the interlocutors and the use of different word order patterns in Hungarian.

The issue of the relationship between information status and word order (in Hungarian and also in other languages) was taken up in the generative tradition by É. Kiss (1995). By this time, the terms *Topic* and *Focus* had been introduced in the literature,⁴ therefore É. Kiss (1995) could formulate her linguistic typological account using these terms. On the one hand, it has been observed that in certain languages “the structural relation [NP, S] can be used to express not only the

3. Translations of Brassai’s text were borrowed from É. Kiss (2008: 28, 30).

4. The functional notions were first defined by linguists of the Prague School. However, these functions were not yet given the names *Topic* and *Focus*. For more on these issues see e.g. Aronoff and Rees-Miller (2017)

function ‘grammatical subject’, associated with the most prominent theta role, but, alternatively, the discourse-semantic function ‘topic’, as well” (É. Kiss 1995: 3). On the other hand, “many languages have a designated structural position for focus, akin to the WH-position of the English sentence” (É. Kiss 1995: 4). In order to capture these cross-linguistic observations, the term *discourse configurationality* was coined. A language is discourse configurational if either one or both of the following two independent descriptions are valid for the given language: (i) the function Topic “is expressed through a particular structural relation”, (ii) the function Focus “is expressed through a particular structural relation” (É. Kiss 1995: 6). Since the current study concentrates on Focus, this notion and the corresponding “structural relation” will be discussed in the following section in more detail.

1.3 The structures investigated in the present study

The two structures investigated in the present paper are (1c) and (1h) repeated as (2a) and (2b), respectively, for convenience.

- (2) a. A kutya meg-harapta a postást.
 the dog-NOM-SG VM-bite-PST.3SG the postman-SG-ACC
 ‘The dog bit the postman.’
- b. A kutya a postást harapta meg.
 ‘It was the postman that was bitten by the dog.’⁵

The most crucial difference between the two structures is the element that appears immediately pre-verbally.

In (2a) it is the VM, *meg*, that sits in the pre-verbal position, and it is incorporated into the verb (V) forming a phonological word with it (É. Kiss 2002). The Noun Phrase (NP), *a kutya*, preceding the VM-V compound functions as a Topic, while the post-verbal NP, *a postást*, is commonly referred to as *information focus* (see e.g. É. Kiss 1998) or *post-verbal focus* (see e.g. Surányi 2011). The different terms in the literature denoting this word order variant are all indicative of the aspect central to the given analysis. In order to remain theory-neutral, the term ‘neutral sentence’ will be used in the present study to refer to structures exemplified in (2a).

In (2b) the VM-V order is reversed, and the NP which is post-verbal in (2a) occupies the immediately pre-verbal position. Accordingly, its status is changed: the immediately pre-verbal NP functions as a Focus. A number of different terms have been used to denote this sentence type, as well (e.g. *identification focus* in É. Kiss 1998, *contrastive focus* in Kenesei 2006), but, in the current study the

5. PreVf sentences are translated into English as clefts based on É. Kiss (1998).

theory-neutral ‘pre-verbal focus’ (henceforth preVf) will be used. Since the current paper aims to investigate the contextual factors that potentially contribute to the use of preVf and neutral sentences, we first review some of the definitions of Focus presented in previous literature, and then we enumerate the contextual requirements for preVf as suggested by theoretical accounts.

1.4 What is Focus?

In modern linguistics, the term Focus has been defined in a number of ways. One of the most often cited earliest definitions is given by Halliday (1967). Halliday (1967: 204) claims that “focus reflects the speaker’s decision as to where the main burden of the message lies.” It is associated with “prominence” and it is a kind of emphasis that helps the speaker mark the most informative part of their message. Halliday adds that “what is focal is ‘new’ information; not in the sense that it cannot have been previously mentioned, [...], but in the sense that the speaker presents it as not being recoverable from the preceding context.” Clearly, Halliday’s approach is a directly functional one: the author identifies functions of focusing which he claims to be the most important. Nevertheless, it must be noted, that Halliday’s treatment is strictly specific to the English language.

Rooth (1985, 1992), and Krifka (2008) using the Alternative Semantics framework and the notion of sets take a more general stance, and as we will see, create a theoretical framework that can be applied cross-linguistically: “Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions” (Krifka 2008: 247). The novelty of the Alternative Semantics approach lies in the fact that it attempts to formulate the most central and cross-linguistically valid function of Focus. To achieve this, the theory has three additional tenets. First, the description does not specify what form a Focus containing construction has to take. It only indicates that focus has to be marked somehow. Consequently, it allows for the variety of different focus marking strategies observed in different languages.⁶ Second, the description does not specify the type of linguistic expression⁷ that can be focused. Third, it does not specify how “different ways of focus marking signal different ways of how alternatives are exploited” (Krifka 2008: 248). That is, the indication of the presence of alternatives can be used to express different aspects of propositional content within or across languages.⁸ In

6. The most typical Focus marking strategies are marking through intonation, word order, particles, clitics etc. (see e.g. Miller 2006).

7. E.g. phrase, sentence, suffix etc.

8. For an explanation see Example (3) and (4) and the corresponding discussion.

sum, the alternative semantics approach “allows for languages to differ in the ways they mark focus and in the specific interpretational effects of focus” (Krifka 2008: 248). In the present study, Rooth’s and Krifka’s approach is adopted. Therefore, we take it that preVf indicates the presence of alternatives. The ways the alternatives indicated by Hungarian preVf are “exploited” will be discussed in the forthcoming sections.

1.5 Contextual factors commonly associated with preVf and neutral sentences

According to theorists, there are two main ways in which the indication of the presence of alternatives is exploited in Hungarian: preVf may express identification or contrast. These two functions will be discussed in the following two subsections.

1.5.1 *Identification*

É. Kiss (1998) defined preVf as “identificational focus” claiming that its function is to represent “a subset of the set of contextually or situationally given elements for which the predicate phrase can potentially hold; it is identified as the exhaustive⁹ subset of this set for which the predicate phrase actually holds” (p. 245). The identificational function of the immediately pre-verbal slot was already recognized by Brassai (1860), who claimed that the very first element in the bulk (i.e. the part following the optional inchoativum) is the most prominent, most emphatic one “which we assume the speaker or listener would ask about” (translation by É. Kiss 2008: 33). Recall, that Brassai (1860) put forward the idea that sentences serve as building blocks of texts or dialogues. With respect to dialogues, it is reasonable to assume that answers given to wh-questions will be felicitous if they identify the entity which corresponds to the wh-element in the question. This identification is achieved through focusing the expression which denotes the entity at hand. Indeed, question – answer congruence has been a test for Focus not only for Hungarian but cross-linguistically, as well (Rooth 1992, Krifka 2001). Rooth (1992), for example claims that “there is a correlation between questions and the position of focus in answers” (p. 9). Consider the sentence pairs in (3) and (4).

- (3) a. Who owns a gorilla? [x owns a gorilla]
 b. [John]_F owns a gorilla.
 c. *John owns a [gorilla]_F.

9. Although exhaustive interpretation is an important aspect of preVf well studied both in the theoretical and experimental literature, this issue will not be directly addressed in the present paper. The question of exhaustive interpretation will be addressed in Section 3.

- (4) a. What does John own? [John owns x]
 b. *[John]_F owns a gorilla.
 c. John owns a [gorilla]_F.

As illustrated by (3a) and (4a), a question introduces a set of alternatives which are related to the potential answers that can be given to them (Hamblin 1973, Rooth 1992, Roberts 1996). (3a) introduces a set of people as potential owners (e.g. [John, Mary, Carl, Susan]), whereas (4a) introduces a set of entities that can be owned (e.g. [a gorilla, a car, a cat, a house]). Focus serves to identify one or more elements of the introduced sets to which the rest of the sentence holds. For this reason, (3c) is an infelicitous answer to (3a), since [gorilla] is not a subset of the set introduced by (3a), and likewise, (4b) is an infelicitous answer to (4a), as [John] is not a subset of the set introduced by (4a). On the contrary, (3b) and (4c) are appropriate answers to their respective questions.

Interestingly, É. Kiss (1998) claims that *wh*-questions “can be answered not only by an identificational focus (i.e. *preVf*) but – less commonly – by a mere information focus (i.e. neutral sentence), depending on whether the answer is intended to be exhaustive” (p. 249).¹⁰ The author reasons that focusing does not take place if the only role of the constituents at hand is the “marking of the novelty of the information they carry”. That is, the referent of the NP is not presented as a member of the contextually available set of alternatives, if it is post-verbal. For this reason, É. Kiss argues that both (5b) and (5c) are compatible with the question in (5a).

- (5) a. Hol jártál a nyáron?
 ‘Where did you go in the summer?’
 b. Jártam Olaszországban.
 go-PST.1SG Italy-in
 ‘I went to Italy (among other places).’
 c. Olaszországban jártam.
 ‘It was Italy where I went.’

According to Surányi (2011) the factor that determines whether a *preVf* or a neutral sentence is used as an answer to a *wh*-question is a function of the “Question Under Discussion” in the sense of Roberts (1996). In Surányi’s account both *preVf* and the neutral sentence can represent a choice between contextually available set members; the neutral sentence is merely a subcase of *preVf*. According to the author, the choice between the *preVf* and neutral sentence types in answers

10. For a similar reasoning see also Brody & Szendrői (2010).

is determined by two mutually exclusive imperatives assumed by the hearer: (i) “Mention all”, or (ii) “Mention some!” If the former is applicable, preVf is used, if the latter, then the neutral sentence type wins. Surányi’s (2011) account thus allows both sentence types as felicitous answers to wh-questions suggesting that the choice is a pragmatic one. In sum, according to findings in the theoretical literature, the question types in (3a), (4a) and (5a) primarily trigger answers of the preVf word order, where the focused element corresponds to the wh-expression in the question.

The question now arises: what type of questions trigger the use of neutral sentences? According to Kenesei (2006) and Skopeteas and Fanselow (2011), answers given to questions of the type in (6a) will be neutral sentences.

- (6) a. Miért örülnek az emberek a hídon?
‘Why are the people on the bridge happy?’
- b. (Mert) Matyi fogott egy halat. proposition: [p]
(because) Mathias.NOM catch-PST.3SG a fish.ACC
‘(Because) Mathias has caught a fish.’
- c. (Mert) megérkezett a postás. proposition: [q]
(because) arrive-PST.3SG the postman.NOM
‘(Because) the postman has arrived.’
- d. (Mert) kisütött a nap. proposition: [r]
(because) start-to-shine-PST.3SG the sun.NOM
‘Because the sun started to shine.’

The sentence type exemplified in (6b), (6c) and (6d) is also called broad focus or sentential focus, since now it is the whole sentence (or rather the proposition denoted by it) that gets focused. For ease of reference, this question type (sometimes also called broad-wh-question) will be referred to as a sentential focus triggering question (SFT-question, henceforth). The SFT-question in (6a) evokes a set of alternatives at the propositional level: the potential answers to (6a) might be the propositions p, q, r, as indicated. Since now the focus is placed on the clauses corresponding to the proposition, the word order will be such that it expresses information focus as in É. Kiss (1998); or in our terminology, it will be a neutral sentence.

Regarding word order variants as potential responses to the question type at hand, Kenesei (2006) makes an interesting point. Consider example (7) (Kenesei 2006: 18).

- (7) a. What’s new?
- b. [Jelcin]_F nyerte meg az orosz választásokat.
Jeltsin-NOM win-PST.3.SG VM the Russian election-PL.ACC
‘Yeltsin has won the Russian elections.’

- c. Jelcin meg-nyerte az orosz választásokat.
'Jeltsin has won the Russian elections.'

According to Kenesei (2006), the answers in (7b) and (7c) are both congruent with the question in (7a). The choice is extra-linguistic, and depends on speaker expectations about the beliefs of the hearer: if the speaker assumes that the hearer is aware of the candidates for presidency (i.e. assumes a set of alternatives present in the universe of discourse), then (s)he will use a preVf sentence as an answer to identify the element of the implicitly assumed set. If, however, the speaker assumes that the hearer knows nothing about the options, broad focus (i.e. a neutral sentence) will be used to designate a proposition among the potential alternatives (e.g. [there was an earthquake], [it is snowing in the mountains], [there will be a pay raise for linguists] etc.). In sum, if the speaker assumes that the hearer is not aware of a set whose elements are smaller than a clause, a neutral sentence will be used as an answer to questions like those exemplified in (6a) and (7a). Otherwise, it is highly likely that the speaker will use a preVf sentence.

The second way preVf exploits the marking of the presence of alternatives is the expression of contrast, to which we now turn.

1.5.2 Contrast

Contrast is commonly associated with focus (see e.g. Rooth 1992; Krifka 1992; É. Kiss 1998; Kenesei 2006; Zimmerman 2008; Destruel & Velleman 2014). É. Kiss (1998), for example claims that “the identificational foci of different languages are specified for the positive value of either or both of the features [\pm exhaustive] and [\pm contrastive]” (p. 267). However, what is categorized as the contrastive use or function of focus is highly variable depending on the frame of analysis. Krifka (2008),¹¹ for example, takes a strict stance on contrastivity in the case of focus: the author claims that contrast is only present if the alternatives are directly mentioned and contrasted in a corrective or additive way. Consider Example (8), (Krifka 2008: 259) in which the alternatives are names for people, and the contrast is made explicit: the proposition [x wants coffee] is not only true for John, but for Sue, as well.

- (8) a. A: [John] wants coffee.
b. B: [Sue] wants coffee, too.

Krifka (2008) adds that the contrastive use of focus is instantiated through certain grammatical and prosodic features (see e.g. Selkirk 2002; Molnár 2002,

11. For a similar account of contrastive focusing see Selkirk (1984).

Gussenhoven 2004). Hence, contrastive focus should be considered as a separate case of focusing.

In the present paper, Rooth's alternative semantics approach is adopted as intuitively formulated in Section 1.4. Since, according to this formulation, focus marks the presence of alternatives, its contrastive function is directly derivable: if focus serves to identify elements in a contextually available set, it follows, that the elements that are not identified by the focus (i.e. the complementary set) can be or will be contrasted. Indeed, according to É. Kiss (1998), "an identificational focus (i.e. preVf, in our terminology) is [+ contrastive] if it operates on a closed set of entities whose members are known to the participants of the discourse" (p. 267). Kenesei (2006) goes further. According to the author, both preVf and neutral sentences perform identification on a given set; the difference is the type of set identified with relation to the contextually available set of all alternatives. While a neutral sentence identifies a subset, preVf identifies a *proper* subset. If a proper subset is identified, it naturally follows that there will be at least one entity to which the proposition of the focus-containing sentence does not hold. Consequently, Kenesei (2006) terms preVf "*contrastive focus*". In sum, if a set is contextually available (implicitly or explicitly), and an utterance is made regarding this set, it is highly likely that a speaker would use a preVf sentence. In this case, the use of preVf may be contrastive. Otherwise a neutral sentence is used.

1.6 The goal of the present study, hypotheses

The goal of the present study is to examine contextual factors that facilitate the use of preVf sentences. More specifically, we are interested in two of the main factors identified in the theoretical literature: the presence of (i) identification and (ii) contrast.

Regarding identification, we hypothesize that in a context where identification is carried out the preVf sentence type is preferred. If the context does not support identification, either a preVf or a neutral sentence is used.

Regarding contrast, it is hypothesized that the presence of contrast facilitates the use of preVf sentences. If no contrast is present, a neutral sentence is preferred.

2. Experiments

In order to test the hypotheses formulated above, we carried out two guided production studies in the form of an online survey. The details of these experiments are outlined in the following sections.

2.1 Experiment 1

2.1.1 Participants

68 native adult Hungarians (58 female, 10 male, mean age: 50.32, SD = 10.48) participated in the first experiment. Participants were recruited through a Facebook page available for this purpose. Since the experiment was carried out online, data was gathered from speakers from different regions of Hungary.

2.1.2 Procedure

The experiment was an online survey containing a series of trials. Each trial consisted of a panel containing a context text in the upper part of the screen (either one or two sentences depending on the experiment or condition (e.g. *Mit tört el Márty?* ‘What did Márty break?’ in Figure 1.), a cloud of words in the middle of the screen (e.g. *törte* ‘break’, *Márty* ‘Martha’, *egy ujját* ‘a finger’, *el* ‘VM’ in Figure 1), and a blank field between the two.



Figure 1. Example panel of the survey

The words in the cloud were randomly arranged in an area under the blank field. The task of the participant was to produce sentences using the words in the cloud so that the resulting sentence is a natural continuation of the context text. Participants could produce a sentence by clicking on the words: once a word was selected by clicking, it appeared in the field above the cloud. Participants were asked to use all the words in the cloud. Once the sentence was built, the next trial was initiated by a click on a button at the bottom of the screen. Before proceeding to the next

trial, participants were encouraged to reread the context text and the sentence together to make sure that their solution was natural and coherent. Responses could be modified as needed.

The experiment contained two blocks: a practice block following the instructions, and the test block. The practice block contained 5 trials similar to those in the test block. The test block contained three types of trials: (i) test, (ii) filler and (iii) hidden control trials (see details for (i) and (ii) in sections 2.1.3 and 2.2.3). In the hidden control trials, participants were given an instruction to put the words in the cloud in a given order; one that would have been very unlikely to be produced by chance. The inclusion of these trials was a practical one. In order to encourage participation, respondents were entered a draw in which a winner would be picked randomly at the end of data collection. The prize was a small fee; a voucher worth of 5000 HUF sent via post to the winner. The hidden control trials were included so that we could filter out respondents who participated only in the hope of winning the voucher: erroneous responses in the hidden control trial indicate that the participant did not read the instruction in these trials based on which it can be assumed that they did not read the context sentences in the rest of the trials either.

The experimental interface was developed by relying on the d3-cloud JavaScript package (Davies 2013) and was implemented in the in the IBEX online experimental environment (Drummond 2010).

2.1.3 *Materials*

Experiment 1 investigated whether identificatory and contrastive contexts have the effect of facilitating the use of preVf word order sentences. For the two different context types, two different test trial types were created.

To test the effect of identification, questions were introduced, as shown in (9). For contexts of identification (identification-condition), we used directive wh-questions (or narrow-wh-questions) (9a); and for contexts where identification was not present (no-identification-condition), SFT-questions (or broad-wh-questions) (9b) were used.

- (9) a. Mit vesztett el Péter?
‘What did Peter lose?’
b. Miért szomorú Péter?
‘Why is Peter sad?’

The set of words comprising the cloud for the example in (9) is given in (10a) separated by commas. The elements of the word cloud were the same types of linguistic units in each critical trial: (i) a name in the nominative case, (ii) an NP comprising of an indefinite article and a noun in the accusative case, (iii) a verb in past tense and (iv) a corresponding VM. (10b) and (10c) are two example sentences that can be produced by ordering the words in the word cloud for the contexts in (9). The

order and location of the elements in the cloud was randomly allocated in each trial.

- (10) a. Péter, egy órát, vesztett, el
Peter-NOM, a watch-SG.ACC, lose-PST.3SG, VM
'Peter', 'a watch', 'lost'
- b. Péter egy órát vesztett el.
'Peter lost [a watch]_F.'
- c. Péter el-vesztett egy órát.
'Peter lost a watch.'

To test the effects of contrast, two-sentence contexts were created, as in (11). The first sentence designated an explicit set (11a) in both conditions. The experimental manipulation was introduced in the second sentence. In the contrast-condition, the second sentence identified one element of the designated set (11b). To avoid a syntactic priming effect, these sentences had a structure that could not be built with the words in the cloud: the NP referring to the identified element was in the nominative case, and the VP did not contain a transitive verb. In the no-contrast-condition, the second sentence did not identify any member of the contextually given set, but it always referred to some state of affairs that was irrelevant with respect to that set (11c). The structure of these sentences also differed from focus constructions for the aforementioned reasons.

- (11) a. Ebéd után maradt desszertnek egy körte, egy alma és egy barack.
'After lunch there was a pear, an apple and an apricot for dessert.'
- b. János desszertje az alma volt.
John-NOM dessert-his-NOM.SG the apple-NOM.SG be-PST.3SG
'John had the apple for dessert.'
- c. János desszert helyett szundított
John-NOM dessert-NOM.SG instead take-a- nap-PST.3SG
inkább kicsit.
rather little.ACC
'John didn't have dessert, he took a nap instead.'

The set of words comprising the cloud for the example in (11) is given in (12a) separated by commas. Just as in the previous case, the elements in the word cloud were of the same type in each trial: (i) a name in the nominative case, (ii) an NP consisting of a definite article and a noun in the accusative case, (iii) a verb in past tense and (iv) a VM. The two canonical word order variants that can be produced using elements listed in (12a) are provided in (12b) and (12c).

- (12) a. Péter, a körtét, ette, meg
Peter-NOM, the pear-SG.ACC, eat-PST.3SG, VM
'Peter', 'the pear', 'ate'

- b. Péter a körtét ette meg.
'Peter ate [the pear]_F'
- c. Péter megette a körtét.
'Peter ate the pear.'

For the investigation of the possible effects of both identification and contrast, 16 critical trials were presented. The conditions were distributed equally: 8 trials in the identification condition, 8 trials in the no-identification condition, 8 trials in the contrast condition, and 8 trials in the no-contrast condition. This amounted to a total of 32 critical trials.

The experiment also contained 64 filler trials. In order to “conceal” the actual purpose of the experiment, we identified four key features for fillers. These features were the following: (i) whether the trial contains a one- or two-sentence context, (ii) the presence or absence of a question in the context, (iii) whether the sentence that fits the context (i.e., the one that the participant produces) should contain a separated or a non-separated VM-V complex, and (iv) the presence or absence of a VM in the sentence to be produced. The features listed were selected to correspond to all the manipulated variables (feature (i), (ii)), and the outcome variable (feature (iii)) in the experiment. Feature (iii) was especially important, since we wanted to encourage participants to regard VMs as elements that could appear in any potential position within the sentence. Additionally, feature (iv) was crucial to eliminate the potential impression that the experiment had anything to do with VMs. The distribution of features in filler trials is presented in Table 1.

Table 1. The distribution of key features in filler trials.

Feature			Number
2-sent. context	Question	VM – V separ.	4
		VM – V non-separ.	4
	No question	VM – V separ.	4
		VM – V non-separ.	4
3-sent. context	Question	VM – V separ.	4
		VM – V non-separ.	4
	No question	VM – V separ.	4
		VM – V non-separ.	4
2-sent. context	Question	No VM	8
	No question		8
3-sent. context	Question		8
	No question		8

2.1.4 Predictions

In line with our hypotheses in Section 1.6, we predicted that in the case of wh-question types (9a), participants would produce preVf word order sentences (10b), whereas in the case of SFT-questions (9b), preVf (10b) and neutral sentences (10c) would be produced.

Regarding the effect of contrast, it was predicted that in the contrast-condition dominantly preVf sentences would be produced, whereas in the no-contrast-condition we expected the dominance of neutral sentences.

2.1.5 Results

PreVf response rates were analyzed in a Friedman's ANOVA model with four within-subjects conditions (narrow-wh-question, SFT-question, contrast context and no-contrast context). The results indicate that context has a significant effect on the choice of word order ($X^2 = 91.24$, $p < .001$). Specifically, there is a sharp contrast in preVf response rates between the narrow-wh ($M = .83$, $SD = .17$) and SFT ($M = .17$, $SD = .15$) question conditions ($z = 7.20$, $p < .001$) and between the contrast ($M = .92$, $SD = .10$) and no-contrast ($M = .63$, $SD = .21$) context conditions ($z = 4.35$, $p < .001$). Further analyses using a series of one-sample Wilcoxon Signed Ranks tests reveal that the median rate of preVf responses significantly differs from chance in all four conditions: in the narrow-wh-question, contrast context and no-contrast context conditions it is above chance level ($z = 5.08$, 5.49 and 3.32 respectively; $p \leq .001$) while in the SFT-question condition it is below chance level ($z = -5.25$, $p < .001$).

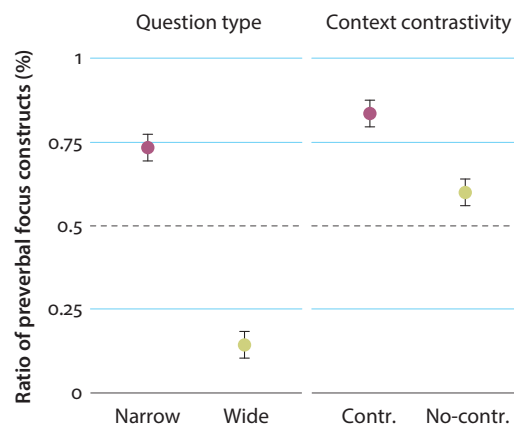


Figure 2. Proportion of sentence types in the two contexts (error bars: 95% CI). The horizontal dashed line represents the level of chance.

2.1.6 Discussion

The data obtained in Experiment 1 partly corroborated our hypotheses. The high proportion of preVf sentences produced in both contexts of identification and contrast are in line with our expectations, and clearly demonstrate that these two contextual factors facilitate the use of the sentence type at hand. The below-chance level rate of preVf sentences in the SFT-condition, and the above-chance level rate of preVf sentences in the no-contrast-condition, however, present an interesting puzzle. Note that in the former case, we predicted that this rate would be around chance, whereas in the latter, we predicted that it would be below chance. Based on the observed patterns, we conjectured that the answer may lie in the availability of sets. In the forthcoming discussion, the issue of sets will be taken up.

Let us return to our starting point regarding the function of focus: focus evokes alternatives (Rooth 1992, Krifka 2008). Using sets to represent what focus does, this formulation suggests that there must be a contextually available set from which a subset is identified by focus, which entails that there is also a complementary set of alternatives. Recall that this is exactly what Kenesei (2006) claims: focus identifies a *proper subset* of the contextually available set. More importantly, Kenesei adds that “whether or not the contrasting complementary set is explicit, in case of contrastive focus a complementary set is always created” (p. 137). If this is indeed true, preVf is triggered in contexts that either explicitly or implicitly contain a set whose members are potential referents of the focus of the proposition to be made. This explanation can straightforwardly account for the observation that in all of the conditions which contained a set or in which a set was evoked, the proportion of preVf structures was above chance level. Let us consider the conditions used in Experiment 1 from the point of view of sets. In contexts with a narrow-wh-question a set is evoked. Consider (3a) and (4b) again: the former evoked a set of potential owners ([x owns a gorilla]), whereas the latter evoked a set of entities that can be owned ([John owns x]). In the case of SFT-questions no such set is evoked. Correspondingly, the observed rate of preVf sentences was well below chance level, and participants reliably produced neutral sentences. In the contrast- and no-contrast-conditions the set was not evoked; it was explicit, since the first sentence always designated a three-element set in the context. Note that apart from the SFT-question-condition, in which no set was available, the proportion of preVf sentences was above chance level in all conditions.

Considering the above reasoning, an important factor contributing to the production of preVf is the availability of sets. Thus, we hypothesized that if a set is either explicitly or implicitly present, a preVf sentence is used. Otherwise either a preVf or a neutral sentence is produced (cf. Examples (7b) and (7c) in Section 1.5.1). In order to test this hypothesis, we designed and ran Experiment 2.

2.2 Experiment 2

2.2.1 Participants

118 native adult Hungarians participated in the experiment (99 female, 19 male, mean age: 48.37, SD = 12.43). Participants were recruited through a Facebook page dedicated to obtaining participants. Since the experiment was carried out online, data was gathered from speakers from different regions of Hungary.

2.2.2 Procedure

The procedure of Experiment 2 was identical that of Experiment 1. For details see Section 2.1.2.

2.2.3 Materials

In order to test the potential effect of the contextual availability of sets, two-sentence contexts were created. The first sentence in the context either (i) designated a set explicitly (Table 2, sentence a), or (ii) implicitly (Table 2, sentence b), or (iii) designated no set at all (Table 2 expression c). In the case of (i), the set was made explicit through an enumeration, while in the case of (ii), the implicit set was implied. In case (iii), we used a label for a particular situation or place irrelevant with respect to the potential sets for the third (i.e. test) sentences. The second sentences were such that they either introduced a contrast with respect to the third sentence (Table 2, sentence d), or not (Table 2, sentence e).

The combination of sentence types in Table 2 resulted in altogether six conditions. These are presented in Table 3.

Example (13) presents the elements of the cloud for the example trial in Table 2 separated by commas (13a), and two potential word order variants that could be produced using these words: a preVf sentence (13b) and a neutral one (13c). The elements of the word cloud were the same type of linguistic units in each critical trial: (i) a name in the nominative case, (ii) an NP comprising of an indefinite article and a noun in the accusative case, (iii) a verb in past tense, and (iv) a corresponding VM.

- (13) a. Misi, egy várat, rakott, össze
 Mike-NOM, a castle-SG.ACC, put-PST.3SG, VM (~together)
 ‘Mike’, ‘a castle’, ‘built’, ‘up’
- b. Misi egy várat rakott össze.
 ‘Mike built [a castle]_F’
- c. Misi össze-rakott egy várat.
 ‘Mike built a castle.’

Table 2. Sentences of the contexts used in Experiment 2

Label	Number in context	Condition	Example
a.	1	Expl. set	A gyerekeknek kiosztottak egy csomó játékkockát. Ezekből mindenfélét össze lehetett rakni: például tornyot, várat, házat. 'Kids were handed out lots of building blocks. So many things could be built out of them: for example a tower, a castle, a house.'
b.	1	Impl. set	A gyerekeknek kiosztottak egy csomó játékkockát, amikből mindenfélét össze lehetett rakni 'Kids were handed out lots of building blocks. So many things could be built out of them.'
c.	1	No set	A játéksarokban: 'In the playing corner.'
d.	2	Contrast	Bence már korábban is sok tornyot épített, így most is ezt csinálta. 'Ben had built a lot of towers before, so now he did so again.'
e.	2	No contrast	Bence álmos volt, így csak nézte, ahogy a többiek játszanak. 'Ben was sleepy, so he just watched the others play.'

Table 3. Conditions in Experiment 2

Experimental manipulation introduced in		
First sentence		Second sentence
explicit set	–	contrast
explicit set	–	no-contrast
implicit set	–	contrast
implicit set	–	no-contrast
no set	–	contrast
no set	–	no-contrast

In Experiment 2, the availability of sets (explicit, implicit or no set) was a between-subject factor, and the presence or absence of contrast was a within-subject factor. Each run of the experiment contained 12 test trials (6 in the contrast condition and 6 in the no-contrast condition) and 24 filler trials. Respondents were assigned to one of the three versions of the experiment randomly. As in Experiment 1, three types of fillers were used in equal proportion: the third sentence (i) had to be produced in a way that it contained a VM-V compound, (ii) the VM and V had to be

separated, or (iii) the sentence did not contain a VM at all. Also as in Experiment 1, this variation was introduced to encourage participants to regard VMs as elements that could appear in any potential position within the sentence.

2.2.4 Predictions

In line with our hypothesis formulated at the end of Section 2.1.6, we predicted the following. If contrast is present, dominantly preVf sentences will be produced. The availability of a set (either explicit or implicit) will facilitate the production of preVf sentences even if no contrast is present. If neither a set is available nor contrast is to be expressed, preVf sentences will not be produced reliably.

2.2.5 Results

The results were in line with our predictions. A Kruskal-Wallis test for the contrast-condition and one for the no-contrast-condition revealed that the presence or absence of an explicit or implicit set had no significant effect on word order choices in the contrast context condition ($H(118) = 2.63, p = .27$), while in the no-contrast context condition it did ($H(118) = 13.35, p = .001$). Specifically, in the no-contrast condition, preVf was less likely to be used when no set was specified ($M = .49, SD = .26$) than when a set was either explicitly given ($M = .69, SD = .24; z = 3.67, p = .001$), or implicitly given ($M = .62, SD = .24; z = 2.54, p = .05$). There was no significant difference between the explicit and implicit conditions. Wilcoxon Signed Rank tests comparing the medians to chance level revealed that in the presence of a contrast context, the probability of choice of preVf in all three set conditions (explicit, implicit, none) was significantly higher than chance ($z = 5.39, 5.27$ and 5.36 ; all $ps < .001$). In the absence of contrast, participants produced a preVf word order with

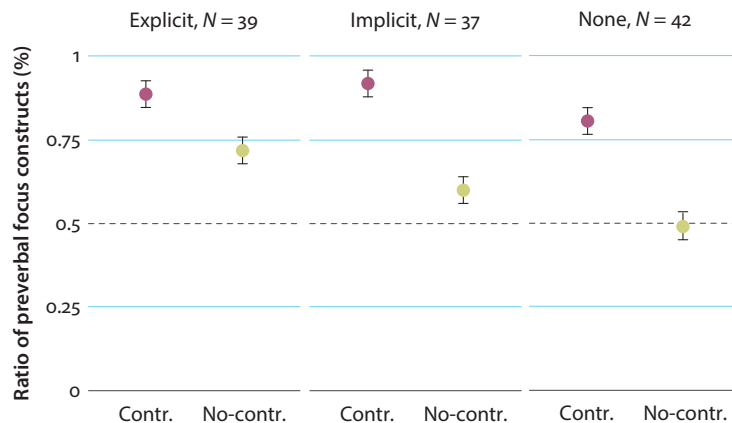


Figure 3. Proportion of sentence types in the three contexts (error bars: 95% CI). The horizontal dashed line represents the level of chance

higher than chance probability when an explicit or implicit set was given (explicit: $z = 3.64$, $p < .001$; implicit: $z = 2.72$, $p = .006$); but the choice between preVf and neutral word order did not differ from chance when no set was given.

2.2.6 Discussion

Data obtained in Experiment 2 corroborated our hypotheses: the availability of sets facilitates the production of preVf sentences. The facilitation effect is present irrespective of the expression of contrast for both explicit and implicit sets. However, in the case of implicit sets, this facilitation effect diminishes if no contrast is to be expressed. Altogether, contrast also facilitates the production of preVf sentences. Additionally, lack of contrast makes the production of preVf sentences less likely irrespective of the type (i.e. explicitness or implicitness) of the contextually available set. However, in the case of contrast, preVf is still reliably produced relative to chance level even if no set is available in the immediately preceding context. If both a contextually available set and contrast are missing, preVf is not produced reliably.

In sum, two important and intimately related factors have been identified which facilitate the production of preVf sentences: (i) the availability of a set and (ii) the expression of contrast.

3. General discussion

The present study investigated the factors that potentially facilitate the use of sentences in the preVf word order. Based on findings in the theoretical literature, it was hypothesized that two of these factors are the expression of (i) identification and (ii) contrast.

These factors were examined in a guided production study: in the experimental trials, participants were presented a context text and a cloud of words. The task was to produce sentences using the words in the cloud in a way that the newly constructed sentences fit the context and thus create a coherent text.

The results of the experiment showed that the aforementioned factors facilitate the use of preVf. However, it was also observed that even when no contrast was to be expressed, participants still reliably produced preVf sentences at a rate above chance level. In the case of SFT-questions, however, the rate of preVf sentences was reliably under chance level. The authors of the present study believe that the formulation of the function of focus as outlined in Section 1.4 is the most appropriate one to account for the observed results. This formulation states that focus is a means of indicating the presence of alternatives. Furthermore, alternatives can only be marked if there is a contextually available set (see e.g. Kenesei

2006). In terms of sets, focus creates a distinction: it designates a proper subset of the contextually available set, and consequently, creates a complementary set. The former is an instance of identification; the latter is an instance of contrast. Note that a set was present in all but one conditions in Experiment 1. In the identification-condition, the narrow-wh-question created an implicit set of alternatives for the NP (cf. Examples (3) and (4)). In the contrast and no-contrast conditions, the first sentence of the contexts always contained an explicit set. In the SFT-question-condition, however, in which a below chance level rate of preVf sentences was produced, neither an implicit nor an explicit set was present. For this reason, we conjectured that the availability of sets may be an underlying factor facilitating preVf use. In order to test this conjecture, a second experiment was run.

Experiment 2 tested conditions in which a set was either explicitly or implicitly present as well as conditions that referred to no set at all. In all cases, there were trials which contained contrast and trials which did not. It was predicted that in conditions that contained a set, preVf would be used. Likewise, if contrast is to be expressed, the rate of preVf sentences would be increased. The results of Experiment 2 were in line with the predictions. PreVf sentences were reliably produced in conditions where a set was present, irrespective of whether the set was explicit or implicit. Also, contrast in general facilitated the rate of preVf sentences with respect to the no-contrast conditions. The only conditions in which the rate of preVf sentences was around chance level was the no-set-no-contrast-condition. Interestingly, we also observed the dominance of preVf sentences in the no-set-contrast-conditions. This result deserves a mention. We believe that the high proportion of preVf sentences can be accounted for in terms of sets, as well. Consider the context for one such trial as shown in (14a).

- (14) a. IN THE PLAYING CORNER:
Ben had built a lot of towers, so he decided to build one now, as well.
b. [Mike built x]
c. [a castle, a house, a bridge, etc]

Although the context does not designate a set either explicitly or implicitly as in the other two conditions, a set is still evoked with respect to the second sentence if contrast is to be expressed. The mechanism is analogous to the one where identification is triggered via a wh-question (cf. example (3) and (4)) as indicated in (14b) and (14c): by the contrastive use of focus a set of alternatives with respect to what can be built out of building blocks is introduced or evoked.

To conclude, the results of Experiment 1 and 2 jointly support the view that there is a general underlying factor determining the use of preVf sentences, namely, the contextual availability of a relevant set. The uses of preVf examined in the current paper are derivable from this factor.

Another, independent observation may account for the dominating proportions of preVf sentences in the conditions containing sets in the present experiments. Káldi and Babarczy (2017) and Káldi and Babarczy (2018), in a series of visual world experiments, presented preVf, neutral and lexically marked focus sentences in contexts containing either explicit or implicit sets, and studied the rate of exhaustive interpretations of the presented sentence types. It was observed that in the case of preVf sentences, the rates of exhaustive interpretation were around 90% (where a chance level was 50%) in all conditions. This is a high ratio compared to those found in the literature where the rate of exhaustive reading was measured in out-of-context preVf and neutral sentences. For example, an earlier study (Káldi, Babarczy & Bende-Farkas 2016) found that the rate of exhaustive reading was 65% for out-of-context preVf sentences. In sum, the results of the experiments reported in the present paper and of those cited above show that the contextual availability of sets matters for two reasons. On the one hand the present experiments showed that the availability of a set facilitates preVf use. On the other hand, the results of Káldi and Babarczy (2017) and Káldi and Babarczy (2018) demonstrate that if this set is either explicitly or implicitly given, the likelihood that preVf is interpreted exhaustively increases. To approach the question from a different perspective, note that according to É. Kiss (1998), preVf sentences have an exhaustive interpretation: for example if an answer to a narrow-wh-question is to be understood exhaustively, a preVf is used. Now, if we postulate that in scenarios where identification and/or contrast is to be expressed, a felicitous utterance has to contain a structure that has exhaustive interpretation (i.e. the speaker obeys the implicit “Mention all!” imperative), it follows that this structure will be a preVf sentence. Apparently, the availability of sets and the typically exhaustive interpretation of preVf are in an intricate relationship. More experimental work is needed, however, to clarify the mechanisms through which these factors interact.

4. Conclusion

The results of the two experiments in the present paper showed that two important functions of the preVf word order sentences are identification and contrast. A crucial factor underlying both of these functions is the availability of sets. Sets can be present in a number of ways: they can either be explicitly mentioned (e.g. by an enumeration), implicitly referred to (e.g. by a category name) or evoked (e.g. by a question or contrast). If a set is present and one or more of its elements are to be identified or contrasted, a preVf sentence is most likely used. Furthermore, it is highly likely that the availability of sets is a critical factor contributing to the typically exhaustive interpretation of preVf sentences.

Acknowledgement

The research was supported by NKFI 115544. Our thanks go to Balázs Surányi for his precious help.

References

- Aronoff, Mark & Janie Rees-Miller (eds.). 2017. *The handbook of linguistics*. Wiley, Blackwell. <https://doi.org/10.1002/9781119072256>
- Brassai, Sámuel. 1852. Tapogatózások a magyar nyelv körül [Exploring the Hungarian Language]. *Pesti Napló* 802.
- Brassai, Sámuel. 1860. A magyar mondat [The Hungarian Sentence]. *Magyar Akadémiai Értesítő. A széptudományi Osztály közlönye* 1. 279–399.
- Brassai, Sámuel. 1885. A mondat dualismusa [The dualism of the sentence]. *Értekezések a Magyar Tudományos Akadémia Nyelv- és Széptudományi Osztály Köréből* 12.
- Brody, Mihály & Kriszta Szendrői. 2010. Exhaustive focus is an answer. Unpublished Manuscript. <http://ling.auf.net/lingbuzz/001113>.
- Davies, Jason. 2013. d3-cloud. <https://github.com/jasondavies/d3-cloud>
- Destruel, Emilie J. & Leah Velleman. 2014. Refining contrast: Empirical evidence from the English it-cleft. *Empirical Issues in Syntax and Semantics* 10. 197–214.
- Drummond, Alex. 2010. Ibex (Internet Based Experiments). <http://spellout.net/ibexfarm/>
- É. Kiss, Katalin. 1995. *Discourse configurational languages*. Oxford: Oxford University Press.
- É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 72(2). 245–273. <https://doi.org/10.1353/lan.1998.0211>
- É. Kiss, Katalin. 2002. *The syntax of Hungarian*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511755088>
- É. Kiss, Katalin. 2008. A pioneering theory of information structure. *Acta Linguistica Hungarica* 55. 23–40. <https://doi.org/10.1556/ALing.55.2008.1-2.2>
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511616983>
- Halliday, Michael. 1967. Notes on transitivity and theme in English: Part 2. *Journal of linguistics* 3(2). 199–244. <https://doi.org/10.1017/S0022226700016613>
- Hamblin, Charles L. 1973. Questions in montague English. *Foundations of language* 10(1). 41–53.
- Káldi, Tamás, Anna Babarczy & Ágnes Bende-Farkas. 2016. Hungarian focus: Presuppositional content and exhaustivity revisited Language and Linguistic structure. In Joseph Edmonds, Michaela Martinková & Marketa Janebova (eds.), *Proceedings of the Olomouc Linguistics Colloquium 2016*. Olomouc: Palacký University Olomouc.
- Káldi, Tamás & Anna Babarczy. 2017. A kontextus hatása a magyar preverbális fókusz értelmezésére: egy szemmozgás-követéses vizsgálat [The effect of context on the interpretation of the Hungarian pre-verbal focus: an eye-tracking study] In Zoltán Bánréti & István Kenesei (eds.), *Általános Nyelvészeti Tanulmányok. XXIX*, 99–126. Budapest: Akadémiai Kiadó.
- Káldi, Tamás & Anna Babarczy. 2018. Linguistic exhaustivity inference is context dependent: A visual-world eye-tracking study on Hungarian focus. *Acta Linguistica Academica* 65(4). 547–595. <https://doi.org/10.1556/2062.2018.65.4.2>

- Kálmán, László. 1985. Word order in neutral sentences. In István Kenesei (ed.), *Approaches to Hungarian 1*, 13–23. Szeged: JATE.
- Kenesei, István. 2006. Focus as identification. In Valéria Molnár & Susanne Winkler (eds.), *The architecture of focus*, 137–168. Mouton de Gruyter. <https://doi.org/10.1515/9783110922011.137>
- Krifka, Manfred. 1992. A framework for focus-sensitive quantification. *Semantics and Linguistic Theory 2*. 215–236. <https://doi.org/10.3765/salt.v2i0.3024>
- Krifka, Manfred. 2001. For a structured meaning account of questions and answers. In Caroline Féry & Wolfgang Strenfeld (eds.), *Audiatur Vox Sapientiae: A Festschrift for Arnim von Stechow* (Studia grammatica, Band 52), 287–319. De Gruyter Akademie Forschung.
- Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica 55*(3–4), 243–276. <https://doi.org/10.1556/ALing.55.2008.3-4.2>
- Miller, Jim. 2006. Focus. In Keith Brown (ed.), *Encyclopedia of language and linguistics* (2nd ed.), 511–518. Elsevier. <https://doi.org/10.1016/B0-08-044854-2/00575-7>
- Molnár, Valéria. 2002. Contrast – from a contrastive perspective. *Language and Computers 39*(1), 174–161.
- Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. In Jae-Hak Yoon & Andreas Katjöl (eds.), *Working Papers in Linguistics-Ohio State University Department of Linguistics 49*, 91–136. Columbus: The Ohio State University Department of Linguistics.
- Rooth, Mats. 1985. *Association with focus*. Ph.D. thesis. Massachusetts Institute of Technology.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics 1*(1). 75–116. <https://doi.org/10.1007/BF02342617>
- Selkirk, Elisabeth. 1984. *Phonology and syntax*. Cambridge: The MIT Press.
- Selkirk, Elisabeth. 2002. Contrastive FOCUS vs. presentational focus: Prosodic evidence from right node raising in English. In Bernard Bel & Isabelle Marlin (eds.), *Speech Prosody 2002: Proceedings of the First International Speech Prosody Conference*, 643–646. Laboratoire Parole et Langage, Université de Provence.
- Skopeteas, Stavros & Gisbert Fanselow. 2011. Focus and the exclusion of alternatives: On the interaction of syntactic structure with pragmatic inference. *Lingua 121*(11). 1693–1706. <https://doi.org/10.1016/j.lingua.2011.05.005>
- Surányi, Balázs. 2011. A szintaktikailag jelöletlen fókusz pragmatikája [The pragmatics of syntactically unmarked focus]. In Huba Bartos (ed.), *Általános Nyelvészeti Tanulmányok 23*, 281–313. Budapest: Akadémiai Kiadó.
- Zimmermann, Malte. 2008. Contrastive focus and emphasis. *Acta Linguistica Hungarica 55*. 347–360. <https://doi.org/10.1556/ALing.55.2008.3-4.9>