

Long split focalization in Hungarian and the typology of A'-dependencies[☆]



Adrienn Jánosi, Jeroen van Craenenbroeck, Guido Vanden Wyngaerd^{*}

KU Leuven, Warmoesberg 26, 1000 Brussels, Belgium

Received 21 November 2012; received in revised form 13 May 2014; accepted 18 July 2014
Available online

Abstract

We show that hitherto unstudied constructions of long split focalization (LSF) in Hungarian provide evidence for a double derivation analysis, in that they can involve either a movement derivation or a base-generation one. These derivations correlate with different patterns of verbal definiteness agreement and case marking. The evidence for the double derivation analysis comes from two sources. First, there is systematic speaker variation in the acceptance of movement derivations: whereas all speakers accept base-generation sentences, only a subgroup of them accepts movement sentences. Second, island phenomena support the analysis: we find that base-generation sentences show no sensitivity to islands, whereas movement ones do.

© 2014 Elsevier B.V. All rights reserved.

Keywords: Long split focalization; Scope-marking; A'-dependencies; Hungarian

1. Introduction

This paper discusses long split focalization (LSF) in Hungarian, a construction in which the NP-part of a focused DP is found in the matrix clause, while the functional layers (including adjectival modifiers) are stranded in the embedded clause. An example is given in (1).¹

- (1) AUTÓT mondott, hogy újat vett.
car.acc said.3sg.indef that new.acc bought.3sg.indef
'He said that he had bought a new CAR.'

This construction has so far not been described or analyzed in the generative literature on Hungarian. This example contrasts with a case of long unsplit focalization (LUF) as in (2):

[☆] The authors would like to thank Anikó Lipták for her comments on an earlier version of this paper, and much constructive discussion, as well as the anonymous reviewers for *Lingua*. All errors are our own.

^{*} Corresponding author. Tel.: +32 26098883.

E-mail addresses: ad.janosi@gmail.com (A. Jánosi), jeroen.vancraenenbroeck@kuleuven.be (J. van Craenenbroeck), guido.vandenwyngaerd@kuleuven.be (G. Vanden Wyngaerd).

¹ Hungarian has a number of related constructions, which we will not discuss here. These are short split focalization, and long and short split topicalisation. For more discussion of these, we refer to Jánosi (2014).

- (2) ÚJ AUTÓT mondott, hogy vett.
 new car.acc said.3sg.indef that bought.3sg.indef
 'He said he had bought a NEW CAR.'

LUF has been studied extensively in the generative literature on Hungarian (e.g. [É. Kiss, 1987](#); [Lipták, 1998a](#); [Lipták, 1998b](#); [Puskás, 2000](#)). Our main goal is to defend a double derivation for LSF, one involving movement and the other base-generation. Each type of derivation can be diagnosed by looking at the patterns of case assignment and definiteness agreement in the matrix clause. We provide two important pieces of evidence for this double derivation analysis. The first involves a split between speakers: whereas all speakers accept the base-generation derivation for LSF, only a subgroup of them accepts the movement derivation. Second, we show that the base-generation derivation is not sensitive to the adjunct island constraint, whereas the movement derivation is. Both the speaker split and the selective island sensitivity are straightforwardly accounted for under the double derivation analysis we propose, whereas these facts are hard to make sense of under alternative proposals. Our double derivation analysis is heavily inspired by a similar proposal made in [Den Dikken \(in press\)](#), who develops a typology of A' -dependencies that also distinguishes movement dependencies from base-generation ones.

The paper is organized as follows. In [section 2](#) we introduce and discuss [Den Dikken's](#) typology of A' -dependencies. [Section 3](#) presents a comparison between split and unsplit A' -dependencies and argues that LSF-examples such as (1) are indeed closely related to their unsplit counterparts in (2). In [section 4](#) we provide an analysis of the LSF-data in terms of a double derivation analysis (movement and base-generation). In [section 5](#) we present data concerning speaker variation and island sensitivity in support of this analysis. These data are based on a questionnaire testing the acceptability of various kinds of focus constructions with 83 native speakers of Hungarian. [Section 6](#) sums up and concludes.

2. Den Dikken's typology of long A' -dependencies

Although long A' -movement, in particular wh-movement and focus movement, has received continuous attention in the generative literature on Hungarian over the past decades ([É. Kiss, 1987](#); [Lipták, 1998a](#); [Lipták, 1998b](#); [Puskás, 2000](#)), it has recently become a highly debated issue. In particular, while all traditional accounts assumed without discussion that long-distance A' -dependencies are derived via movement (e.g. [É. Kiss, 1987, 1998](#); [Marácz, 1987](#); [Horváth, 1998](#); [Lipták, 1998a](#); [Lipták, 1998b](#)), a number of recent analyses argue that in addition to the movement derivation of long focus/wh-constructions, there is also the possibility of deriving such structures by base-generating the 'long-moved' DP in the matrix clause. In this paper we focus on one such account, i.e. that of [Den Dikken \(in press\)](#).²

Building on [Gervain's](#) resumptive prolepsis account, [Den Dikken](#) proposes three different analytical possibilities for long-distance A' -dependencies in Hungarian (and cross-linguistically):

- (3) a. successive-cyclic movement via vP -edges
 b. resumptive prolepsis
 c. scope marking (with no concord, partial concord or full concord)³

Let us discuss and illustrate each of these in turn. The first one is the well-known movement strategy, the only twist being that this movement operation has no stopover in specCP. This is part of [Den Dikken's](#) broader theoretical program, in which specCP only serves as a final landing site, never as an intermediate one. Technically, this is achieved by having the

² [Gervain \(2009\)](#) also defends a double derivation analysis, and in addition ties the two derivations to two different speaker groups, an issue we come back to extensively below. While we take [Den Dikken's](#) analysis as the point of departure for our own, we briefly point out two differences between [Gervain's](#) analysis and our approach to long focus constructions. In her 2009 analysis [Gervain](#) splits speakers into 'movement' speakers and 'base-generation' (i.e. resumptive) speakers, and argues that the two dialects do not overlap, i.e. speakers either use only the base-generation or only the movement strategy. In our experiment, which was conducted with mainly speakers from North-Eastern Hungary, we found that there are informants who can make use of both strategies. Another difference is that what we call long focus constructions 'with a potential of case mismatch' (cf. [section 4.1](#)) are derivable both via long-distance movement and via a resumptive dependency in [Gervain's](#) account. In contrast, such examples can only be derived via a base generation strategy in our analysis. See [Jánosi \(2014\)](#) for a more detailed discussion of [Gervain's](#) account and a comparison with the one defended here.

³ The proposal that in long focus constructions the 'long-focused' constituent occupies a scopal position has been put forward by various authors, e.g. [É. Kiss \(1987, 1998\)](#), [Lipták \(1998a\)](#), [Lipták \(1998b\)](#), [Horváth \(1995, 1998, 2000\)](#). Where [Den Dikken's](#) analysis differs from the ones mentioned above is the introduction of the operation 'concord' in the derivation of the scope-marking construction. See below for discussion.

matrix verb—or more precisely, its *v*—agree with the embedded CP as proposed by Rackowski and Richards (2005). As a result of this Agree-relation, any subsequent (Agree-and-)movement operations are free to ignore the locality boundary imposed by CP (cf. the Principle of Minimal Compliance of Richards, 1998) and move directly to specvP, the next locality boundary. A schematic representation of this type of derivation is given in (4), and an example from Hungarian in (5) (Den Dikken, in press, p. 5).⁴

- (4) XP... [_{VP} ~~XP~~ [_{VP} V [_{VP} V [_{CP} C... [_{VP} ~~XP~~...]]]]]
- (5) ??Hány lány akarod, hogy eljöjjön?
 how.many girl.nom.sg want.2sg.def that pv.come.3sg
 'How many girls do you want to come?'

In order to properly analyze this example, we need to pay close attention to its case and agreement morphology. Note first of all that *hány lány* 'how many girls' bears nominative case, which is the case assigned to the foot of the chain, i.e. the subject position of the embedded clause, rather than the case of the object position of the matrix verb *akarod* 'want'. Den Dikken takes this to be an indication that movement has taken place. Secondly, the matrix verb *akarod* 'want' bears definite agreement, i.e. it shows agreement with a definite object. Given that *hány lány* 'how many girls' is indefinite, this cannot be the trigger for this agreement. Den Dikken argues that the matrix verb—or more precisely, the matrix *v*—agrees with the entire embedded clause, thus voiding the need for the wh-phrase to make an intermediate stopover in specCP (see above). In support of the idea that finite clauses trigger definite agreement, Den Dikken (in press, p. 5) presents the following contrast:

- (6) Szeretek PRO szerelmes lenni.
 love.1sg.indef PRO in.love be.inf
 'I love being in love.'
- (7) Szeretem, hogy szerelmes vagyok.
 love.1sg.def that in.love am.1sg
 'I love that I am in love.'

This pair of examples shows that finite and infinitival complement clauses behave differently with respect to the definiteness agreement found on the matrix verb. While finite complement clauses occur with a verb in the definite conjugation, infinitival ones co-occur with verbs in the indefinite conjugation. Den Dikken takes this to mean that finite clauses are definite and hence trigger definite agreement on their selecting verb, while infinitival clauses are indefinite (see also Kenesei, 1992).

The third and final point to notice about the example in (5) is the fact that the embedded verb bears singular agreement. This is the normal type of agreement with a quantifier like *hány lány* 'how many girls': although semantically plural, they trigger singular morphology on the verb that agrees with them, a type of agreement called 'formal agreement'. In other words, the singular ending on *eljöjjön* 'comes' signals that it is agreeing with *hány lány* 'how many girls' rather than, for example, a resumptive pronoun (see below). Taking together the case of the wh-phrase, the definite agreement on the matrix verb, and the singular agreement on the embedded verb, Den Dikken arrives at the following movement analysis: the wh-phrase *hány lány* 'how many girls' starts out in the embedded clause, receives nominative case there, triggers agreement on the embedded verb, and subsequently moves to the specvP of the matrix clause without an intermediate stopover in specCP. This lack of an intermediate landing site is licensed because the matrix *v* agrees with the embedded clause (yielding definite agreement on *akarod* 'want'). In other words, the example in (5) constitutes an illustration of the first option in (3), successive cyclic movement via vP-edges.

The second strategy is the resumptive prolepsis strategy (described in detail by Gervain, 2009). It starts out from the idea that the 'long-moved' DP is base-generated in the matrix clause and that it is resumed in the embedded clause by a resumptive pronoun. If this resumptive pronoun is covert, then the resulting surface structure looks like a movement

⁴ As indicated by the grammaticality diacritic, this example is not fully acceptable. Indeed, as pointed out by Den Dikken (in press, p. 10), the movement strategy "is generally deemed rather marginal by Hungarian speakers for cases involving *argument* extraction". When it comes to adjunct extraction, however, the movement strategy is the only one available. Given that we will focus on argument extraction (and predicative XP extraction) in our LSF-data below, however, we gloss over this complication here. As will become clear in section 5, LSF shows the same preference for non-movement as wh-extraction of arguments. See also the next section for some more discussion of the colloquial status of the constructions under discussion here.

dependency, even though the only movement operations involved are strictly clause-bound.⁵ The schematic structure for prolepsis is given in (8), and an example in (9) (Den Dikken, in press, p. 6).⁶

(8) XP... [_{VP} XP [_{VP} V [_{VP} V XP [_{CP} C... pro...]]]]

(9) %Hány lányt akarsz, hogy eljöjjenek?
 how.many girl.acc want.2sg.indef that pv.come.3pl
 'How many girls do you want to come?'

This example differs from the one in (5) in three ways. First of all, the left-peripheral wh-phrase now bears accusative rather than nominative case. This suggests that it is receiving case inside the matrix clause (from the matrix verb *akarsz* 'want'), not as the subject of the embedded clause. This is further corroborated by the indefinite agreement on the matrix verb, i.e. the matrix verb not only assigns case to *hány lányt* 'how many girls', it also agrees with this DP in definiteness (see also Den Dikken, 2006). Thirdly, notice the plural agreement on the embedded verb *eljöjjenek* 'come'. This suggests that this verb is not agreeing with the wh-phrase *hány lányt* 'how many girls', but rather with a pronominal element that is, according to Den Dikken, in a semantic (and not grammatical or formal) agreement relation with its antecedent, the wh-phrase. As a result, this pronominal element can only trigger plural agreement on the embedded verb (see Gervain, 2009, for a different view). These three differences lead Den Dikken to conclude that the example in (9) differs from the one in (5) in that it does not involve long-distance movement.

With one long-distance A'-dependency analyzed as movement and another one as involving base-generation/resumption, all the relevant options seem to be exhausted. However, Den Dikken proposes a third option, scope-marking (cf. (3c)), which is based on so-called partial wh-movement constructions. The idea here is that a scope marker is merged in the matrix clause, and a full DP in the embedded clause.⁷ Moreover, there can be a concord relationship between the two, and this leads to a further tripartition within this category.⁸ In the first case, there is no concord at all and what we derive are 'plain' scope marking constructions: a neuter wh-pronoun sits in the matrix specCP, while the fully contentful wh-phrase surfaces in the embedded specCP. Such constructions are familiar from certain varieties of German, and are found in Hungarian as well:

(10) Was glaubst du wen er gesehen hat?
 what believe you who.acc he seen has
 'Who do you believe he has seen?'

(11) Mit akarsz, hogy hány lány jöjjön el?
 what.acc want.2sg.indef that how.many girl.nom come.3sg pv
 'How many girls do you want to come?'

The second option is scope-marking with partial concord, the scope marker taking over a subset of the features of the DP in the embedded clause. Typically, this subset only consists of the ϕ -features of the wh-constituent, and, in the German example below, also its case feature.

(12) Wen glaubst du wen er gesehen hat?
 who.acc believe you who.acc he seen has
 'Who do you believe he has seen?'

This construction goes by the name of wh-copying in the literature, but Den Dikken argues that this is a misnomer, as there is no literal copying involved in such examples, only partial concord. This partial concord involves ϕ -features (and possibly

⁵ Gervain (2009) argues that the resumptive is always covert, whereas Den Dikken (in press) argues that some speakers marginally accept an overt resumptive pronoun. The Hungarian author of this paper likewise accepts this structure with an overt subject resumptive pronoun. See also Jánosi (2014) for examples and discussion.

⁶ Once again we see that the example is not grammatical for all speakers of Hungarian. See footnote 4.

⁷ As shown in Horváth (1995, 1997, 1998, 2000), recently revived in Haegeman and Ürögdi (2010), the wh-expletive pronoun *mit* 'what' in wh-scope marking constructions is most likely not an associate of the embedded question word but of the entire embedded clause and as such, no chain exists between the two wh-phrases. This is compatible with our analysis of LSF.

⁸ See Haegeman and Ürögdi (2010) for an alternative view and a detailed discussion of the availability of wh-expletives vs. long-movement structures.

also case), but crucially does not involve the features residing under D, i.e. the features relating to the quantificational properties of the *wh*-constituent. This is why, contrary to surface appearances, a case like (12) involves only partial, not full, concord in Den Dikken's analysis.⁹

The third option, scope-marking with full concord, is the most interesting one for our purposes, as it leads to a surface representation which, at least at first sight, gives the impression that long-distance movement has taken place, because the lower DP does not surface in the embedded clause and we find singular agreement in that clause. Its schematic representation is given in (13) and an example in (14) (Den Dikken, *in press*, p. 10).

(13) SM_i... [_{VP} SM [_{VP} V [_{VP} V SM [_{CP} XP_i C... XP...]]]]

(14) ?Hány lányt akarsz, hogy eljöjjön?
 how.many girl.acc want.2sg.indef that pv.come.3sg
 'How many girls do you want to come?'

Let us first take a look at the schematic representation in (13). The contentful XP (corresponding to *hány lányt* 'how many girls' in (14)) is merged in the embedded clause, and moves to the specCP of that clause. The scope marker (abbreviated as SM) is merged in the matrix clause inside VP, where it enters into a local concord relation with the contentful XP. This is a full concord relation in that all the formal and semantic features of the lower XP are copied onto the SM, except for those uninterpretable features that have already been checked in the embedded clause. In the example given, nominative case has been checked in the embedded clause and is therefore not copied under concord. Given their (near-)identity, the cooccurrence of SM and XP will yield a linearization problem at PF (Nunes, 2004) and as a result, one of them, the lower one, i.e. XP, will have to be deleted. The result is a structure in which there is only one *wh*-phrase (SM), which surfaces in the matrix clause but thematically belongs in the embedded clause, i.e. a configuration resembling that of long-distance movement. This, Den Dikken argues, is how the derivation of (14) proceeds.

Note that as far as its nominal and verbal morphology is concerned, this example is a mix of (5) and (9). On the one hand, the embedded verb bears singular agreement, suggesting that it is the *wh*-phrase *hány lányt* 'how many girls' itself, and not a resumptive pronoun, which has agreed with this verb, as in a movement derivation. On the other hand, the matrix verb shows indefinite agreement and the *wh*-phrase bears accusative case, which suggests that the *wh*-phrase is base-generated inside the matrix clause. The scope marking analysis manages to reconcile these apparently contradictory requirements: in a way, the *wh*-phrase originates both in the embedded clause and in the matrix clause, without the two instances being linked via movement.

Now that we have introduced and illustrated Den Dikken's classification of long-distance A'-dependencies in Hungarian, it is time to link it to our central topic, i.e. LSF. In the remainder of this paper, we focus exclusively on the movement analysis on the one hand, and scope-marking with full concord (i.e. base-generation) on the other.¹⁰ These were illustrated by (5) and (14), respectively, both repeated below.

(15) ??Hány lány akarod, hogy eljöjjön?
 how.many girl.nom want.2sg.def that pv.come.3sg
 'How many girls do you want to come?'

(16) ?Hány lányt akarsz, hogy eljöjjön?
 how.many girl.acc want.2sg.indef that pv.come.3sg
 'How many girls do you want to come?'

The first task we face is to apply Den Dikken's analysis to LSF. It is clear that LSF cannot work in quite the same way as just described, since there is an adjectival remnant in the embedded clause. Both under the movement and under the base-generation analysis, something will need to be said about this additional element. Before embarking on an analysis

⁹ Den Dikken argues that *wh*-copying is absent from Hungarian because Hungarian spells out moved *wh*-phrases in specFocP instead of specCP, so that the phase head C intervenes between the two constituents involved in the concord relationship. Although *wh*-phrases can raise terminally to specCP, and must so do in the derivation of full-concordial scope marking constructions (see below), they are never pronounced in specCP in Hungarian. We refrain from a fuller discussion of this issue as it is not the central concern of this paper.

¹⁰ Note that while this means that on the surface we are setting aside the resumptive prolepsis analysis, this option will continue to play a role behind the scenes, as the account we will propose takes over certain characteristics of the resumptive prolepsis approach (see in particular the discussion surrounding (42b) below). For expository purposes, however, we stick to the binary opposition between movement and base-generation at this point.

of LSF, however, we need to establish that LSF and unsplit A'-movement are indeed part of the same family of operations. This is the focus of the next section.

3. Split and unsplit A'-dependencies

In this section, we discuss two types of focalization constructions, the long unsplit focus (LUF) construction and the long split focus (LSF) construction. We show that they share basic properties with the two types of wh-movement constructions discussed by Den Dikken.

3.1. Information structure

Both LSF and LUF are part of spoken, colloquial or non-standard Hungarian. Their standard Hungarian counterpart is the so-called expletive-associate construction (henceforth EA), which is exemplified in (17).

- (17) AZT mondta, hogy ÚJ AUTÓT kért el.
 expl.acc said.3sg.def that new car.acc asked.3sg.indef pv
 'He said he had borrowed a NEW CAR.'

In this example the complement clause is introduced by the expletive pronoun *azt* and the DP *új autót* 'new car' is focused inside the embedded clause. Movement to specFocP is signalled by the use of verb–preverb inversion. The meaning of this sentence can be paraphrased in English with a pseudocleft: *What he said he borrowed was a new car*. In this respect, (17) is very similar to a sentence which has undergone long focus movement.¹¹

- (18) ÚJ AUTÓT mondott, hogy elkért.
 new car.acc said.3sg.indef that pv. asked.3sg.indef
 'He said he had borrowed a NEW CAR.'¹²

In an EA-example such as (17), we can also choose to focus only a subpart of the nominal projection rather than the entire DP.

- (19) Azt mondta, hogy új AUTÓT kért el.
 expl.acc said.3sg.def that new car.acc asked.3sg.indef pv
 'He said he borrowed a new CAR.'

This example can be paraphrased as follows: *The new thing which he said he borrowed is a car*. As such, it is very similar in meaning to an instance of LSF (with the same provisos that we just mentioned for the case of unsplit focus).¹³

¹¹ It should be clear in light of the preceding discussion that we are using the term 'long focus movement' in a non-theoretical sense here, i.e. as synonymous with 'long-distance A'-dependency involving focus'. It is not meant to imply anything about the syntactic analysis of this example.

¹² Although the same element is focused in (18) and (17), the information structure of the two sentences is not entirely identical. In (17), everything up to *új autót* 'new car' can be new information, including the matrix verb. In contrast, in (18), the matrix verb has to be background information, because it follows the focus. Lipták (p.c.) also points out that the scope of the two foci is distinct, as it corresponds to their c-command domains. If there is a modal verb in the matrix clause, that modal will scope over and under the focus in the expletive-associate and long focus movement construction, respectively.

¹³ In (20) we change the embedded verb to one without a preverb, as the example is impossible with a preverb, whether it is inverted or not. A true minimal pair between EA and LSF thus cannot be constructed. The LSF case with a preverb in the embedded clause requires either an additional focus on the adjective (and verb–preverb inversion, as shown in (i)), or the addition of the indefinite article to the adjective in the embedded clause (see (ii)).

- (i) AUTÓT mondott, hogy ÚJAT kért el.
 car.acc said.3sg.indef that new.acc borrowed.3sg.indef pv
 'He said that he had borrowed a NEW CAR.'

- (ii) A: Mit mondott, hogy elkért egy újat?
 what.acc said.3SG.indef that pv. asked.3sg.indef a new.acc
 'What new thing did he say he had borrowed?'
 B: AUTÓT mondott, hogy < egy újat > elkért < egy újat >.
 car.acc said.3sg.indef that a new pv. asked.indef a new.acc
 'It's a new CAR that he said he had borrowed.'

- (20) AUTÓT mondott, hogy újat vett.
 car.acc said.3sg.indef that new.acc bought.3sg.indef
 'He said that he had bought a new CAR.'

Comparing LUF and LSF constructions from the perspective of information structure, we see that they are close analogues, differing merely in the scope of what is in focus: in LUF examples both the adjective and the noun are, while in LSF, only the noun is focused.¹⁴

3.2. Syntactic similarities between *wh*-movement, LSF and LUF

A number of syntactic restrictions that apply both to LUF and LSF confirms their membership to the same general class as *wh*-movement (for detailed discussion, see Brody, 1995). First and foremost, there is the fact that all types of focus movement in Hungarian, like *wh*-movement, trigger verb–preverb inversion, suggesting that focused elements and *wh*-words move to (or through) the same position (Lipták, 2001). The examples in (22) show this for short *wh*-movement, short unsplit focus movement, and short split focus movement, respectively. The example in (21) is a baseline case without verb–preverb inversion.

- (21) Mari felhívta Jánost.
 Mari.nom pv.called.def.3sg János.acc
 'Mary called Janos.'
- (22) a. Kit hívott fel Mari?
 who.acc called.indef.3sg pv Mari.nom
 'Who did Mary call?'
 b. JÁNOST hívta fel Mari.
 Janos.acc called.def.3sg pv Mari.nom
 'Mary called JÁNOS.'
 c. AUTÓT tört össze újat.
 car.acc broke.3sg.indef pv new.acc
 'She broke a new CAR.'

With the long versions of these movements, verb–preverb inversion takes place in the matrix clause, as illustrated for LSF in (23).

- (23) AUTÓT híresztelte el, hogy újat tört össze.
 car.acc announced.3sg.def pv that new.acc broke.3sg pv
 '(S)he announced that (s)he broke a new car.'

A second restriction concerns the incompatibility of LUF, LSF, and *wh*-movement with the (otherwise optional) matrix expletive *azt* (see also Kenesei, 1992, for discussion of the conditions on the omission of the expletive). This is true regardless of whether the matrix verb shows definite or indefinite agreement.

- (24) *Azt ÚJ AUTÓT mondott/mondta, hogy vett.
 expletive.acc new car.acc said.3sg.indef/def that bought.3sg.indef
 INTENDED: 'He said he had bought a NEW CAR.'

¹⁴ In LSF the adjective can optionally be contrastively stressed as well, in addition to the noun, e.g. when it forms an answer to the question in (i):

- (i) A: Mit mondott, hogy újat kért el, és nem régít?
 What.acc said.3sg.indef that new.acc asked pv and not old.acc
 'What did he say that he borrowed new, and not old?'
 B: AUTÓT mondott, hogy ÚJAT kért el.
 car.acc said.3sg.indef that new asked pv
 'It's a CAR that he said that he borrowed NEW.'

Such contrastive focus on the adjective is optional, however, so that this case need not concern us here.

- (25) *Azt AUTÓT mondott/mondta, hogy újat vett.
 expletive.acc car.acc said.3sg.indef/def that new.acc bought.3sg.indef
 INTENDED: 'He said that he had bought a new CAR.'
- (26) *Mit hány lányt akarsz/akarod, hogy eljöjjön?
 what.acc how.many girl.acc want.2sg.indef/def that pv.come.3sg
 INTENDED: 'How many girls do you want to come?'

Thirdly, the finite complementizer *hogy* 'that', which is normally optional at the beginning of complement clauses (even if this clause contains embedded *wh*- or focus-movement), cannot be left out in the case of a long-distance A'-dependency, regardless of whether the fronted XP is split or unsplit:

- (27) *ÚJ AUTÓT mondott vett.
 new car.acc said.3sg.indef bought.3sg.indef
 'He said he had bought a NEW CAR.'
- (28) *AUTÓT mondott újat vett.
 car.acc said.3sg.indef new.acc bought.3sg.indef
 'He said that he had bought a new CAR.'

The same restriction is found with long *wh*-movement:

- (29) *Hány lányt akarsz eljöjjön?
 how.many girl.acc want.2sg.indef pv.come.3sg
 'How many girls do you want to come?'

Further syntactic similarities between *wh*-movement, LUF, and LSF involve the case of the fronted constituent and the definiteness agreement on the matrix verb. Den Dikken (in press) observes that accusative (matrix) case on the fronted *wh*-constituent correlates with indefinite agreement on the matrix verb, and embedded case on the fronted *wh*-constituent with definite agreement on the matrix verb. The relevant examples are repeated here:

- (30) ??Hány lányt akarsz, hogy eljöjjön?
 how.many girl.nom want.2sg.def that pv.come.3sg
 'How many girls do you want to come?'
- (31) ?Hány lányt akarsz, hogy eljöjjön?
 how.many girl.acc want.2sg.indef that pv.come.3sg
 'How many girls do you want to come?'

Recall that for Den Dikken the definite marking on the matrix verb in (30) is due to the fact that it agrees with the finite embedded clause. In contrast, the indefinite agreement in (31) results from agreement of the matrix verb with the fronted *wh*-constituent. This agreement correlates with matrix case assignment. These two sentences have two different derivations for Den Dikken: matrix case and agreement are the result of full concordial scope-marking, i.e. base-generation of the fronted constituent, whereas embedded case on the fronted constituent is the result of movement.

We find exactly the same correlation between case and agreement with LSF and LUF. The examples in (32) (LSF) and (33) (LUF) feature accusative case on the fronted constituent, and they require indefinite agreement on the matrix verb (the embedded verb assigns dative, as indicated by the case marking on the adjective in the LSF construction):¹⁵

¹⁵ In (33) the embedded clause has a dative resumptive pronoun in the place where (32) has the dative-marked stranded adjective. This resumptive pronoun is required when it is assigned an oblique case. Nominative and accusative resumptive pronouns are possible but not required in LUF. This follows from the rules for pronoun drop in Hungarian. The resumptive pronoun is also not required in a case like (35) below, where the fronted constituent in focus position bears embedded case. Although we think the distribution of resumptive pronouns in LUF fits in rather straightforwardly with the analysis we shall develop below, we shall not go into it any further here, as this would lead us too far afield.

- (32) AUTÓT hallott/*hallotta, hogy újnak örülnének.
 car.acc heard.3sg.indef/def that new.dat be.pleased.cond.3pl
 'He heard that they would be pleased with a new CAR.'
- (33) ÚJ AUTÓT hallott/*hallotta, hogy annak örülnének.
 new car.acc heard.3sg.indef/def that it.dat be.pleased.cond.3pl
 'He heard that they would be pleased with a NEW CAR.'

In contrast, when the fronted constituent takes embedded (i.e. dative) case, only definite agreement is possible (for those speakers who accept this derivation; see below for discussion):¹⁶

- (34) %AUTÓNAK *hallott/hallotta, hogy újnak örülnének.
 car.dat heard.3sg.indef/def that new.dat be.pleased.cond.3pl
 'He heard that they would be pleased with a NEW CAR.'
- (35) ÚJ AUTÓNAK *hallott/hallotta, hogy örülnének.
 new car.dat heard.3sg.indef/def that be.pleased.cond.3pl
 'He heard that they would be pleased with a NEW CAR.'

In sum, constructions of long wh-movement, LUF, and LSF share a large number of characteristics. The case and agreement data furthermore suggest a double derivation analysis, one derived by movement, the other an instance of base-generation. This being established, we now proceed to developing such an analysis. In so doing, we focus on the hitherto unstudied LSF construction. Although our analysis of LSF straightforwardly extends to LUF, we shall restrict ourselves to the proper analysis of LSF here, referring the reader to Jánosi (2014) for a more extensive discussion of LUF.

4. The analysis

4.1. Introduction

In this section we propose an analysis for LSF in Hungarian. In line with Den Dikken's proposal, we argue that LSF constructions may be either derived by movement, or by base-generation. As in Den Dikken's analysis of wh-movement, these different derivations are diagnosed by the case marking and definiteness agreement patterns in the matrix clause. Notably the case-marking patterns are interesting in the case of LSF, because, in contrast to the A'-dependencies discussed by Den Dikken, LSF can show a visible case-mismatch between the noun in the matrix clause and the adjective in the embedded clause. This is illustrated by the example in (36).¹⁷

- (36) AUTÓT hallott, hogy újnak örülnének.
 car.acc heard.3sg.indef that new.dat be.pleased.cond.3pl
 'He heard that they would be pleased with a new CAR.'

The potential for case mismatch correlates systematically with the presence of agreement between the matrix verb and the noun in the matrix clause, as we shall show in more detail in the following section.¹⁸ We take examples such as (36) to instantiate the base-generation derivation.

As we saw in the previous section, example (36) has a variant with matching case and definite agreement marking in the matrix clause:¹⁹

- (37) %AUTÓNAK hallotta, hogy újnak örülnének.
 car.dat heard.3sg.def that new.dat be.pleased.cond.3pl
 'He heard that they would be pleased with a new CAR.'

¹⁶ 30.12% of our informants accepted the sentence in (34), while 69.88% rejected it.

¹⁷ We return to case mismatches of this sort more extensively below.

¹⁸ We use the term 'potential for case mismatch' because actual case mismatches can only be observed if the matrix verb assigns a case that is different from the case assigned by the embedded verb.

¹⁹ We return to the %-mark below.

We take such examples to be derived via movement. At the same time, however, we need to account for the presence of the adjective inside the embedded clause, an element that was missing in Den Dikken's data. For this we will make use of Lipták and Saab's theory of NP-ellipsis and Ott's analysis of split DPs (Lipták and Saab, 2012; Ott, 2011).

A further peculiarity about LSF constructions is the fact that the adjective in the embedded clause bears a case ending. As shown in (38), in a regular (i.e. unsplit) DP, case is marked only once, and obligatorily on the noun, never on the adjective.²⁰

- (38) egy(*nek) új(*nak) autó*(nak)
 a.dat new.dat car.dat
 'a new car'

This is also a property of the construction that our analysis has to provide an account for.

The remainder of this section is organized as follows. We first discuss the structure that we assume underlies both types of LSF constructions. This structure is largely similar to the one that Ott (2011) assumes underlies split topics in German (section 4.2). Next, we discuss the case-mismatching type of LSF illustrated in (36) and argue that it arises as the combination of base-generation and concord (section 4.3). In section 4.4 we turn to data such as (37) and propose a movement account.

4.2. The underlying structure of LSF

Both types of LSF start out from the same underlying structure, which is a predication structure as proposed by Ott (2011). Ott discusses split topic constructions in German of the following type:

- (39) Gute Zeitungen kennt Fabian nur eine e aus Berlin
 good newspapers knows Fabian only one from Berlin
 'As far as good newspapers are concerned, Fabian only knows one from Berlin.'

He proposes that the two portions of a split topic start out as separate constituents in a predication structure as shown in (40):

- (40)
-
- ```

graph TD
 Root[] --- DP[DP]
 Root --- NP[NP]
 DP --- DP1[eine e aus Berlin]
 NP --- NP1[gute Zeitungen]

```

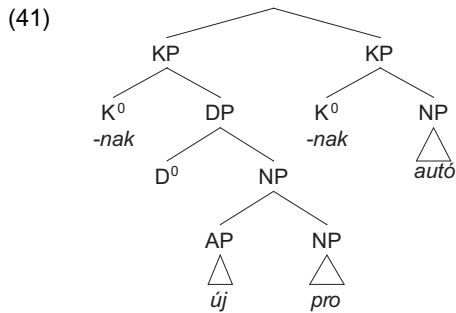
The subject of this predication, DP, contains an elided noun, represented by *e*.<sup>21</sup> The problem with this structure is that neither of its two composing parts is a lexical item that could provide a label for the dominating node. In order to break the symmetry between DP and NP in (40), NP has to move out, so that DP can provide the label for the dominating node (see also Moro, 2000). Once labelled as DP, this constituent satisfies the selection requirements of V (albeit at the phase level only, i.e. after movement).

Applied to the case of LSF in Hungarian, we assume that the underlying structure is essentially the same as for German split topics, i.e. a complex predication structure with a DP subject and an NP predicate:

<sup>20</sup> The adjective does bear a case ending in those cases in which the noun is missing, such as in the following elliptical exchange:

- (i) A: Melyik autónak örülsz?  
 Which car.dat be.pleased.2sg?  
 'Which car are you pleased with?'  
 B: Az újnak örülök.  
 the new.dat be.pleased.1sg.  
 'I'm pleased with the new one.'

<sup>21</sup> Ott remains agnostic about the exact nature of NP-ellipsis in German.



There are a number of obvious differences between this tree and Ott's in (40). For one thing, both the subject and the predicate node of the predication structure are topped by a KP (for concreteness, we have added the dative suffix *-nak* in the tree). This relates to the presence of case morphology on both the fronted noun and the adjective in Hungarian LSF constructions, an issue to which we return below. For another, the subject contains an adjective rather than an article (cf. *eine* in (40)), and we represent the elided noun as *pro*. These differences aside, the structure is entirely parallel to the structure assumed by Ott.

As in Ott's system, we assume that the node dominating this structure cannot be labelled because its two components are too symmetric with respect to one another. This symmetry can be remedied in either of two ways: through ellipsis of one of its members, or through movement. As will become clear in the following sections, we argue that Hungarian LSF instantiates both these options.

In both types of LSF, the adjective is case-marked, and it receives embedded case. Following Ott (2011), we assume that when the structure in (41) is merged as the sister of V, both parts of it simultaneously get case-marked under multiple Agree (Ott, 2011, pp. 70–73). In the subject DP of the predication structure, however, case morphology cannot be lowered onto the noun, as this noun is elliptical. As a last resort rescue operation, the case ending gets appended to the adjective, which surfaces with case morphology. In this analysis, we follow Lipták and Saab (2012), who discuss plural number marking on Hungarian adjectives, which remains absent on the adjective if the noun is realised, but appears on the adjective if the noun is elided, exactly like the case morphology discussed here. In this way, we account for the fact that the adjective in the embedded clause systematically displays case morphology in both types of LSF.

Before we turn to a more detailed discussion of these two derivations, we discuss a prediction that this analysis makes. In particular, we expect the subject of the predication to be able to host D-elements, like articles and demonstratives. This prediction is borne out, as in the following examples with the indefinite article *egy* 'a' (see also note 14), and the demonstrative *azt*, respectively:<sup>22</sup>

- (42) a. AUTÓT hallott, hogy vettek egy újat.  
 car.acc heard.3sg.indef that bought.3pl.indef a new.acc  
 'He heard that they had bought a new CAR.'
- b. AUTÓT mondott, hogy azt újat vett.  
 car.acc said.3sg.indef that dem.acc new.acc bought.3sg.indef  
 'He said that he bought a new CAR.'

We now proceed to a discussion of the two derivations possible on the basis of the predication structure in (41), beginning with the concord derivation.

#### 4.3. LSF type 1: base-generation and concord

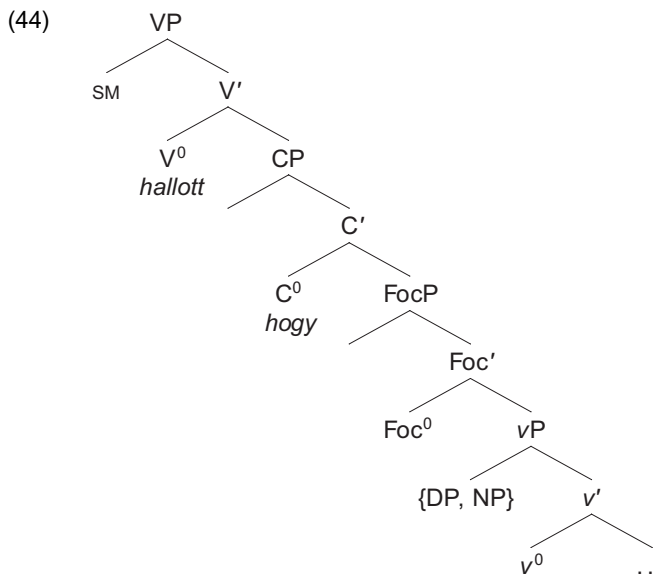
In this section we provide an analysis of the example in (36), repeated here as (43):

- (43) AUTÓT hallott, hogy újnak örülnének.  
 car.acc heard.3sg.indef that new.dat be.pleased.cond.3pl  
 'He heard that they would be pleased with a new CAR.'

<sup>22</sup> In the b-example the embedded verb surprisingly shows indefinite agreement, despite the demonstrative being formally definite. This could be explained by assuming that *azt* spells out *pro* rather than D, as it can in other cases of NP-ellipsis. We shall not pursue this matter any further here.

We take the case mismatch in this example (accusative in the matrix clause, dative in the embedded) to be an indication that the dependency between *autót* ‘car’ and *újnak* ‘new’ is not one that is created by movement, but rather that both parts are independently base-generated inside their own clause. In so doing, we follow Den Dikken’s analysis in terms of scope marking.

An analysis in terms of scope marking implies that it is the lower of the two DPs which is the truly contentful one. The higher one is essentially a scope marker which has undergone full concord with (a portion of) the lower DP, to the extent that the two become indistinguishable, which in turn leads to ellipsis of the lower copy. This means that the derivation of the embedded clause starts out with the complex predication structure in (41) in the complement position of the verb, which then raises to specvP.<sup>23</sup> Ignoring further functional projections that may be merged in the embedded clause, this structure is then merged with CP (the head of which is spelled out as *hogy* ‘that’) and this CP is in turn selected by and merged with the matrix verb *hallott* ‘hear’. Moreover, following Den Dikken (in press), we assume that the specifier of the matrix VP is occupied by a scope marker. The relevant portion of this stage of the derivation is represented in (44) (where the scope marker is indicated as SM). We represent the unlabelled node of the complex predication as the unordered set {DP, NP}.



In this configuration, the scope marker undergoes concord with the constituent in specvP. Den Dikken assumes that even in cases of full concord, there is a part of the nominal constituent that does not partake in this process (viz. its case features, see above, section 2). This feature of Den Dikken’s analysis will allow us to account for the case mismatch. Moreover, given that in our analysis the scope marker enters into a concord relation with the NP-portion of the predication structure (rather than with its DP-subject), the focused XP that shows up in the matrix clause in LSF is necessarily indefinite, since definiteness is encoded in the DP-domain. Following Den Dikken (in press), this process of full concord leads to ellipsis of the lower member of the concord relationship, in this case the NP-predicate in the lower specvP. Subsequently, the scope marker (which is now indistinguishable from the NP *autó* ‘car’) raises to specvP in the matrix clause, triggers indefinite agreement on the matrix verb and receives accusative case. It then raises to the specFocP of the matrix clause, thus successfully concluding the derivation of (43).

This analysis allows us to account for the fact, observed earlier, that LSF (and LUF, for that matter) are incompatible with the matrix expletive *azt*. We only have to make one additional (plausible) assumption, which is that object agreement in the matrix clause, and concomitant assignment of matrix object case, can only take place once. Since in the case-mismatching type of LSF the base-generated noun is assigned matrix object case, the expletive cannot be assigned case, and must therefore remain absent.

Summing up, in this subsection we have provided an analysis for the case-mismatching LSF-example in (43). We have argued that it involves concord between a scope marker in the specVP of the matrix clause, and a subpart of a complex

<sup>23</sup> As we saw earlier, the embedded adjective may, under certain discourse conditions, undergo further movement to the embedded specFocP, both in the concord and the movement type of LSF (see e.g. note 15 above). On the other hand, the movement to specvP is not (always) obligatory either, as the adjective may be stranded in a postverbal position as well, provided it is preceded by an indefinite article, as shown in e.g. (42a). We ignore these complications here, as they are not crucial to our concerns.

predication structure in the embedded clause. This concord relationship results in NP-ellipsis in the embedded clause, which in turn eliminates the local instability in the predication structure. In the next subsection we turn to the case-matching variant of LSF.

#### 4.4. LSF type 2: movement driven by local instability

In the LSF-example in (37) (repeated here as (45)) the noun found in the matrix clause shows up with a morphological case ending that cannot have been assigned in this clause.

- (45) %AUTÓNÁK hallotta, hogy újnak örülnének.  
 car.dat heard.3sg.def that new.dat be.pleased.cond.3pl  
 'He heard that they would be pleased with a new CAR.'

It stands to reason, then, to analyze this type of example as involving movement rather than base-generation. As discussed in section 4.2, this movement is driven by the need to make the underlying predication structure asymmetric, so that it can be labeled. The movement derivation is identical to the base-generation one in all relevant respects, except that the NP-predicate of the complex predication structure is not elided as a result of concord, but instead subextracts and moves into the specvP of the matrix clause on its own, and from there on to the matrix specFocP. The resulting structure is one in which the dative-marked NP sits in the specFocP of the matrix clause, while the AP, which is part of the subject DP of the predication structure, is stranded in the embedded clause, i.e. the example in (45). Recall that according to [Den Dikken \(in press\)](#) this movement operation is only possible because the matrix verb shows (definite) agreement with the finite clause. We agree with this analysis in that it is local agreement of the embedded CP with the matrix *v* that renders extraction of the NP out of the embedded clause possible: this directly accounts for the fact that the movement derivation is only compatible with definite agreement on the matrix verb. Moreover, this also allows us to account for the obligatory absence of the expletive in LSF (see above, section 3) by our earlier assumption that object agreement in the matrix clause, and concomitant assignment of matrix object case, can only take place once. In the case of movement that we are discussing here, the matrix *v* agrees with the embedded CP and therefore cannot agree with the expletive *azt*.<sup>24</sup> As a result, *azt* cannot receive matrix object case, and is ruled out. Agreement with the moved constituent is ruled out for the same reason, but since this constituent has been case-marked in the embedded clause, this is not a problem. If the verb were to agree with the expletive in the matrix clause, it could no longer agree with the embedded CP, and movement out of that CP would be blocked. There is, in other words, no way in which the expletive can appear in the movement derivation of LSF. The same reasoning applies to LUF, where the same restriction is found.<sup>25</sup>

#### 4.5. A remaining issue

We have argued above that the complex predication structure underlying LSF represents a labelling problem, which may be solved by either eliding or moving the predicate NP of the predication structure. There is, however, an alternative resolution of the labelling issue that we have not considered so far, which would involve elision or movement of the subject DP of the predication structure. Ott considers this issue, and observes that in the symmetric structure in (40) the symmetry cannot be lifted by moving out the subject DP:

- (46) \*Eine e aus Berlin kennt Fabian nur gute Zeitungen.  
 a from Berlin knows Fabian only good newspapers

He assumes that this restriction is to be explained on pragmatic grounds, i.e. movement of the subject of the predication (DP) is possible in principle, but ruled out by general pragmatic constraints on topic–comment structure. The idea is that a

<sup>24</sup> See [De Cuba and Ürögdi \(2009\)](#) and [Haegeman and Ürögdi \(2010\)](#) for an alternative view. They claim that the clausal expletive only appears in the matrix sentence of embedding constructions when it stands in for the entire embedded clause; that is, when the embedded clause is associated with either the verb-modifier position or an A'-position (focus or topic) of the matrix clause. As such, sentences involving long-distance constituent focus are not expected to feature the clausal expletive.

<sup>25</sup> This account predicts that *azt* should not be incompatible with a focused constituent in the matrix clause as long as this constituent does not originate in the embedded clause, and does not compete with *azt* for accusative case. Such a case would be exemplified by a short-moved subject bearing nominative case. This prediction is indeed borne out:

- (i) Azt MARI mondta, hogy új autót vett.  
 Expl.acc Mary said.3sg.def that new car.acc bought.3sg.indef  
 'It was Mary that said that she had bought a new car.'

fronted topic sets a frame, and that the comment must be interpretable as being about that frame. Such a pragmatically well-formed information structure does not arise in (46), Ott argues (Ott, 2011, 84ff).

Before considering the question if this analysis carries over to Hungarian LSF, we need to review the facts first. In LSF, the adjectival portion of the focused DP cannot occur in the matrix clause (cf. (47)), but it looks like it is allowed in short-distance cases (cf. (48)).

(47) \*ÚJNAK hallotta, hogy autónak örülnének.  
new.dat heard.3sg.def that car.dat would.be.pleased.3pl

(48) Újnak AUTÓNAK örülnének.  
new.dat car.dat would.be.pleased.3pl  
'Of new things they would be pleased with a CAR.'

The reason why (48) is subject to an analysis as a split construction rather than as a run-of-the-mill DP with an adjective inside it is that the adjective is case-marked.<sup>26</sup> This suggests an analysis as a split construction, along the lines of our analysis of LSF, but differing from it in that the subject DP (which contains the adjective *újnak* and an elided noun) has moved out of the predication structure rather than the predicate NP *autónak*. At the same time, however, it is clear that *autónak* has also undergone movement, as it occurs in the preverbal focus position. It is also striking that the information structure of (48) is different from that of (47), in that the adjective is in focus in (47) but a topic in (48), the noun *autónak* being the focus. The generalisation seems to be that the adjective cannot become the focus of a split construction. This in turn suggests that in this case too, the reason for the absence of split constructions like (47) may be pragmatic. This is confirmed by the fact that, when we try to background the noun and focus the adjective, as in the following discourse, the noun must remain silent:

(49) A: Milyen autónak hallottad, hogy örülnének?  
What.kind.of car.dat heard.2sg.def that would.be.pleased.3pl  
'What kind of car did you hear they would be pleased with?'

B: ÚJNAK hallottam, hogy örülnének.  
new.dat heard.1sg that would.be.pleased.3pl  
'I heard that they would be pleased with a new one.'

So while it is possible to have the adjective in focus and the noun in the background, this leads to the obligatory elision of the noun, so that one is hard pressed to find any evidence that there ever was a complex predication structure in the first place, and if so, what has moved out to rescue the structure. We shall not pursue this matter any further here, but leave it as a topic for further research.

#### 4.6. Summary

We have provided an analysis for the two types of LSF. While the finer details of those analyses have led us into various directions, the basic distinction between the two types of LSF should be clear: the first involves base-generation and concord, the second movement. In the next section we provide two pieces of evidence in support of this double derivation analysis.

<sup>26</sup> It is not only the presence vs. absence of case marking that differentiates split nominal phrases from unsplit ones. Jánosi (2014) argues that at least one of the subparts of split nominal phrases must occur as contrastive topic in Spec, CTopP (i) or as focus in Spec, FocP (ii).

(i) Autót azt vettem újat, de házat azt nem.  
car.acc res.pro.acc bought.1sg new.acc but house.acc res.pro.acc not  
'I bought a new car but I didn't buy a new house.'

(ii) Csak AUTÓT vettem újat.  
only car.acc bought.1sg new.acc  
'I only bought a new CAR.'

Unsplit nominal phrases, on the other hand, are not restricted to these two preverbal projections. As (iii) shows, unsplit nominal phrases can appear post-verbally, which is not an option for split ones.

(iii) Vettem egy új autót / \*autót újat tegnap.  
Bought.1sg a new car.acc / car.acc new.acc yesterday  
'I bought a new car yesterday.'

For more discussion of the differences between split and unsplit nominal phrases see (Jánosi, 2014, pp. 39–42).

## 5. Testing the analysis

### 5.1. Methodology

We present LSF-data which result from the field investigations carried out by the first author in March–April 2011. She distributed a questionnaire containing 91 test items (and 32 fillers) among 83 native speakers. Of the informants, 94% were students at the College of Nyíregyháza, approximately 80% permanently lived in North-Eastern Hungary, and 94% were between 19 and 25 years old. The informants were asked to judge each sentence on a scale of 1 (unacceptable) to 5 (perfectly acceptable).<sup>27</sup>

Before discussing the data proper, we first need to make an important methodological remark. As was pointed out in [section 3](#) (and as also became clear by looking at the data taken from [Den Dikken \(in press\)](#) in [section 2](#)), both LSF and LUF are colloquial, non-standard counterparts of the expletive-associate construction in standard Hungarian. This was also reflected in the survey we conducted: while EA was considered to be perfect (i.e. 5 on a scale of 1–5) by the majority of the informants, the judgments on LSF showed much more variation and were on average lower than those for EA. This means that in interpreting the data we have attached greater weight to relative rather than absolute judgements. Moreover, in what follows we have determined a cutoff point between LSF-examples that were deemed grammatical and those that were not, as follows: we considered all examples grammatical that scored at least 3 on a scale of 1 (unacceptable) to 5 (acceptable). This cutoff point to our mind accurately reflects the marked status these examples have in present-day Hungarian. To determine whether the difference between the ratings of two sentences is significant or not, we performed paired two-sample student's *t*-tests. As is customary, we took a *p*-value below 0.05 to indicate a significant difference.

Finally, we excluded from consideration those speakers who gave a score lower than 3 to a pair of LSF examples where both the matrix and the embedded verb assign accusative case (24% of the informants); examples are given in (52) and (53) below. The fact that a small group of speakers rejects long focus constructions altogether was reported in [Gervain \(2009\)](#) in connection with LUF as well. This seems to be the case for LSF, too, which is not surprising in light of the non-standard status of both LUF and LSF.<sup>28</sup>

### 5.2. Results

With this in mind, we turn to the actual LSF-data. The results that we obtained provide support for the distinction between the scope marking and the movement analysis proposed in [Den Dikken \(in press\)](#) for the long *wh*-movement examples in (50) (base generation) and (51) (movement):

(50) ?Hány lányt akarsz, hogy eljöjjen?  
 how.many girl.acc want.2sg.indef that pv.come.3sg  
 'How many girls do you want to come?'

(51) ??Hány lány akarod, hogy eljöjjen?  
 how.many girl.nom want.2sg.def that pv.come.3sg  
 'How many girls do you want to come?'

These examples correspond to the basic pair of LSF-sentences in (52) (base-generation) and (53) (movement). The %-mark preceding (53) indicates a split in the judgments to which we return in [section 5.3](#).<sup>29</sup>

(52) AUTÓT mondott, hogy újat vett.  
 car.acc said.3sg.indef that new.acc bought.3sg.indef  
 'He said that he had bought a new CAR.'

(53) %AUTÓT mondta, hogy újat vett.  
 car.acc said.3sg.def that new.acc bought.3sg.indef  
 'He said that he had bought a new CAR.'

<sup>27</sup> The entire questionnaire, fully glossed and with a summary of the scores, is available online ([http://www.crispp.be/questionnaire\\_lsf.pdf](http://www.crispp.be/questionnaire_lsf.pdf)).

<sup>28</sup> Note, however, that the 'matrix oblique – embedded accusative' pattern (see below, example (58)) is acceptable for 85.54% of our informants. This means some of the speakers we excluded do seem to have a limited acceptance of the LSF construction, it being essentially restricted to examples with a matrix oblique assigning verb and an embedded accusative assigning one. We leave the further investigation of this pattern for future research.

<sup>29</sup> 43.37% of our informants accepted the sentence in (53), and 56.63% rejected it.

These examples differ only in the definiteness marking on the matrix verb. In (52) the verb bears indefinite agreement, i.e. it agrees with the left-peripheral noun *autót* 'car'. In (53) on the other hand, *mondta* 'said' is definite and so does not agree with *autót* 'car'. As such, this pair of examples is a close analogue to the ones in (50)–(51), in that the first one corresponds to a base generation derivation and the second one to a movement derivation. What makes these particular LSF-examples less than completely informative, however, is the fact that the case assigned by the matrix verb is identical to the one found in the embedded clause. This means that we have no way of telling which case is assigned where. We therefore now turn to LSF-examples with an actual case mismatch potential. First consider the following:<sup>30</sup>

(54) %AUTÓT hallott, hogy újnak örülnének.  
car.acc heard.3sg.indef that new.dat be.pleased.cond.3pl  
'He heard that they would be pleased with a new CAR.'

(55) \*AUTÓT hallotta, hogy újnak örülnének.  
car.acc heard.3sg.def that new.dat be.pleased.cond.3pl  
'He heard that they would be pleased with a new CAR.'

The verb *hall* 'hear' assigns accusative case and its embedded counterpart *örül* 'to be pleased' dative. This means that in the example in (54) the DP-remnant *autót* 'car' is assigned case in the matrix clause, while the rest of this DP (the adjective *újnak* 'new') receives its case in the embedded clause. Moreover, as the contrast in judgments in (54) and (55) ( $t(82) = 4.89$ ,  $p = 4.81 \times 10^{-6}$ ) indicates, this particular pattern of case assignment was found to be only compatible with indefinite agreement on the matrix verb. Given that *autót* 'car' is also indefinite, we take this to indicate that the matrix verb has to Agree with the fronted DP-remnant.

The case mismatch data in (54)/(55) support the base-generation analysis in a way that the A'-dependencies studied by Den Dikken do not, because Den Dikken's data have no remnant material in the embedded clause. Our data indicate that embedded case is assigned in the embedded clause, and therefore that the embedded case-assigning head is not just simply defective, as one might be tempted to think on the basis of the LUF data alone. The fact that both parts in LSF dependencies each have their own independent, possibly different, case strongly confirms the base-generation analysis.

In (56)/(57) we have a case-matching type of LSF:

(56) %AUTÓNAK hallotta, hogy újnak örülnének.  
car.dat heard.3sg.def that new.dat be.pleased.cond.3pl  
'He heard that they would be pleased with a new CAR.'

(57) \*AUTÓNAK hallott, hogy újnak örülnének.  
car.dat heard.3sg.indef that new.dat be.pleased.cond.3pl  
'He heard that they would be pleased with a new CAR.'

Here, the fronted DP-remnant *autónak* 'car' bears the case that is assigned by the embedded verb (just like the stranded adjective). If the matrix verb does not agree with this (indefinite) phrase, as in (56), then some speakers accept the example, while others do not. If there is indefinite agreement, the example is rejected by all LSF-speakers.<sup>31</sup>

Note that case mismatches can come in various flavors, in that 'matrix case assignment' is not necessarily synonymous with accusative case. Consider in this respect the following two examples:

(58) AUTÓRA számított, hogy újat kap.  
car.subl counted.3sg.indef that new.acc receive.3sg.indef  
'He expected to receive a new CAR.'

(59) ?AUTÓRA számított, hogy újjal dicsekedhet.  
car.subl counted.3sg.indef that new.instr boast.can.3sg.indef  
'He expected to be able to boast of a new CAR.'

<sup>30</sup> 49.39% of our informants accepted the sentence in (54), and 50.61% rejected it.

<sup>31</sup> A caveat is in order in that verbs in Hungarian never show definiteness agreement with dative-marked objects, but take the default indefinite agreement form. This could be how the verb in (57) ends up with indefinite agreement marking. The crucial point, however, is that under our analysis movement requires definite agreement marking and hence, that the pattern in (57) is correctly predicted to be ruled out.



Here, the matrix verb *számítani* ‘to expect’ assigns sublative case, while the embedded verbs (*kapni* ‘receive’ in (58) and *dicsekedni* ‘to boast’ in (59)) require accusative and instrumental, respectively. The adjectival part of the split DP shows up with the case of the embedded clause, while the fronted portion shows sublative case assigned by the matrix verb. Moreover, the verb once again appears in the indefinite conjugation. These data further support Den Dikken’s proposal of concordial scope-marking, in particular the idea that the matrix noun is base-generated there as a scope marker and receives the case assigned to a matrix object independently of the case assigned in the embedded clause.

A movement analysis would be hard-pressed to explain this data pattern. For one thing, it would have to assume that a movement chain may be assigned case twice. But as Den Dikken (in press, p. 10) notes, “allowing DPs to possess more than one structural case feature undermines the Case Filter, forfeiting a principled explanation for the distribution of NP–movement”. For another, allowing double case marking in movement chains flies in the face of well-attested patterns of case marking in A’-movement chains, where case on the moved element is invariably that of the foot of the chain.<sup>32</sup> The examples with mismatching cases on the raised noun and the stranded adjective are particularly relevant here. Under the approach we have adopted, the double accusative examples would easily be amenable to an Ott-style (long-distance) movement analysis like the one discussed above (see the structure in (40) and surrounding discussion). In that analysis, NP and AP both receive case, the NP being subsequently extracted. Provided one assumes that the NP can be raised into a higher clause, such an analysis would straightforwardly account for examples with accusative marking on both the noun in the matrix and the adjective in the embedded clause. But such an analysis would crucially fail for examples with different case marking on noun and adjective, such as the ones in (58) and (59).

### 5.3. Further confirmation: speaker variation

The LSF-data presented in the previous subsection provide substantial corroboration of the classification of A’-dependencies outlined in Den Dikken (in press). More importantly, however, the results of the empirical fieldwork reveal that there is systematic variation among the speakers of Hungarian we investigated, in a way that strongly supports the double derivation analysis that we have proposed for LSF. It turns out that all LSF-speakers (63 of our informants) accept base-generation (as diagnosed by the case and agreement patterns in the matrix clause), but only a subgroup of these speakers (32 informants) accept the movement derivation.<sup>33</sup> This split between speaker groups suggests that both LSF types are subject to a different analysis.

If pairs like (52) and (53) were really just one and the same construction, this type of inter- and intra-speaker variation would be highly unexpected. Under the account we have proposed for LSF (following the account of Den Dikken, in press), however, where one type of construction is due to movement and the other to scope marking, this type of variation can be tied to the presence or absence in individual grammars of a particular type of derivation, viz. the movement derivation.<sup>34</sup>

In short, in this section we have shown that the split we have found within speakers of Hungarian supports the double derivation analysis of LSF, thus strengthening the more general view on long-distance A’-dependencies outlined in Den Dikken (in press).<sup>35</sup> In the next section we discuss another argument for this double derivation analysis.

<sup>32</sup> A reviewer points out that the analysis of French ECM-constructions with wh-movement (e.g. *Qui crois-tu être intelligent?*) in Kayne (1984) involves case assignment to an intermediate link in the A’-movement chain. A reanalysis of Kayne’s data in terms of resumptive prolepsis or concordial scope marking would elegantly avoid case assignment to an intermediate chain link. In any event, the construction as analyzed by Kayne was never an instance of double case assignment, as the embedded infinitive does not assign case to its subject.

<sup>33</sup> As one reviewer points out, the pattern of variation reported here may not hold for the dialect spoken in the capital region. This would not be unexpected as our survey involved mainly speakers of North-Eastern Hungary. The investigation of geographical variation concerning LSF throughout Hungary would be an interesting topic for further research.

<sup>34</sup> See also Gervain (2009) for a different split between movement and non-movement speakers. Her system is different from ours in that individual speakers either have the movement or the base-generation derivation in their grammars, but do not mix the two. This implies that what we have taken to be the (more universally accepted) base-generation pattern can be derived either by movement or by base-generation in her system, thus allowing it to become the more widely attested pattern. We shall not attempt a further comparison with Gervain’s theory here, since this would leave us too far afield.

<sup>35</sup> A reviewer invites us to speculate on the reasons for this speaker variation. In this respect, it is worth pointing out that Den Dikken (in press) observes that the movement derivation is the only long A’-fronting strategy available in Hungarian if the fronted constituent is a non-argument (i.e. a predicate nominal or measure phrase). He attributes the fact that with long argument extraction movement is dispreferred to the availability of derivationally simpler ways of forming long A’-dependencies, such as resumptive prolepsis and scope marking (see also note 5). The speculation we can offer at this point is that this preference for nonmovement derivations for certain speakers goes as far as a complete rejection of the movement derivation in cases where both are available. Clearly, this speculation raises all sorts of questions concerning competition between derivations, which go well beyond the scope of the present paper. We shall therefore not attempt to address them here.

#### 5.4. Island effects

Our final piece of evidence in support of the double derivation analysis involves the fact that the case-matching type of LSF, i.e. the one we have proposed involves movement, is sensitive to islands. In contrast, the case-mismatching type of LSF, which we have analyzed as a scope marking dependency, shows no sensitivity to island effects.

In our questionnaire we included both adjunct and complex NP-islands. Our expectation was that the two types of LSF should give different results: LSF type 1, the base-generation type, with matrix case assignment and definiteness agreement, should not display island effects. LSF type 2, the movement type, with embedded case on the raised NP and no definiteness agreement, was expected to show island sensitivity. These expectations were borne out to a large extent.

Consider first what happens when we let the base-generation type of LSF span across an adjunct island:<sup>36</sup>

- (60) LEPKÉT       mondott,       hogy már vége lett       a versenynek       mikor zöldet találtak.  
butterfly.acc said.3sg.indef that already end became.3sg the competition.dat when green.acc found.3pl  
'He said that by the time they found a green BUTTERFLY, the competition had ended.'

That this is a base-generation example is shown by the definiteness agreement in the matrix clause: the indefinite noun *lepke* 'butterfly' triggers indefinite agreement on the matrix verb. As predicted, this example is considered grammatical by an overwhelming majority of our informants (of both groups of speakers as defined earlier). This fact provides strong support for our earlier analysis of this type of LSF as involving base-generation and no cross-clausal movement. At the other end of the spectrum we find the example in (61), which forms a perfect minimal pair with (60) in that the only difference between the two is the definiteness marking on the verb. The fact that the matrix verb bears definite agreement in (61) shows that this is a movement example, and as the judgement indicates this type of LSF is sensitive to the adjunct island, exactly as expected.

- (61) \*LEPKÉT       mondta,       hogy már vége lett       a versenynek       mikor zöldet találtak.  
butterfly.acc said.3sg.def that already end became.3sg the competition.dat when green.acc found.3pl  
'He said that by the time they found a green BUTTERFLY, the competition had ended.'

The examples in (60) and (61) involve accusative case both in the matrix and in the embedded clause. When we start mixing cases, island effects continue to be absent when the higher LSF-remnant has its case and agreement determined in the matrix clause. This is shown in (62).<sup>37</sup>

- (62) SÜLYT       mondott,       hogy már vége lett       a versenynek       mikor  
weight.acc said.3sg.indef that already end became.3sg the competition.dat when  
jóra       emlékezett.  
correct.subl remembered.3sg.indef  
'He said that by the time he remembered the correct WEIGHT the competition had ended.'

Sensitivity to the complex NP island is less easy to test, because the complex NP also occupies an accusative case position inside the matrix clause. This makes LSF type 1, i.e. the base-generation type, with matrix case assignment to the focused noun, illicit for independent reasons. As we have argued above, a verb can assign accusative case only once, and in a biclausal complex NP structure the complex NP and the focused noun are both competing for accusative case. A relevant example is shown in (63):

<sup>36</sup> This example was presented in the following context: 'A competition is organized in a forest. The task is to collect as many green animals as possible in one hour. After the competition Eve says that Robi from the other group has said that he found a green bug as soon as the competition finished. Peter thinks that Robi said he found a butterfly (and not a bug) and says:'

<sup>37</sup> What is unexpected is the behavior of examples where both LSF-remnants are obliquely marked, as in (i):

- (i) %SÜLYRA       mondta,       hogy már vége lett       a versenynek       mikor jóra       emlékezett.  
weight.subl said.3sg.def that already end became.3sg the competition.dat when correct.subl remembered.3sg.indef  
'He said that by the time he remembered the correct WEIGHT the competition had ended.'

There is a subset of our informants (viz. the movement speakers), who judge this example to be acceptable. That they should find the movement derivation good is not surprising, but that they should not mind about the adjunct island in this case is. Clearly, this selective island sensitivity of movement speakers is a problem for any account. We leave a full exploration of this issue for further research.

- (63) \*ΑΥΤΌΤ hallott olyan híreket, hogy újat vettek.  
 car.acc heard.3sg.indef such news.acc that new.acc bought.3pl.indef  
 '(S)he heard some news that they had bought a new CAR.'

This sentence was indeed rejected by an overwhelming majority of our informants.

LSF type 2, which we analyzed as involving movement, is likewise expected to be bad, since this movement violates the complex NP constraint. This prediction is confirmed, as (64) was also judged to be unacceptable by our informants.<sup>38</sup>

- (64) \*ΑΥΤΌΤ hallotta a hírt, hogy újat vettek.  
 car.acc heard.3sg.def the news.acc that new.acc bought.3pl.indef  
 '(S)he heard the news that they had bought a new CAR.'

In other words, biclausal complex NP examples are bad in both the movement and the base-generation derivation, for all speakers: the movement derivation is ruled out because of subjacency, the base-generation derivation for case reasons.

In an attempt to circumvent this problem, we also tested tricausal structures with a complex NP in the middle clause. We tested six minimal pairs that differed in the definite/indefinite agreement form of the matrix verb.<sup>39</sup> One such pair is illustrated in (65) and (66). As expected, we found a significant contrast between the acceptability of the two types of LSF in this configuration, i.e. LSF type 1 scored better than LSF type 2. However, both sets of sentences were found to be considerably worse than the baseline structures (52) and (53), probably because of the added complexity of the tricausal structure.<sup>40</sup>

- (65) ?ΑΥΤΌΤ mondott, hogy hallotta a hírt, hogy újat vettek.  
 car.acc said.3sg.indef that heard.3sg.def the news.acc that new.acc bought.3pl  
 'He said that he heard the news that they had bought a new CAR.'

- (66) \*?ΑΥΤΌΤ mondta, hogy hallotta a hírt, hogy újat vettek.  
 car.acc said.3sg.def that heard.3sg.def the news.acc that new.acc bought.3pl  
 'He said that he heard the news that they had bought a new CAR.'

In (65) we find an LSF type 1 structure, with indefinite agreement on the main verb indicating matrix case assignment and thus base-generation. In contrast, (66) instantiates a movement derivation since there is no definiteness agreement in the matrix clause. Examples like (65) were judged to be significantly better than examples like (66), i.e. we find the contrast reproduced that we found with the adjunct island cases.

All in all, then, both the adjunct island data and the tricausal complex NP island data clearly support the analyses we have proposed, while the biclausal complex NP island data appear to be unrevealing for independent reasons.

## 6. Conclusion

In this paper, we have discussed constructions of long split focalization (LSF) in Hungarian. We have argued that they provide evidence for a double derivation analysis, and more generally for Den Dikken's typology of A'-dependencies. Simplifying somewhat, A'-dependencies in his view can involve either a movement derivation or a base-generation derivation. We have proposed a similar double derivation analysis for LSF. Type 1 LSF is a base-generated scope-marking dependency, with a noun in the matrix clause that acts as a scope marker triggering the deletion of the corresponding contentful noun in the embedded clause. Type 2 LSF involves movement, i.e. the long focus raising of the NP-part of a complex predication structure into the matrix clause. Both types of LSF were diagnosed by characteristic patterns of case marking and definiteness agreement in the matrix clause. We provided empirical evidence for this double

<sup>38</sup> In (64) the definiteness on the noun that heads the complex NP was changed in comparison with (63) because we did not want informants to rule out the example because of a definiteness agreement mismatch between the verb *hallott(a)* 'hear' and its direct object *a hírt* 'the news'.

<sup>39</sup> The test results for these six pairs are the following:  $t(62) = 19.23$ ,  $p = 8.05 \times 10^{-28}$ ,  $t(62) = 14.09$ ,  $p = 5.31 \times 10^{-21}$ ,  $t(62) = 15.13$ ,  $p = 1.69 \times 10^{-22}$ ,  $t(62) = 12.23$ ,  $p = 3.51 \times 10^{-18}$ ,  $t(62) = 16.88$ ,  $p = 7.08 \times 10^{-25}$ ,  $t(62) = 12.02$ ,  $p = 7.56 \times 10^{-18}$ .

<sup>40</sup> For LSF type 1 we compared the ratings of the baseline example (52) with the average ratings of the six LSF type 1 tricausal complex NP structures, yielding  $t(62) = 17.37$ ,  $p = 1.65 \times 10^{-25}$ . Analogously for LSF type 2, i.e. comparing (53) with the average of the six LSF type 2 complex NP structures, we obtained  $t(30) = 8.39$ ,  $p = 2.33 \times 10^{-9}$ . When testing LSF type 1 in complex NP island configurations, both dialects were taken into account, as both accept the baseline LSF type 1 structure. When testing LSF type 2 in this configuration, only the dialect accepting the baseline LSF 2 structure was considered.

derivation analysis based on a questionnaire filled out by 83 native speakers of Hungarian. The results confirmed the correlation between case and agreement patterns in a way that is consistent with the double derivation analysis we proposed. Second, we have shown there to be systematic speaker variation in the acceptance of LSF sentences: whereas all speakers tolerate base-generation derivations, only a subgroup of them accepts movement derivations. Finally, wh-island phenomena provided additional support for the analysis: the base-generation derivation showed no adjunct island and complex NP island sensitivity, whereas the movement derivation did.

## References

- Brody, M., 1995. *Lexico-Logical Form: A Radically Minimalist Theory*. MIT Press, Cambridge, MA.
- De Cuba, C., Úrögdi, B., 2009. Eliminating factivity from syntax: sentential complements in Hungarian. In: Den Dikken, M., Vago, R. (Eds.), *Approaches to Hungarian: Vol. 11: Papers from the 2007 New York Conference*. John Benjamins, Amsterdam, pp. 29–64.
- Den Dikken, M., 2006. When Hungarians Agree (to Disagree): The Fine Art of ‘Phi’ and ‘Art’. ms., CUNY.
- Den Dikken, M., in press. On the strategies for forming long A’-dependencies: evidence from Hungarian. In: Surányi, B. (Ed.), *Minimalist Approaches to Syntactic Locality*. Cambridge University Press, Cambridge.
- É. Kiss, K., 1987. *Configurationality in Hungarian*. Reidel, Dordrecht.
- É. Kiss, K., 1998. On generic and existential bare plurals and the classification of predicates. In: Rothstein, S. (Ed.), *Events and Grammar*. Kluwer Academic Publishers, Dordrecht, pp. 145–162.
- Gervain, J., 2009. Resumption in focus(-raising). *Lingua* 119 (4), 687–707.
- Haegeman, L., Úrögdi, B., 2010. Referential CPs and DPs: an operator movement account. *Theoret. Linguist.* 36 (2), 111–152.
- Horváth, J., 1995. Partial wh-movement and wh ‘scope markers’. In: Kenesei, I. (Ed.), *Levels and Structures (Approaches to Hungarian 5)*. JATE Press, Szeged, pp. 89–124.
- Horváth, J., 1997. The status of ‘wh-expletives’ and the partial wh-movement construction of Hungarian. *Nat. Lang. Linguist. Theory* 15 (3), 509–572.
- Horváth, J., 1998. Multiple wh-phrases and the wh-scope-marker strategy in Hungarian interrogatives. *Acta Linguist. Hung.* 45 (1–2), 31–60.
- Horváth, J., 2000. On the syntax of ‘wh-scope marker’ constructions: some comparative evidence. In: Lutz, U., Müller, G., von Stechow, A. (Eds.), *Wh-Scope Marking*. Amsterdam, John Benjamins, pp. 271–316.
- Jánosi, A., 2014. *Long Split Focus Constructions in Hungarian with a View on Speaker Variation*. KU Leuven (Ph.D. thesis).
- Kayne, R., 1984. *Connectedness and Binary Branching*. Foris Publications, Dordrecht, Holland.
- Kenesei, I., 1992. Az alárendelt mondatok szerkezete [The structure of subordinate clauses]. In: Kiefer, F. (Ed.), *Strukturális magyar nyelvtan [A Structural Grammar of Hungarian]*. Akadémiai Kiadó, Budapest, pp. 529–714.
- Lipták, A., 1998a. Focus and ECM. In: Cambier-Langeveld, T., Lipták, A., Redford, M. (Eds.), *Proceedings of ConSOLE VI*. HIL, Leiden, pp. 153–168.
- Lipták, A., 1998b. A fókuszemelések egy minimalista elemzése [A minimalist analysis of Hungarian long focus movements]. In: Büky, L., Maleczki, M. (Eds.), *A mai magyar nyelv leírásának újabb módszerei III [New Methods of Describing Present-Day Hungarian III]*. JATE Press, Szeged, pp. 93–115.
- Lipták, A., 2001. *On the Syntax of Wh-Items in Hungarian*. Leiden University (Ph.D. thesis).
- Lipták, A., Saab, A., 2012. Licensing by Inflection Reconsidered. The Ellipsis-Morphology Connection in the Nominal Domain. ms., University of Leiden.
- Marácz, L., 1987. Wh-strategies in Hungarian: data and theory. In: Ruzsa, I., Szabolcsi, A. (Eds.), *Proceedings of the 1987 Debrecen Symposium on Logic and Language*. Akadémiai Kiadó, Budapest, pp. 33–47.
- Moro, A., 2000. *Dynamic Antisymmetry*. MIT Press, Cambridge, MA.
- Nunes, J., 2004. *Linearization of Chains and Sideward Movement*, Linguistic Inquiry Monographs. MIT Press, Cambridge, MA.
- Ott, D., 2011. *Local Instability: The Syntax of Split Topics*. Harvard University (Ph.D. thesis).
- Puskás, G., 2000. *Word Order in Hungarian: The Syntax of A’-Positions*. John Benjamins, Amsterdam.
- Rackowski, A., Richards, N., 2005. Phase edge and extraction: a Tagalog case study. *Linguist. Inquiry* 36 (4), 565–599.
- Richards, N., 1998. The principle of minimal compliance. *Linguist. Inquiry* 29 (4), 599–629.