Long split focus constructions in Hungarian: a base-generation approach*

Adrienn Jánosi
CRISSP/HUB/KUL
janosiad@yahoo.com

TOPIC OF THIS TALK: long-distance split focalization (LSF) in Hungarian

(1) **AUTÓT$_{Foc}$** mondott/mondta hogy **ÚJAT$_{Foc}$** vett.
Car.ACC said.3Sg.Indef./Def. that new.ACC bought.3Sg.Indef
'(S)he said that(s)he had bought a new CAR.'

MAIN CLAIMS:
1. LSF is derived via base generation, not movement
2. speaker variation pertaining to LSF can be reduced to whether or not a speaker allows case concord (Den Dikken 2010)

CENTRAL DATA: a questionnaire of 91 questions distributed among 83 native speakers

OUTLINE OF THE TALK

1. Long-distance split focalization (LSF): the basic data

→ LSF can be classified according to the following three criteria:

(i) The case of the higher DP
There are two options for case-marking the higher DP (i.e. *AUTÓT* in (1)): case is either assigned by the matrix verb (2) or by the embedded verb (3). (Note that *hall* 'hear' takes an accusative complement and *örül* 'be pleased' a dative one.)

(2) **AUTÓT$_{Foc}$** hallott hogy **ÚJNAK $_{Foc}$** örülnének.
Car.ACC heard.3Sg.**Indef.** that new.DAT be.pleased.Cond.3Pl.
'(S)he heard that they would be pleased with a new car.'

(3) **AUTÓNAK $_{Foc}$** hallotta hogy **ÚJNAK $_{Foc}$** örülnének.
Car.DAT heard.3Sg.**Def.** that new.DAT be.pleased.Cond.3Pl.
'(S)he heard that they would be pleased with a new car.'

(ii) Object definiteness agreement in the matrix clause
A transitive matrix verb can either agree in definiteness (i.e. show indefinite agreement as in (4)) or not agree (and display definite agreement as in (5)) with the higher (indefinite) DP in LSF.

(4) **AUTÓT$_{Foc}$** mondott hogy **ÚJAT $_{Foc}$** vett.
Car.ACC said.3Sg.**Indef.** that new.ACC bought.3Sg.Indef
'(S)he said that (s)he had bought a new CAR.'

(5) **AUTÓT$_{Foc}$** mondta hogy **ÚJAT $_{Foc}$** vett.
Car.ACC said.3Sg.**Def.** that new.ACC bought.3Sg.Indef
'(S)he said that (s)he had bought a new CAR.'

* I'm indebted to Jeroen van Craenenbroeck, Guido Vanden Wyngaerd, Anikó Lipták and Marcel Den Dikken for discussions and encouragement. All remaining errors are my own.

correlation between (i) and (ii): indefinite agreement correlates with a case ending on the higher DP that is assigned by the matrix verb (see (2) and (4)) while definite agreement correlates with a case ending that is determined by the embedded verb (see (3) and (5)).

(iii) The case of the lower DP

The case of the lower DP (i.e. *újat* in (1)) is always determined by the embedded verb (i.e. ACC in (4)/(5) and DAT in (2)/(3)).

→ The criteria in (i)-(iii) yield a fourfold classification of LSF structures:

	Case of the higher DP	**Obj.agr. on the matrix V**	**Case of the lower DP**
a.	ACC	definite	ACC
b.	ACC	indefinite	ACC
c.	ACC	indefinite	OBL
d.	OBL	definite	OBL

table 1. The four surface types of LSF

2. Theoretical background: movement vs. base generation

traditional view: long focus constructions involve movement of the focused XP from its base position (via intermediate landing sites) to specFocP in the matrix clause (É.Kiss 1987, Lipták 1998, Puskás 2000)

(6) **AUTÓT** mondott hogy vett.
car.ACC said.3Sg.Indef. that bought.3Sg.Indef.
'He said that he had bought a car.'

(7) **A FIÚT** mondtad hogy láttad.
the boy.ACC said.2Sg.Def. that saw.2Sg.Def.
'You said that you had seen the boy.'

more recent view: in addition to the movement derivation of long focus constructions, a group of speakers derives such structures by base-generating the focused DP in the matrix clause. (Gervain 2007, Den Dikken 2010)

Gervain (2007): native speakers of Hungarian can be subdivided into 'movement' and 'non-movement' speakers based on the case and agreement patterns found in long-distance focus movement of quantified subjects

(8) Az összes **LÁNYT** mondtad hogy jön.
The all girl.Sg.ACC said.2Sg.Def .that come.3Sg
'You said that all the girls were coming.'

→ accusative marking on the focus, singular agreement in the embedded clause: **acceptable for all speakers**

(9) Az összes **LÁNY** mondtad hogy jön.
The all girl.Sg. said.2Sg.Def. that come.3Sg
'You said that all the girls were coming.'

→ nominative marking on the focus, singular agreement in the embedded clause: **acceptable only for movement speakers**

(10) Az összes **LÁNYT** mondtad hogy jönnek.
The all girl.Sg.ACC said.2Sg.Def. that come.3Pl.
'You said that all the girls were coming.'

→ accusative marking on the focus, plural agreement in the embedded clause: **acceptable only for non-movement speakers**

(11) Az összes **LÁNY** mondtad hogy jönnek.
The all girl.Sg. said.2Sg.Def. that come.3Pl.
'You said that all the girls were coming.'

→ nominative marking on the focus, plural agreement in the embedded clause: **unacceptable for all speakers**

	case on the focus/embedded agreement			
	ACC/SG	NOM/SG	ACC/PL	NOM/PL
movement speakers	ok	ok	*	*
non-movement speakers	ok	*	ok	*

table 2. Movement and non-movement speakers according to Gervain (2007)

3. Empirical results from the questionnaire[1]

3.1 Set-up and methodology

Date: April-March 2011

Venue: College of Nyíregyháza

Number of informants: 83

Basic profile of the informants:

Age: 19-25

Major: English/Marketing

Place of living: 80% in North-Eastern Hungary

Number of test items: 91 (including 32 fillers on backward gapping)
Type of questions: acceptability judgements on a scale of 1 (ungrammatical) to 5 (fully grammatical)

Note on the grammaticality scale: given that even baseline LSF-structures are marked compared to the neutral expletive-associate constructions (see below), the threshold for acceptability is taken to be 3.

[1] The entire questionnaire, fully glossed and with a summary of the scores, is available at http://crissp.be/adrienn/questionnaire_lsf.pdf

3.2 Revisiting the Gervain-sentences

→ the split between movement and non-movement speakers was not entirely confirmed by the young North-East-Hungarian informants of the questionnaire

similarities:
- ACC/SG (cf. (8)) is acceptable for all speakers
- NOM/PL (cf. (11)) is unacceptable for all speakers
- NOM/SG (cf. (9)) is subject to speaker variation

difference:
- ACC/PL (cf. (10)) is acceptable for all speakers

	case on the focus/embedded agreement			
	ACC/SG	NOM/SG	ACC/PL	NOM/PL
NE Hungarian informants	ok	%	ok	*

table 3. the Gervain-examples according to the questionnaire

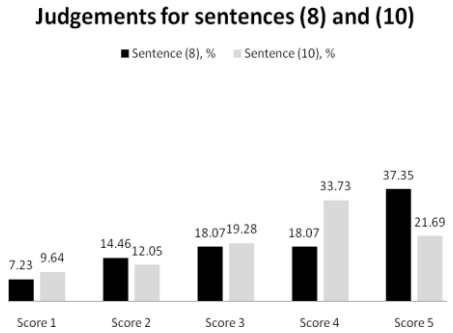


Figure 1. Comparison of scores for (8) and (10)

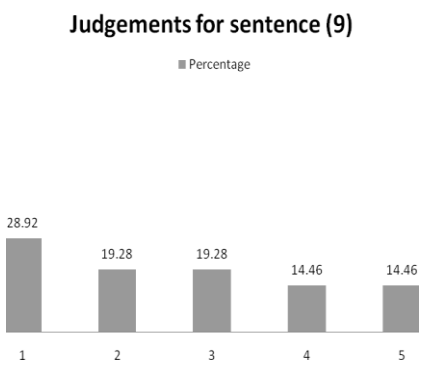
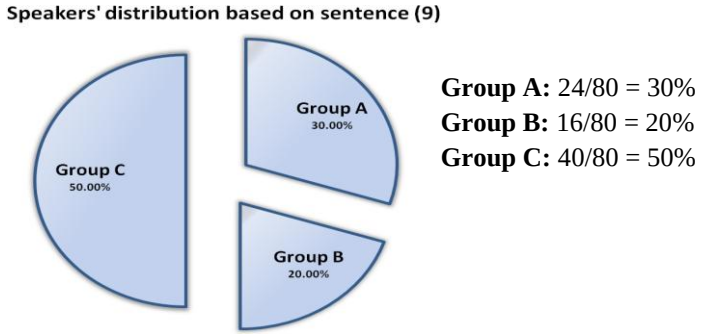


Figure 2. Distribution of scores for (9)

proposal: based on the score distribution for example (9), I split up my informants into three groups:

Group A: score for (9) ≥ 4 (grammatical)
Group B: score for (9) = 3 (not fully grammatical)
Group C: score for (9) ≤ 2 (ungrammatical)



Group A: 24/80 = 30%
Group B: 16/80 = 20%
Group C: 40/80 = 50%

figure 3. Proportional representation of groups A, B and C[2]

[2] 3 out of the total of 83 informants did not indicate their score concerning this sentence and were excluded from the analysis of the results.

methodological note: in order to investigate the speaker variation raised by (9), I henceforth focus on the two extreme cases, i.e. groups A and C

hypothesis: the difference between group A and group C seems to boil down to whether or not the focused XP in the matrix clause can bear the case assigned in the embedded clause → if correct, this difference should show up elsewhere in the questionnaire as well

(2) **hypothesis:**
When case on the focused element in the matrix clause is determined by the embedded clause, group A should score the sentence better than group C.

3.3 Revisiting the basic LSF-patterns

	Case of the higher DP	**Obj.agr. on the matrix V**	**Case of the lower DP**	**How many more informants of *group A* than of *group C* find the structure grammatical? (%)**
a.	ACC	definite	ACC	**16,66%** (A:54,16%, C:37,5%)
b.	ACC	indefinite	ACC	**12,5%** (A: 87,5%, C:75%)
c.	ACC	indefinite	OBL	**24,16%** (A:66,66%, C:42,5%)
d.	OBL	definite	OBL	**21,66%** (A:41,66%, C:20%)

Table 4. The 4 types of LSF with judgement patterns

→ the difference in judgements in a/d (case determined by the embedded clause) is **expected** under the hypothesis in (2)

→ the difference in judgements in b and c is **not expected** under the hypothesis in (2) → see section 5 for discussion

3.4 Islands

(i) complex NP-island with case determined by the embedded clause:

(12) **Autót** 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.Indef.
'He said that he heard the news that they had bought a new car.'

difference in judgement between group A and group C: 20,83%
(A: 45,83%, C: 25%)

(ii) complex NP-island with case determined by the matrix clause:

(13) **Autót** 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.Indef.
'He said that he heard the news that they had bought a new car.'

difference in judgement between group A and group C: 4,16%
(A: 54,16%, C: 50%)

(iii) adjunct island with case determined by the embedded clause:
(Context used in the questionnaire: "... Az említett versenyen a lepkék különböző adatairól is kapnak kérdést a versenyzők. Éva szerint az ellenfél csapatából Robinak pont akkor jutott eszébe a jó válasz a lepkék hosszáról, amikor már letelt a gondolkodási idő. Péter nem hosszra, hanem súlyra emlékszik, és ezt mondja:")

(14) **SÚLYRA** mondta hogy már vége lett a versenynek mikor **jóra** emlékezett.
Weight.**ONTO** said.3Sg.**Def .** that alreadyend became.3Sg. the competition.Dat when correct.**ONTO** remembered.3Sg.Indf
'He said that by the time he remembered the correct weight the competition had ended.'

difference in judgement between group A and group C: 33,33%
(A: 70,83%, C: 37,5%)

(iv) adjunct island with case determined by the matrix clause:

(15) **SÚLYT** mondott hogy már vége lett a versenynek mikor **jóra** emlékezett Weight.**ACC** said.3Sg.**Indef.** that already end became.3Sg. the competition.Dat when correct.**ONTO**

remembered.3Sg.Indef.
'He said that by the time he remembered the correct weight the competition had ended.'

difference in judgement between group A and group C: 2.5%,
(A: 50%, C: 52,5%)

3.5 Data summary

→ two groups of speakers can be discerned in the questionnaire data; they differ in whether or not they allow the focused XP in the matrix clause to bear the case assigned in the embedded clause

4. The analysis

4.1 Starting assumption: base generation

→I take the fact that LSF can cross island boundaries (cf. example (15), repeated below) to be an indication that this construction involves base-generation, not movement.

(16) **SÚLYT** mondott hogy már vége lett a versenynek mikor **jóra** emlékezett
Weight.**ACC** said.3Sg.**Indef.** that already end became.3Sg. the competition.Dat when correct.**ONTO** remembered.3Sg.Indef.
'He said that by the time he remembered the correct weight the competition had ended.'

→I assume that LSF is derived from the expletive-associate construction (cf. 17)), the difference being that in (18) the expletive is spelled out as the core of the focused NP in the embedded clause

(17) **Azt** mondta hogy **ÚJ AUTÓT$_{Foc}$** vett.
Expl.ACC said.3Sg.**Def.** that new car.ACC bought.3Sg.Indef.
'(S)he said that (s)he had bought a new car.'

(18) **AUTÓT$_{Foc}$** mondott/mondta hogy **ÚJAT$_{Foc}$** vett.
Car.ACC said.3Sg.**Indef./Def.** that new.ACC bought.3Sg.Indef
'(S)he said that(s)he had bought a new CAR.'

supporting evidence: LSF is incompatible with the matrix expletive:

(19) *Azt **AUTÓT$_{Foc}$** mondott/mondta hogy **ÚJAT$_{Foc}$** vett.
Expl.ACC Car.ACC said.3Sg.**Indef./Def.**that new.ACC bought.3Sg.Indef
'(S)he said that(s)he had bought a new CAR.'

4.2 Background for the analysis: Den Dikken (2010)'s hidden scope marking

Den Dikken (2010): long-distance focus movement is in fact hidden scope marking: the higher DP is base-generated in the position of the expletive, that is, in SpecVP. From there, under closest c-command, the higher DP and the lower DP (situated in the lower Spec,CP) engage in a ***concord*** relation.

Example:

(20) a. ?hány lány-t akar-sz, hogy eljöjjön?
how.many gir.l.ACC want.2Sg.Indef. that PV-come-3SG
b. SM+FFDP=*hány lányt* ... [*v*P ~~SM+FFDP~~ [*v*P *v* [VP *akar-sz* ~~SM+FFDP~~ [$_{CP}$ DP=0 *hogy* ~~DP~~ *eljöjjön*]]]]
(Den Dikken 2010: 10)

Main steps of the derivation:

- the wh-constituent (lower DP) checks NOM case in the embedded clause. Then
- it A'-moves to the SpecCP position of the embedded clause where
- it engages in a concord relation with the wh-scope marker generated in the Spec of the matrix VP.
- The lower DP deletes as all of its features are shared (under concord) with the wh-scope marker *except for case*.

→ As the scope marker has an accusative case feature of its own, the higher DP is spelled out bearing the accusative case marker.
(Den Dikken 2010)

4.3 The proposal: case concord vs. basic concord

→ the relation between the two DP-portions in LSF is established through concord:

(21) **Concord:** Transfer of features from one constituent (DP) to another constituent of the same category (from a lower contentive to a higher semantically empty scope marking element) resulting in overlapping features to the extent of near-identity. Near-identity causes the contentive (lower) DP to delete (under c-command) while the scope marking element that was originally semantically empty surfaces in the form of the contentive.

Note: this corresponds to Den Dikken's (2010) 'full concord' dependency with the difference that a c-command relation is proposed instead of 'closest c-command' between the two DPs.

→ there are two types of concord:

1. Basic concord: concord not involving case, but enough overlapping features for the lower DP to delete.
2. Case concord: same as basic concord, but also involving case

Conditions on basic concord

1. The matrix DP must be an argument of the embedded verb (subject, object, prepositional object/oblique argument).
2. Featural near-identity between the higher and the lower DP
3. C-command relation between the two DPs (i.e. The beneficiary DP must c-command the 'donor' DP)
4. Availability of the expletive-associate (standard) variant of the LSF construction

Conditions on case concord

1. The presence of basic concord.
2. Lack of (definiteness) agreement between the matrix verb and the higher DP (as a consequence, it cannot be case-marked, either).
3. The lower DP must be in SpecFocP (or higher) in the embedded cause.
4. Case concord is in complementary distribution with
   - subject-verb (number, person)
   - object-verb (definiteness)
   - other agreement (involving case assignment).

→ **Case concord**, therefore, is an alternative way of agreement when the higher DP is not a real argument of the matrix verb AND does not agree with the verb of its own clause.

speaker variation: the difference between the two groups now reduces to the presence or absence of case concord

	types of concord	
	basic concord	**case concord**
group A	√	√
group C	√	*

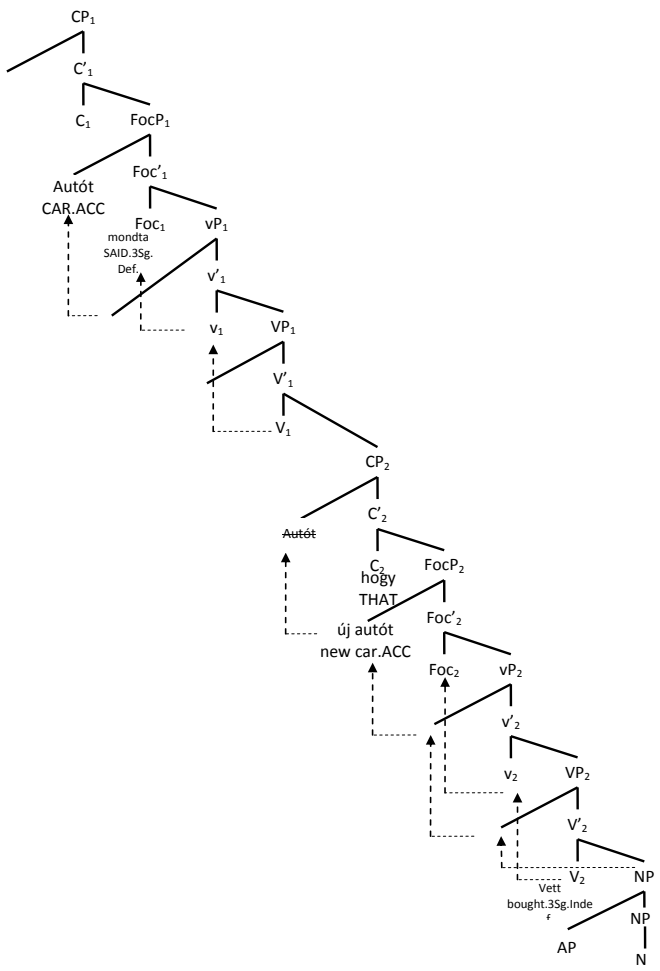
Table5. Speaker variation reduced to case concord

4.4 Sample derivations

4.4.1 LSF with case concord

(22) **AUTÓT** mondta hogy újat vett.
Car.ACC said.3Sg.**Def.** that new.ACC bought.3Sg.Indef
'(S)he said that (s)he had bought a new CAR.'

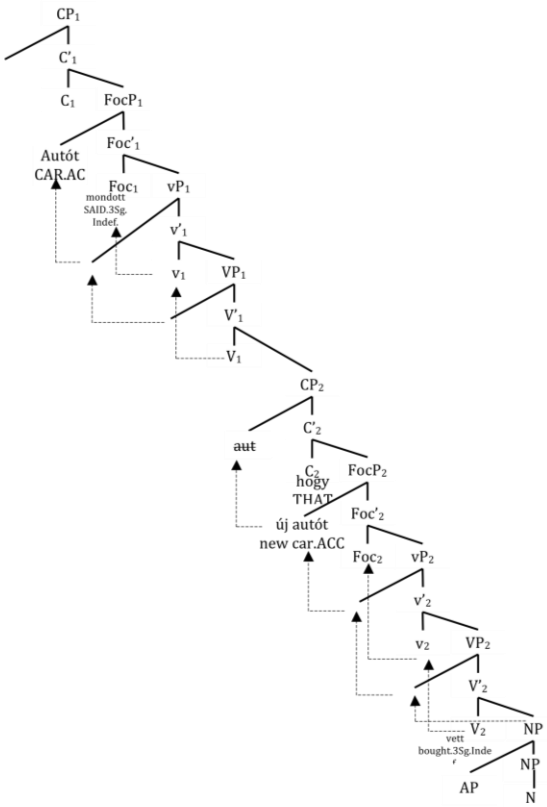
(23)



4.4.2 LSF with basic concord

(24) **AUTÓT** mondott hogy újat vett.
Car.ACC said.3Sg.**Indef.** that new.ACC
bought.3Sg.Indef
'(S)he said that (s)he had bought a new CAR.'

(25)



• The higher DP is base-generated in Spec vP_1

• Phrasal movement of the lower DP to Spec$FocP_2$

• Subphrasal movement of the core of the lower DP to SpecCP_2

• *Concord involving case* between the two DPs

• Deletion of the lower DP

• The higher DP is base-generated in Spec VP_1

• Phrasal movement of the lower DP to Spec$FocP_2$

• Subphrasal movement of the core of the lower DP to SpecCP_2

• *Concord except for case* between the two DPs

• Deletion of the lower DP

5. Extensions and problems

5.1 The ACC/OBL-pattern in group C

recall: one of the four basic LSF-patterns did not follow from the case concord hypothesis:

	Case of the higher DP	**Obj.agr. on the matrix V**	**Case of the lower DP**	**How many more informants of *group A* than of *group C* find the structure grammatical? (%)**
c.	ACC	indefinite	OBL	**24,16%** (A:66,66%, C:42,5%)

Table 6. The 4 types of LSF with judgement patterns (partially)

→ the difference in judgements in c is **not expected** under the hypothesis in (2)

hypothesis: the markedness of this pattern is due to the fronting of the oblique in the embedded clause

supporting evidence: in LSF where the adjective remains in postverbal position, the data distribution is exactly as expected:

example:

(26) **AUTÓT** hallott hogy számítanak egy újra.
Car.ACC heard.3Sg.Indef. that count.3Pl. a new.ONTO
'(S)he heard that they had bought a new car.'

	Case of the higher DP	**Obj.agr. on the matrix V**	**Case of the lower DP**	**How many more informants of *group A* than of *group C* find the structure grammatical? (%)**
a.	ACC	definite	ACC	**32,5%** (A:50%, C:17,5%)
b.	ACC	indefinite	ACC	**7%** (A: 62%, C: 55%)
c.	ACC	indefinite	OBL	**2%** (A:50%, C:52%)
d.	OBL	definite	OBL	**25,83%** (A:33,33%, C:7,5%)

Table 7. Judgement patterns for the 4 types of LSF with the lower DP in postverbal position

5.2 Embedded plural marking in the Gervain-examples

→ if group A allows for case concord, why is Gervain's fourth example (repeated below) universally disallowed?

(27)* Az összes **LÁNY** mondtad hogy jönnek.
The all girl said.2Sg.Def. that come.3Pl.
'You said that all the girls were coming.'

hypothesis: embedded plural marking is an indication of a resumptive pronoun (*pro*), but for case concord to take place (cf. the nominative marking on *lány*), the embedded clause has to contain a full DP

6. Summary and conclusions

It has been shown that - based on complex NP island effects and adjunct island effects – LSF constructions can be thought of as instances of base-generation rather than movement.
Concord between the two DPs can account for case mismatches (basic concord) and obligatory case matches (case concord), as well.
Basic concord proved to be a generally available strategy, while speaker variation concerning LSF can be reduced to the presence vs. absence of the *case concord* strategy.

References

Den Dikken, Marcel 2010. On the strategies for forming long A'-dependencies: Evidence from Hungarian

É. Kiss, Katalin 2002. *The Syntax of Hungarian*, CUP

Gervain, Judit 2007. 'Resumption in focus(-raising)', in Lingua 2007.11.008, Elsevier B.V.

Kenesei, István 1994. Subordinate clauses. In Ferenc Kiefer & Katalin É. Kiss (eds), *The syntactic structure of Hungarian. Syntax and Semantics 27.* New York: Academic Press. 275-354.

Lobeck, Anne 2005. 'Ellipsis in DP', in Everaert, Martin and van Riemsdijk, Henk (eds.) 2005, *The Blackwell Companion to Syntax*, Blackwell Publishing, Malden MA.

Lipták, Anikó 1998. 'Focus and ECM' Approaches to Hungarian long focus movements.

Lipták, Anikó. 2001. *On the syntax of* wh-*items in Hungarian*. Ph.D. dissertation, Leiden University/HIL.

Ott, Dennis 2009. 'Multiple NP split. A Distributed Deletion analysis', in: Groninger Arbeiten zur germanistischen Linguistik 48, pp. 65-80.

Rizzi, Luigi. 2006. On the form of chains: Criterial positions and ECP effects. In Lisa Cheng & Norbert Corver (eds), *WH-movement: Moving on*. Cambridge, MA: MIT Press. 97–133.

Puskás, Genovéva 2000. *Word Order in Hungarian: The Syntax of A'-positions*, Amsterdam: John Benjamins.

Van Hoof, Hanneke 2005. 'Split topicalization', in Everaert, Martin and van Riemsdijk, Henk (eds.) 2005, *The Blackwell Companion to Syntax*, Blackwell Publishing, Malden MA.

Van Riemsdijk, Henk 1989. 'Movement and Regeneration', in: *Dialect Variation and the Theory of Grammar*, Paola Beninca (ed.), Dordrecht: Foris, pp. 105-136.