Adrienn Jánosi: Long-distance split focalization in Hungarian: a base-generation approach

Introduction and main claims
Although long focus movement has received continuous attention in the Hungarian generative literature over the past decades (É.Kiss 1987, Lipták 1998, Puskás 2000), it has recently become a highly debated issue. Based on new data, Gervain (2007) and Den Dikken (2010) show that 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. This paper uses another novel set of data, long focus constructions involving split bare NPs (LSF for short) as illustrated in (1), to settle debated issues concerning long focus movement:

(1) AUTÓT\(_{\text{Foc}}\) mondott/mondta hogy ÚJAT\(_{\text{Foc}}\) vett. Car.ACC said.3Sg./Def. that new.ACC bought.3Sg

'(S)he said that(s)he had bought a new CAR.'

Based on a large-scale paper-and-pencil questionnaire study of LSF constructions involving 83 speakers, the talk will argue for the following claims:

1. Wh-movement cannot cross clause boundaries in Hungarian.
2. Base-generation with (case) concord (cf. Den Dikken (2010)) is a widely applied strategy to establish long A'-dependencies in non-standard Hungarian.
3. Speaker variation can be reduced to the following two facts:
   a. 'Movement speakers' (cf. Gervain 2007) tend to apply the movement strategy also in long-distance cases which are principally derivable by movement (i.e. 'long-extraction'). This, however, causes the derivation to fail (see claim (1)). b. 'Non-movement speakers', on the other hand, tend to apply the case-concord strategy in such cases, which results in an acceptable surface string since this strategy does not involve long-distance movement.

A considerable part of all speakers accept LSF constructions involving complex NP islands, which constitutes empirical evidence in favour of claims 1 and 2.

Generalizations about LSF
There are four surface patterns of LSF when it comes to definiteness agreement and case marking. This classification is based on the following three factors:

1. 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 (3) or by the embedded verb (4).

(3) AUTÓT\(_{\text{Foc}}\) hallott hogy ÚJNAK\(_{\text{Foc}}\) örülhénék. Car.ACC heard.3Sg.Indef. that new.DAT be.pleased.Cond.3Pl.

'(S)he heard that they would be pleased with a new car.'

(4) AUTÓNAK\(_{\text{Foc}}\) hallotta hogy ÚJNAK\(_{\text{Foc}}\) örülhénék. Car.DAT heard.3Sg.Def. that new.DAT be.pleased.Cond.3Pl.

'(S)he heard that they would be pleased with a new car.'

2. Object definiteness agreement in the matrix clause A transitive matrix verb can either agree in definiteness (i.e. 'indefinite agreement' (5)) or not agree (i.e. 'definite agreement' (6)) with the higher (indefinite) DP in LSF.

(5) AUTÓT\(_{\text{Foc}}\) mondott hogy ÚJAT\(_{\text{Foc}}\) vett. Car.ACC said.3Sg.Indef. that new.ACC bought.3Sg

'(S)he said that (s)he had bought a new CAR.'

(6) AUTÓT\(_{\text{Foc}}\) mondtta hogy ÚJAT\(_{\text{Foc}}\) vett. Car.ACC said.3Sg.Def. that new.ACC bought.3Sg

'(S)he said that (s)he had bought a new CAR.'
Indefinite agreement correlates with a case ending on the higher DP that is assigned by the matrix verb (see (3) and (5)) while definite agreement correlates with a case ending that is determined by the embedded verb (see (4) and (6)).

(3) 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 (5)/(6) and DAT in (3)/(4)). The above facts yield a fourfold classification of LSF structures, summarized in table 1.

<table>
<thead>
<tr>
<th>Case of the higher DP</th>
<th>Obj.agr. on the matrix V</th>
<th>Case of the lower DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ACC</td>
<td>definite</td>
<td>ACC</td>
</tr>
<tr>
<td>b. ACC</td>
<td>indefinite</td>
<td>ACC</td>
</tr>
<tr>
<td>c. ACC</td>
<td>indefinite</td>
<td>OBL</td>
</tr>
<tr>
<td>d. OBL</td>
<td>definite</td>
<td>OBL</td>
</tr>
</tbody>
</table>

**Theoretical analysis**

The data of LSF point to two conclusions:

1. LSF constructions in which the higher DP agrees with the matrix verb (line b and line c in table 1) involve two DPs base-generated in their own clause along the lines of the ‘concordial scope marking dependency’ without (case) concord introduced in Den Dikken (2010). The main argument for this position comes from case mismatches. As is clear from table 1, case mismatches between the higher and the lower DP are only allowed in cases where the higher DP agrees in (in)definiteness with the matrix verb. Since A'-chains can only have one case, in contexts where case mismatches are allowed a movement analysis is excluded.

2. LSF constructions in which the higher DP does not agree with the matrix verb (i.e. line a and d in table 1) also involve two base-generated DPs which, however, are linked by case concord (c.f. Den Dikken 2010). Evidence for this comes from complex NP islands (7).

(7) AUTÓRA mondta hogy azt a hírt hallotta

hogy újra számítanak

CAR.ONTO said.3Sg.Def. that Dem.ACC the news.ACC heard.3Sg.Def. that new.ONTO count.3Pl.Indef.

‘(S)he said that (s)he heard the news that they expect a new car.’

The significantly increased acceptance of structures of line d with a complex NP island (cf. (7)) compared to LSF structures without a complex NP island (cf. (4)) shows that when there is a complex NP dividing the two DPs (and hence extraction is excluded), base-generation with case concord becomes a widely applied strategy to form an A’-dependency between the two DPs divided by a clause boundary.

**Speaker variation**

LSF structures support the finding of Gervain (2007) that speakers of Hungarian fall into two groups when it comes to long-extraction: so-called movement speakers and ‘non-movement’ speakers.

Movement speakers do not accept structures like a and d as they attempt to derive such structures by long-extraction, which fails because wh-movement is clause-bound in Hungarian.

Non-movement speakers, however, accept these examples (i.e. line a and d) because they base-generate the two DPs in their respective clause and link them via case concord (cf. Den Dikken 2010)