Morphosyntactic coding and local coherence in Hungarian

Kata Balogh
Heinrich-Heine-Universität Düsseldorf

The main aim of this paper is introducing a study on the relation of the morphosyntactic coding and the discourse functions of two strategies in Hungarian: *zero (argument) coding* and *structural topic marking*. Both are widely discussed in the Hungarian linguistic literature dating back to the early '80s, however, both phenomena are almost exclusively discussed at the sentence-level syntax, in terms of the 'pro-drop property' and the 'topic position' respectively. An investigation of their role in discourse interpretation and discourse structure is generally missing. This study contributes to this latter issue.

Both of our target strategies are associated with some kind of *salience*, which is understood as referring to things that "*stand out from the ground, can be easily recognized, are in the focus of attention, or foremost to a person's state of mind*" (Chiarcos et al. 2011, page 1; see also von Heusinger 1997). Structural topicalization corresponds to some kind of highlighting, identifying a referent for predication, while zero coded referents are mostly the ones that are identifiable without effort, hence already in the center of attention. The notion of salience plays an important role in various areas of linguistic studies from anaphoric relations to discourse relations. For our purposes, the most important aspects are the role that salience plays in information structure and the topic-comment division, as well as its role in the realization of discourse referents within the discourse. We are particularly interested in the characterization and modeling of the two in terms of *discourse processes*, their differences and also their possible interplay. To achieve these research goals, we analyze and model our data in the framework of *Centering Theory* (Grosz et al., 1983, 1995; Walker et al., 1998).

1 The target strategies

1.1 Structural topic marking

In Hungarian, the word order in the preverbal field is primarily determined by the topic-comment distinction. The discourse-semantic function *topic* is expressed structurally, placed in a clause-initial position. Structural marking of the topic in Hungarian can also be shown by the fact that categorial and thetic judgements are structurally different. Thetic judgments are expressed with a verb-initial structure (1), while in categorial judgments (2), a constituent is selected and placed in the clause-initial, 'topic position' (e.g., É. Kiss, 2004).

(1) *Be-szaladt egy kutya az irodá-m-ba.*
VPRT-ran[3SG] a dog the office-PS1SG-ILL
'A dog ran into my office.'

(2) a. *A fiú kerget egy kutyá-t.*
the boy chase[3SG] a dog-ACC
'The boy is chasing a dog.'

b. *A kutyá-t kerget-i egy fiú.*
the dog-ACC chase-3SG.D a boy
'A boy is chasing the dog.'

Constituents in the 'topic position' are restricted by the requirements of being referential and specific, where specificity is taken as a familiarity condition, the referent of the constituent in the 'topic position' must be known/already present in the discourse (e.g., Kálmán, 2001; É. Kiss, 2004). It is generally assumed that a filled 'topic position' expresses the sentence topic, considered as *aboutness topic* (= the logical subject of predication).[1]

[1]Note that the contrastive topic occupies the same structural position, but it is associated with a distinguished prosody and, furthermore, it does not come with a referentiality constraint.

1.2 Zero coding

In Hungarian, the subject pronoun is generally omitted in unmarked sentences (3a). This phenomenon is referred to as the *pro-drop property* in generative (syntactic) approaches (e.g., Jaeggli & Safir, 1989; Koeneman & Zeijlstra, 2021), and it is present in a great number of languages (e.g., Italian, Chinese, Spanish, Hungarian), while generally not available in English. It is widely accepted that there are different types of 'pro-drop' (e.g., classical/consistent pro-drop versus radical/discourse pro-drop), however, a precise characterization of these types and their typological distribution are still under discussion (see Koeneman & Zeijlstra, 2021).

In Hungarian, the subject pronoun is generally left out in unmarked sentences. When the subject pronoun is overt, it indicates contrast (3b). Next to the subject pronoun, the object pronoun can also be omitted under given circumstances. However, there is an asymmetry: an overt object pronoun itself does not signal contrast, the sentence in (3a) receive the same interpretation with or without overtly expressing the direct object.

(3) a. *Lát-t-ad (ő-t).*
see-PST-2SG.D (s)he-ACC
'You saw him/her.'

b. *Te lát-t-ad (ő-t).*
you see-PST-2SG.D (s)he-ACC
'[You]^{CT} saw him/her.'

Zero argument coding in Hungarian is often considered as 'classical pro-drop', but a detailed discussion of the licensing conditions is missing. If we consider the relevant morphosyntactic properties proposed by Jaeggli & Safir (1989), Koeneman & Zeijlstra (2021) and Neeleman & Szendrői (2007), we find phenomena specific for both agreement-based (classical) and discourse-based (radical) pro-drop. Similar to Japanese and Chinese, various pronouns can be omitted, which is a characteristic property of radical pro-drop. In Hungarian, next to the subject, the object (3a) and the possessive pronoun (4) is also often omitted.

(4) *Peti fel-rakta a kalap-já-t a fej-é-re.*
Pete VPRT-put the hat-PS.3SG-ACC the head-PS.3SG-SUB
'Pete put his hat on his head.'

In Hungarian, we encounter more freedom with respect to zero coding than in classical pro-drop languages (e.g., Italian, Spanish), but this freeness is more restricted than in radical pro-drop languages (e.g., Japanese, Chinese). In this paper, we do not target the issue of the licensing conditions and the type of 'pro-drop' in Hungarian. We are rather interested in discourse interpretation and modeling, and the discourse processes behind the phenomenon of zero coding. Despite the rich agreement and verbal inflection in Hungarian, which points to the direction of classical pro-drop, we rather inquire into the aspects that possibly relate Hungarian to discourse-based pro-drop. We are not investigating what morphosyntactic features it make possible to omit an argument, but pose the question: since this strategy is available in Hungarian, what are its functions (or roles) in the discourse.

2 The study

Our primary target of investigation is which discourse processes correspond to zero coding and structural topic. Our starting point is based on the generalizations of Comrie (1999) and Van Valin (2005), who argue that less marked expressions (e.g., zero coding) lead to topic continuity, while overt marking (e.g., structural topic) leads to changes. Comrie's (1999, p. 342) explanation is that "*in the extended domain, the expectation is for referential continuity, or, as it is often called in the literature, topic continuity*". Van Valin (2005) similarly argues that zero coding represents the most unmarked topic and as such it marks continuing topic. Our initial hypothesis is that the general strategy in Hungarian is to mark continuation by zero elements, and topic shifts by structural topic (see 5), and the expectation is that this pattern is verifiable on a larger amount of data.

(5) A kisfiú kergette a béká-t, *The boy was chasing the dog,*

a. *aztán* ∅ / *#(a kisfiú) el-ugrott* *egy faág-ra.*
and.then ∅ / the boy PRT-jumped[3SG] a branch-SUB
'and then he(=boy) / #the boy jumped away to a branch.' (continue)

b. *aztán* *a béka / az / ő el-ugrott* *egy faág-ra.*
and.then the frog / that / he PRT-jumped[3SG] a branch-SUB
'and then the frog / that one / he(=frog) jumped away to a branch.' (shift)

2.1 Methodology and data

The examples in (5) illustrate the different behaviour of zero coding and structural topic across sentences. In order to check our hypothesis on a larger scale, we investigate data from Hungarian narratives within the framework of *Centering Theory* (Grosz et al., 1983, 1995; Walker et al., 1998). This framework provides the adequate ways for our principal aims, as it is developed for the purposes of exploring the relations between relative salience and the corresponding linguistic expressions. Furthermore, this theory comes with an independent, self-contained terminology, which avoids the terminological diversity and confusion that otherwise occurs with respect to research on topicality and salience.

Centering Theory is the local-level component of the theory of discourse structure proposed by Grosz and Sidner (1986). It relates the *center of attention* within an utterance, the choice of referential expressions and the perceived local coherence between individual utterances. The discourse entities (≈ referents) that are salient at a given point of the discourse are called *centers*. At each utterance, two types of centers are distinguished: the set of *forward-looking centers* (CF) and a distinguished single *backward-looking center* (CB). The latter one establishes the connection to the previous utterance. The CB of the utterance roughly corresponds to the notion of (*aboutness*) *topic* in other theories. The set of forward-looking centers constitutes the local attention state, which is updated at each utterance. This set is partially ordered, based on a ranking of the relative salience of its elements. The ranking indicates the relative likelihood of a referent to be the highest ranked (≈ most salient) center in the subsequent utterance. The highest ranked element within the set of CFs is called the *preferred center* (CP). The CP is the center that is predicted to be the CB of the next utterance. Determining the ranked CF-set and the CB of a given utterance is driven by different constraints. Furthermore, the ranking of the CF-set is subject to language specific properties.

Between the subsequent utterances there are four transition types defined: CONTINUE (CON), RETAIN (RET), SMOOTH-SHIFT (SSH) and ROUGH-SHIFT (RSH). These transitions are determined by the relation between the CB of the utterance U_i and the CB of the previous utterance U_{i-1}, as well as by the relation between the CB and the CP of the utterance U_i. We take the definitions of the basic transition types as given by Walker et al. (1998). The transitions CON and RET both continue the CB (≈ topic). The difference between the two is that in case of RET there is a change of CB predicted in the subsequent utterance: $CB(U_i) \neq CP(U_i)$. In case of CON, the prediction is also keeping the CB. The transitions SSH and RSH both change the CB. The difference is that at SSH the prediction is that this new CB will be kept $[CB(U_i) = CP(U_i)]$, while at RSH, it is not the case. Considering the possibility that an utterance has no CB, we follow the transition definitions given by Walker et al. (1998), and consider transitions where the second utterance has no CB under CON and RET. We take the transitions where the first utterance has no CB as the NULL transition (Poesio et al., 2004), which is considered as a kind of shift at the global-level: it corresponds to the beginning of a new discourse segment.

The application of Centering Theory to Hungarian data is novel, no such analysis was proposed before. Centering Theory has undergone various different developments resulting in differences in setting the basic parameters within the various centering analyses (see also Poesio et al., 2004). Consequently, in a centering analysis of Hungarian data, we first need to decide how to set the

different parameters in this language. In our analysis and modeling, we follow as close as possible the standard practices within Centering Theory for segmenting the texts, for determining the centers, for ranking the CF-set and for determining the transition types. One of the most important parameters in centering is the ranking of the CF-set in each utterance. It is widely accepted that ranking is language specific, depending on the type and the specific properties of the given language. Considering that the main aim of this study is to determine the relation between the centers, the transitions and their morphosyntactic realization, we do not regard overt topic marking or zero coding as a determining feature for ranking. We argue that the best ranking for our study on Hungarian is the one based on logical conceptual structure as proposed by Cote (1998), and we refer to this parameter as the 'LCS position'. As usual, we rank arguments higher than adjuncts, and adjuncts higher than possessors.

In our study, we investigated 12 Hungarian spoken narratives acquired by guided elicitations based on the first three books of the picture book series, the 'Frog Stories', by Mayer (1967, 1969) and Mayer & Mayer (1971). The 12 stories are narrated by 7 different consultants. Each story is told by four different persons, all monolingual native speakers of Hungarian between age 22 and 53. Their style and ways of story-telling showed considerable differences, which provides us a diverse set of data. The 12 recordings contain 602 utterances, that forms our target data of investigation. Following the basic rules of Centering Theory, at each utterance, we determined the CB and the transition type. For the CB, we also registered: (1) the type of the expression (zero, overt NP, overt pronoun etc.), (2) its surface position (postverbal, 'topic position' etc.), (3) its 'LCS position' determined by the predicate and (4) when overt, its case.

2.2 Findings

We investigated 602 utterances, hence 602 transitions. In total, we identified 1295 centers, of which 27,3% were realized by a zero element, and 18,5% were expressed using structural topic. The rest of the centers were expressed overtly in a non-topic position. There were 448 CBs in total. We explored the relation of morphosyntactic coding and its role in the discourse from two perspectives: (1) what does a given morphosyntactic strategy express (CB, CP or none), and (2) which transition corresponds to a given morphosyntactic coding.

Regarding structural topic, we have found that it tends to express a center that is not the CB. This finding is crucial in our discussion. Both notions, the CB and the Hungarian structural topic, are claimed to be (roughly) equivalent to the *aboutness topic* of the utterance (see Walker et al., 1998; É. Kiss, 2004, respectively). Our findings suggest that structural topic marking and the CB of the utterance should not lead to the same (or similar) topic notion. The structural topic cannot simply be equated with the CB, i.e., the most salient or foregrounded referent. If we look at the distribution of topical CPs, we see that there is an even stronger tendency that structural topic rather corresponds to the CP of the utterance. The issue which transition is related to structural topic can be approached in two ways: (1) looking at the transition in all cases where the utterance contains a structural topic, and (2) looking at the transition where the CP is realized by structural topic. Both counts show that structural topic marking tends to correspond to some kind of shift (SSH, RSH, NULL). A relatively large number of NULL transitions associated with the 'topic position' further suggests that the structural topic marking in Hungarian corresponds to the marking of a new discourse segment, hence it is related to the global structure, i.e., the global coherence of the given text.

Regarding zero coding, we see a correspondence between zero coding and the CB. This is supported in both ways: when we look at how the CB is expressed and also when we look at all zero elements and what these realize. In both directions, we see a strong tendency that the CB is zero coded. By these findings, we argue that the semantic notion of (*aboutness*) *topic* or "*the discourse entity that the utterance U_i, most centrally concerns*" (Walker et al., 1998, p. 3) is rather associated with zero coding than with structural topic in Hungarian. As for the relation between a zero coded CB and the transition type, out of all cases where the CB is realized by a zero element, 74,9% corresponds to

some kind of continuation (CON/RET). A zero-marked CB is never associated with the NULL transition. This distribution indicates that there is a strong tendency that the use of zero coding establishes local coherence by continuation of the most salient center in consecutive utterances.

In addition to the above findings, we also explored the general (not Hungarian specific) issue of plural/set referents and the possible subtypes of RET and SSH. Considering set referents needs some elaboration regarding the transition types, that are defined using a 'difference-requirement'. A strict interpretation of this rule takes set referents as strictly different, without considering which individual referents are included in the given set. A more concessive view, however, makes a distinction between cases where the 'difference-requirement' is considered between a set referent and a singular referent, or between two set referents. This view leads to a more fine-grained distinction of transition types, with inclusion (or subset) relation. Take, for example, the SSH transition with the rule $CB(U_i) \neq CB(U_{i-1})$. When one or both CBs is a set, and the two are strictly different, it means that an entirely new topic is chosen. If the current CB includes the previous CB, then the topic is extended, hence the topic is kept, but more is added to it. And finally, if the previous CB includes the current CB, it indicates that the current topic is part of the previous topic, hence the topic is partially the same. In our study, we identified a number of such cases, and based on their distribution and their morphosyntactic coding, we argue that distinguishing these sub-types is meaningful. Considering them further strengthens our findings of the discourse roles of the target morphosyntactic marking strategies.

3 Conclusions

Based on our centering analysis and findings above, we argue that the characterization of the sentence topic as 'what the sentence is about' is not sufficient. By such a notion the functions of the two morphosyntactic strategies cannot be fully captured. A proper characterization of their functions must be given above the sentence-level. We conclude from our study that there is a strong tendency to mark larger discourse units (discourse segments) by structural topic, hence this morphosyntactic strategy should also be considered at the level of global coherence. On the level of local coherence, we conclude, that the two morphosyntactic strategies reflect the two-dimensional aspect of salience (e.g., Givón, 1983, 2001) in discourse: zero argument coding has the backward-looking function, establishing local coherence by topic continuity, while structural topic marking has the forward-looking function, indicating a topic shift, which relates both to the local and the global coherence. Our findings meet our initial hypothesis based on Comrie (1999) and Van Valin (2005), of which we provided experimental evidence.

Note that our findings and conclusions do not go against the claims of, e.g., É. Kiss (2004), on the analysis of structural topic in Hungarian. We offer a more elaborate characterization of the function of structural topic marking in discourse, which was merely given by É. Kiss as "*the topic foregrounds an individual*" (É. Kiss, 2004, p. 8), without specifying what the exact process is behind 'foregrounding'. In this study, we provided an extension by investigating such processes within discourse. We also offered an explanation on zero elements along the same lines, and a comparison of the two strategies. As for zero coding, we argue that it is crucial to look beyond the issue what licenses zero elements and rather investigate the question: when zero coding is available in a language, what is its function within discourse and information exchange.

References

Chiarcos, Christian, Berry Claus & Michael Grabski. 2011. Introduction: Salience in linguistics and beyond. In Christian Chiarcos, Berry Claus & Michael Grabski (eds.), *Salience: Multidisciplinary Perspectives on its Function in Discourse*, 1–28. Berlin: De Gruyter Mouton.

Comrie, Bernard. 1999. Reference-tracking: description and explanation. *Sprachtypologie und Universalienforschung STUF* 52(3-4). 335–346.

Cote, Sharon. 1998. Ranking Forward-looking Centers. In Marilyn A. Walker, Aravind K. Joshi & Ellen F. Prince (eds.), *Centering Theory in Discourse*, 55–69. Oxford: Clarendon Press.

É. Kiss, Katalin. 2004. *The syntax of Hungarian*. Cambridge: Cambridge University Press.

Givón, Talmy. 1983. Introduction. In Talmy Givón (ed.), *Topic continuity in discourse: a quantitative cross-language study*, 5–41. Amsterdam: John Benjamins.

Givón, Talmy. 2001. *Syntax*. Amsterdam: John Benjamins 2nd edn.

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In Mitchell Marcus (ed.), *Proceedings of the 21st annual meeting of the Association for Computational Linguistics*, 44–50. Cambridge, MA: Association of Computational Linguistics.

Grosz, Barbara J., Aravind K. Joshi & Scott Weinstein. 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics* 21(2). 203–225.

Grosz, Barbara J. & Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3). 175–204.

von Heusinger, Klaus. 1997. Salience and definiteness. *The Prague Bulletin of Mathematical Linguistics* 67. 5–23.

Jaeggli, Osvaldo A. & Kenneth J. Safir. 1989. The null subject parameter and parametric theory. In Osvaldo A. Jaeggli & Kenneth J. Safir (eds.), *The null Subject Parameter*, 1–44. Dordrecht: Kluwer Academic Publishers.

Kálmán, László (ed.). 2001. *Magyar leíró nyelvtan: Mondattan I* [hungarian descriptive grammar: syntax i]. Budapest: Tinta Könyvkiadó.

Koeneman, Olaf & Hedde Zeijlstra. 2021. Morphology and pro drop. In Rochelle Lieber (ed.), *The oxford Encyclopedia of Morphology*, Oxford: Oxford University Press.

Mayer, Mercer. 1967. *A boy, a dog and a frog*. New York: Dial Books for Young Readers.

Mayer, Mercer. 1969. *Frog, where are you?* New York: Dial Books for Young Readers.

Mayer, Mercer & Marianna Mayer. 1971. *A boy, a dog, a frog and a friend*. New York: Dial Books for Young Readers.

Neeleman, Ad & Kriszta Szendrői. 2007. Radical pro drop and the morphology of pronouns. *Linguistic Inquiry* 38. 671–714.

Poesio, Massimo, Rosemary Stevenson, Barbara Di Eugenio & Janet Hitzeman. 2004. Centering: A Parametric Theory and Its Instantiations. *Computational Linguistics* 30(3). 309–363.

Van Valin, Robert D., Jr. 2005. *Exploring the syntax-semantics interface*. Cambridg: Cambridge University Press.

Walker, Marilyn A., Aravind K. Joshi & Ellen F. Prince. 1998. Centering in Naturally Occurring Discourse: An Overview. In Marilyn A. Walker, Joshi Aravind & Ellen Prince (eds.), *Centering Theory in Discourse*, 1–28. Oxford: Clarendon Press.