Morphosyntactic coding and local coherence in Hungarian

Kata Balogh

Abstract This paper presents a study on the discourse functions of two morphosyntactic strategies in Hungarian: *zero coding* and *structural topic marking*. We investigate these phenomena in naturally occurring narratives, and propose an analysis within the framework of *Centering Theory*. In earlier work on Hungarian, these two strategies are mostly discussed at the sentence-level syntax and semantics, but a detailed investigation of their discourse-level behaviour is generally missing. In this paper, we extend the earlier analyses on these phenomena and provide a more elaborate characterization of them, with special attention to their function in discourse. Next to our primary goal, we also discuss the issue of plural referents and their proper treatment within Centering Theory, and propose an extension to the rules of determining the transition types.

K. Balogh, Heinrich-Heine-Universität Düsseldorf, `balogh@hhu.de`

1 Introduction

The central aim of this paper is the investigation of the discourse functions of two morphosyntactic strategies in Hungarian: *zero coding* and *structural topic marking*. Both strategies are predominantly discussed within the generative tradition, and mainly analyzed at sentence-level syntax, in terms of the 'pro-drop property' and the 'topic position' respectively; however, a detailed investigation of their discourse-level behaviour is generally missing. This paper aims to contribute to filling this gap.

Both our target strategies can be associated with *salience* (von Heusinger 1997; Chiarcos & Claus & Grabski 2011) and with *aboutness topic* (Reinhart 1981; Roberts 2011).[1] Structural topic marking corresponds to a kind of high-

---

[1]In this paper, we only discuss 'sentence topics', and not 'discourse topics' and 'contrastive topics'. To avoid confusion of different terminologies (see Tomioka 2020), we strive to distinguish syntactic and semantic uses consequently. The terms 'topic', 'topicality' and

lighting, 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. We are particularly interested in the characterization and modeling of the two in terms of *discourse processes*. A proper characterization of their functions must be given beyond the sentence-level. To achieve this goal, an analysis in terms of Centering Theory provides an appropriate ground. The work reported here is inspired by Shimojo's (2016) study on Japanese. We mainly adopt the same methodology, but the core parameters of our study are different and adjusted to the specific questions in Hungarian. Furthermore, we tackle the issue of plural reference, which is not present in Shimojo's (2016) analysis. In turn, using the same methodology offers a good ground for a future cross-linguistic comparison.

1.1 Structural topic marking

Hungarian has a rich morphology with an extended case marking system and verbal inflection. These morphosyntactic devices code most grammatical information. Grammatical functions such as 'subject' and 'object' are not marked by syntactic configurations. However, it is not the case that in Hungarian the word order is entirely free. The unconstrained order of the words only concerns the grammatical functions. Word order variations play an important role, too, though from the perspective of discourse-semantics. In Hungarian, the surface order is closely related to the information structure of the utterance, as certain syntactic positions are driven by the discourse-semantic functions: *topic* and *focus*. Languages having this structural behaviour are called *discourse configurational* (É. Kiss 1995; Surányi 2015), and can be further distinguished by virtue of which functions are overtly marked in the morphosyntax. Topic-prominent languages mark the discourse function 'topic' on the surface structure (e.g., Japanese, Korean), while focus-prominent languages have overt focus marking (e.g., Aghem, Basque). Hungarian manifests both topic- and focus-prominence.

In Hungarian, the relative order of the constituents behind the verb does not encode grammatical differences, as all word order variations are associated with the same semantic content. On the other hand, the topic/focus structure of the utterance determines the surface structure of the prever-

---

'topic referent' will be used in interpretational terms, while 'structural topic', 'topic position' and 'topicalization' refer to syntactic notions.

bal elements, motivating the 'topic-' and 'focus-positions' in the syntactic representation.[2]

(1) *Mari-nak tegnap   minden fiú a népmesé-t   mondt-a   el.*
     Mary-DAT yesterday every   boy the folk.tale-ACC told-3SG.D PRT
     'Every boy told [Mary]^TOP [the FOLK TALE]^FOC yesterday.'

As (1) illustrates, an *aboutness topic* is expressed structurally, placed in the clause-initial, 'topic position', preceding the narrow identificational focus and the quantifier positions. Structural topic marking in Hungarian is also shown by the fact that categorical and thetic judgements are structurally different. A verb-initial structure (2) expresses a thetic judgment, while in categorical judgments (3), a constituent is selected and placed in the topic position (e.g., É. Kiss 2004).[3] The referent of the constituent in this position is the one the sentence makes a statement about.

(2) *Kerget-i   egy fiú a   kutyá-t.*
     chase-3SG.D a   boy the dog-ACC
     'A boy is chasing the dog.'   thetic

(3) *A   kutyá-t kerget-i   egy fiú.*
     the dog-ACC chase-3SG.D a   boy
     'A boy is chasing the dog. (≈ The dog, a boy is chasing him.)'   categorical

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 must be present in the discourse (see Kálmán 2001; É. Kiss 2004). It is generally assumed that the constituent in the topic position expresses the *aboutness topic* of the sentence. É. Kiss (2004) also defines the 'function of topic' along these lines.

1.2 Zero coding

In Hungarian, the subject pronoun is generally omitted in unmarked (or neutral) sentences (4a). This phenomenon is referred to as the *pro-drop property* in generative (syntactic) approaches (e.g., Jaeggli & Safir 1989;

---

[2]The preverbal field also hosts universal quantifiers, *also*-phrases and negation.

[3]Note that some simplification is at place here. The structural differences of the thetic and categorical judgements are not always clear-cut (e.g., Gécseg & Kiefer 2009). However, this does not undermine the existence and importance of the Hungarian topic position.

Koeneman & Zeijlstra 2019). Given that this morphosyntactic phenomenon often considers the subject, pro-drop languages are also called null subject languages. This phenomenon is present in a great number of languages (e.g., Hungarian (4a), Italian (4b), Chinese, Spanish, etc.), while this strategy is generally not available in English (4c).[4]

(4)   Did Anna invite Bea?

    a.   *Igen, meg-hívt-a   (Beá-t).*
        yes   PRT-invited-3SG.D (Bea-ACC)
        'Yes, she invited her/Bea.'

    b.   *Sì, (lei) la ha invitata.*
        yes (she) her has invited
        'Yes, she invited her.'

    c.   Yes, she invited her. / *Yes, invited her.

The above examples illustrate that Hungarian and Italian allow for an unpronounced subject without loss of grammaticality, as opposed to English, where both the subject and the object pronouns must be overtly expressed. Despite the fact that pro-drop is used most often on the subject pronoun, it is not restricted to it. In certain languages (e.g., in Hungarian), it is also possible to leave out the object pronoun or other elements.

It is widely accepted that there are different types of pro-drop, however, a precise characterization of these different types, as well as the typological distribution of the phenomenon is still under discussion (Koeneman & Zeijlstra 2019). As for the different types of pro-drop and their distribution across languages, one of the leading questions in morphosyntactic approaches is what the licensing conditions are for omitting pronouns. Along this line, different types of pro-drop are distinguished, such as agreement-based (or consistent) pro-drop (Jaeggli & Safir 1989; Koeneman & Zeijlstra 2019) and discourse-based (or radical) pro-drop (Neeleman & Szendrői 2007). As Koeneman & Zeijlstra (2019) conclude, it is rather difficult to precisely characterize the various pro-drop types merely based on the morphological properties of the given languages. Next to pro-drop, a similar strategy of *topic-drop* is

---

[4]Note that zero coded arguments can appear in English in very specific constructions (e.g., conjunction reduction) or in informal speech. However, a possibility of omitting arguments is generally not present in English, hence it is not considered a pro-drop language.

proposed (e.g., Huang 1984). Evidence from Germanic languages show that 'pro-drop' and 'topic-drop' are distinct phenomena. German, for example, is not considered a pro-drop language, but the subject/object can be omitted from the sentence if it occupies a topic position in the pre-field of root clauses.

Hungarian is considered a 'pro-drop' language, the subject pronoun is generally left out in unmarked sentences (see (4a), (5a)). When the subject pronoun is overt, it indicates contrast (5b). Next to the subject pronoun, the object pronoun can also be omitted under given circumstances, however, there is an asymmetry. Contrary to an overt subject pronoun, an overt object pronoun itself does not signal contrast, the sentence in (5a) receives the same interpretation with or without overtly expressing the direct object.

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

Zero coding of arguments in Hungarian is often taken as agreement-based 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 (2019) and Neeleman & Szendrői (2007), we find phenomena specific for both agreement- and discourse-based pro-drop. For example, similar to Japanese and Chinese, various pronouns can be omitted, which is a characteristic property of discourse-based pro-drop. In the following, we will not target the issue of licensing conditions and the type of pro-drop in Hungarian. We are rather interested in the relation between discourse interpretation, discourse modeling and zero coding, as well as the processes that are behind this phenomenon. Despite the rich agreement and verbal inflection in Hungarian, which strongly points to the direction of agreement-based pro-drop, we inquire here into the aspects that possibly relate Hungarian to the Japanese-/Chinese-type pro-drop. We are not investigating what morphosyntactic features make zero coding possible, but rather pose the question: since this strategy is available in Hungarian, what is its function (or role) in the discourse.

2 Centering Theory

Centering Theory (Grosz, Joshi & Weinstein [GJW] 1983; 1995; Walker, Joshi & Prince [WJP] 1998; Brennan, Friedman & Pollard [BFP] 1987) is the local-level component of the theory of discourse structure proposed by Grosz and Sidner (Grosz 1977; Sidner 1979; Grosz & Sidner 1986). Centering Theory models the *attentional state* of the discourse, explaining local coherence between utterances. In their theory of discourse structure, Grosz & Sidner (1986) make a distinction between *global coherence* and *local coherence*, as the two major angles of discourse interpretation. Local coherence concerns the relation between individual utterances, while global coherence relates discourse segments, larger spans of texts.

Morphosyntactic choices within an utterance are related to the given discourse context. This relation is bidirectional: the discourse context restricts the morphosyntactic choices, while marking of information structure at the sentence-level helps discourse processing, as it reflects the underlying discourse structure. The structure of information in discourse is reflected in different linguistic structures contributing to the topic-comment and focus-background divisions. Such structures are manifested by special constructions (e.g., topicalization, clefts), by prosodic prominence, and by anaphora and ellipsis. Looking at the side of discourse processing, more coherent discourse is easier to interpret. In interpreting the discourse, the central processing tasks are: (1) determining the coherence relations and (2) finding the referents of the expressions in the sentence. Centering Theory concerns the *local discourse coherence* between the utterances. It models the focus of attention of the discourse participants, i.e., a local-level component of the Attentional State. Centering establishes the relation between salient discourse entities and the ways of their linguistic expression.

2.1 Centers and ranking

The salient discourse entities ($\approx$ referents) 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* (CFs) and a distinguished single *backward-looking center* (CB), which establishes the connection to the previous utterance. The CB of the utterance roughly corresponds to the notion of *(aboutness) topic* in other theories. After the initial proposal (GJW 1983, 1995; WJP 1998), Centering Theory has undergone several developments resulting

in differences in setting the basic parameters within the various analyses. One of these differences is whether only a single or also multiple CBs are allowed. In this paper, we follow the classical analysis with a single CB.

The set of forward-looking centers (CF-set) 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 CB in the subsequent utterance. The highest ranked element is the *preferred center* (CP), the center that is most likely to be the CB of the next utterance. Hence, the CP can be seen as the predicted next CB. Grosz, Joshi & Weinstein (1995) argue that various features – syntactic, semantic, lexical – play a role in determining the ranking. They propose that grammatical relations have the major role in ranking, which is sufficient for English (BFP 1987). Nevertheless, it is generally accepted that ranking is language specific and various ranking strategies are proposed. Kameyama (1985) included *-wa* marked topicalization and zero coding above grammatical functions for Japanese, Rambow (1993) added word order for German, Turan (1998) argues for thematic relations for Turkish and Cote (1998) argues for ranking by *lexical conceptual structures* (Jackendoff 1990).

Determining the ranked CF-set and the CB is driven by different constraints (see (6); WJP 1998: 3). For identifying the CB, the ranking of the CF-set in the previous utterance plays a crucial role (see 3. below).

(6) For each utterance U_i, in a discourse segment D, consisting of utterances $U_1, ..., U_m$:

  1. There is precisely one backward-looking center $Cb(U_i, D)$.
  2. Every element of $Cf(U_i, D)$ must be realized in U_i.
  3. The center, $Cb(U_i, D)$, is the highest-ranked element of $Cf(U_{i-1}, D)$ that is realized in U_i.

The first constraint in (6) states that each utterance has exactly one CB. This does not allow utterances without a CB or utterances with multiple CBs. Nevertheless, CB-less utterances are common. They are typical at the beginning of a discourse, but can occur at any place later. Therefore, a weaker version of the constraint is applied that states that there is no more than one CB in an utterance (WJP 1998: ftn. 2). This allows CB-less utterances

different from the initial one, while keeping the restriction that no multiple CBs are allowed. This is the view we adopt within our study.

The 'realize-relation' is based on the particular semantic theory assumed for interpretation. It is a generalization of the relation 'directly realize': "*U directly realizes c iff U is an utterance (of a phrase) for which c is the semantic interpretation.*" (GJW 1995: 9). The precise definition depends on the semantic theory used. In an earlier draft of the same work (GJW 1986), the realize-relation is based on situation semantics (Barwise & Perry 1983).

2.2 Transition types

Between the subsequent utterances, four transition types are defined: Continue (con), Retain (ret), Smooth-shift (ssh) and Rough-shift (rsh). These transitions are determined by (i) the relation between the CB of U_i and the CB of the previous utterance U_{i-1}, i.e., whether the CB is changed, and (ii) by the relation between the CB and the CP of U_i, i.e., whether keeping the CB is predicted. We take the definitions of the basic transition types as given by WJP (1998).

Transition type	relation between $CB(U_i)$ and $CB(U_{i-1})$		relation between $CB(U_i)$ and $CP(U_i)$
Continue	$CB(U_i) = CB(U_{i-1})$ or no $CB(U_{i-1})$	and	$CB(U_i) = CP(U_i)$
Retain	$CB(U_i) = CB(U_{i-1})$ or no $CB(U_{i-1})$	and	$CB(U_i) \neq CP(U_i)$
Smooth-shift	$CB(U_i) \neq CB(U_{i-1})$	and	$CB(U_i) = CP(U_i)$
Rough-shift	$CB(U_i) \neq CB(U_{i-1})$	and	$CB(U_i) \neq CP(U_i)$

Figure 1 Transition types

The first two transition types both continue the backward-looking center (CB ≈ topic). The difference between the two is that in case of Retain there is a change of CB predicted in the subsequent utterance, given that the current CP differs from the current CB. In case of Continue the prediction is also keeping the CB. The last two transitions both change the CB. The difference is that at Smooth-shift the prediction is that this new CB will be kept [$CB(U_i) = CP(U_i)$], while at Rough-shift, it is not the case [$CB(U_i) \neq CP(U_i)$]. Various proposals are made before to capture transitions between utterances where one (or both) of them has no CB. Kameyama (1986) adds CB-Establishment, a transition from a CB-less utterance to one with a CB.

The opposite of that, where a CB-less utterance follows one with a CB, can be considered as a Zero transition, and when a CB-less utterance follows another CB-less utterance it can be considered a Null transition (Poesio et al. 2004). In our analysis, we follow the definitions given by WJP (1998), and take transitions from a CB-less utterance to one with a CB under Continue and Retain. Hence we do not consider CB-Establishment. Furthermore, we take the transitions where the second utterance has no CB, whether or not the first utterance has one, as one case, and indicate them as Null.

3 Centering analysis

In this study, we investigated 12 Hungarian spoken narratives acquired by guided elicitations based on the picture books known as the 'Frog Stories' (Mayer 1967; 1969; Mayer & Mayer 1971). These books each tell a different story using exclusively illustrations, but no words. The 12 stories are narrated by 7 different consultants, each story is told by 4 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 form our target data for the investigation. The core data in this study contain naturally occurring narratives, which primarily represent language production, and as such they are significant for any analysis and explanation of morphosyntactic choices. The data are morphosyntactically annotated and segmented, both carried out by 2 different annotators.

3.1 Predictions and main questions

The starting point of our analysis is based on earlier approaches to Hungarian structural topic marking and zero coding. Both target morphosyntactic strategies are related to the relative *salience* of the given referents and to their *topicality* in terms of *aboutness topic*.

The first prediction concerns the morphosyntactic realization of topicality and the most salient entity. Both notions, the CB and the referent of the expression in the Hungarian topic position, are considered to be (roughly) equivalent to 'aboutness topic' (Reinhart 1981). This predicts that the filled topic position should express the CB of the utterance.[5]

---

[5]Note that Hungarian allows multiple topic positions, however, such constructions are less frequent and raise further issues that go beyond the scope of the paper. We restrict our

The BACKWARD-LOOKING CENTER, Cb(U$_i$, D) is a special member of the Cf, which represents the discourse entity that the utterance U$_i$, most centrally concerns, similar to what is elsewhere called the 'topic' (Walker, Joshi & Prince 1998: 3)

The topic foregrounds an individual (a person, an object, or a group of them) from among those present in the universe of discourse as the subject of the subsequent predication. (É. Kiss 2004: 8)

(…) the landing site of topic movement is assumed to be the specifier position of a functional projection called TopP. (É. Kiss 2004: 12)

The other prediction is based on the generalizations by Comrie (1999: 342): "*(…) in the extended domain, the expectation is for referential continuity, or, as it is often called in the literature, topic continuity*", and by Van Valin (2005), who similarly argues that a zero morpheme is the most unmarked topic and as such it marks continuing topics.

These predict that although the two strategies in Hungarian are both related to topicality, they differ in the discourse processes underlying them. This claim is supported by the different uses of zero coding and structural topic marking, as illustrated in (7) below. Our expectation is that this pattern is verifiable on a larger amount of data.

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

   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 / he$_{=frog}$ jumped away to a branch.'   (shift)

In order to check the above predictions on a larger scale, we investigate data from Hungarian narratives and provide a centering analysis. In our analysis, we investigate the use of structural topic marking and zero coding with respect to (i) the relative salience of the centers (i.e., referents) and (ii) the established local coherence between the utterances (i.e., the respective transition types). The analysis is driven by the following core questions:

study to a single topic position, and leave multiple structural topics for further work.

(1) What is the distribution of the transitions?

(2) Which center (CB, CP, none) is realized by { structural topic / zero }?

(3) What are the correspondences between the expression of the CB and the transition type? In particular: (a) When the CB is realized by { structural topic / zero }, what is the transition? and (b) For a given transition type how is the CB expressed?

3.2 Parameters for Hungarian

Centering Theory is parametric and therefore the analysis is dependent on the ways these parameters are set (Poesio et al. 2004). The application of Centering Theory to Hungarian data is novel, as no comparable analysis has been proposed before.[6] Therefore, we first need to discuss how to set the different parameters for Hungarian. In our analysis, we follow as close as possible to the standard practices within Centering Theory, and propose only necessary changes, for example, which language-specific features are determinant for ranking, or how to implement the 'realize-relation'.

3.2.1 Segmenting

One of the crucial questions for any centering analysis is how to segment the given text into utterances. In the initial proposal (e.g., GJW 1995), utterances are simply identified with sentences. Kameyama (1998) considers intrasentential local coherence, and she proposes a segmenting of complex sentences and takes tensed clauses as the basic discourse units (i.e., utterances) in English. Following this approach, we take each coordinated clause and tensed adverbial clause as a separate utterance. We do not segment inifinitival complements, complement clauses and relative clauses.

3.2.2 The realize-relation

In each utterance, we need to determine the centers, i.e., the referents that are evoked/talked about at that point in the discourse. Referential expressions evoke referents to entities as usual. In Hungarian, we must consider zeros (or zero pronouns) as well, which realize arguments in subject or object

---

[6]Very few works on Hungarian mention Centering Theory in connection with particular phenomena (temporal relations (Fretheim & Vaskó 1996), anaphora resolution (Lejtovicz & Kardkovács 2007)), but none of these works can be considered to be a strict application of Centering Theory to Hungarian data. Furthermore, no texts were analyzed, and the language-specific settings of the theory were not investigated beforehand.

functions[7] (8) and the possessor (9). Zero pronouns occur both in main and embedded clauses.

(8)   *Aztán lát-ják, hogy a   tav-on   van egy béka. Ott   üldögél.   És*
        then see-3PL that the lake-SUP is   a   frog there sit.around[3SG] and
        meg-próbál-ják el-kap-ni.
        PRT-try-3PL.D   PRT-catch-INF
        'Then they see that there is a frog at the lake. He is sitting around there. And they try to catch him.'

(9)   He thought to go to them following the footsteps of the boy and the dog.
        Meg is   érkezett a   lakás-uk-ba.
        PRT also arrived the flat-PS.3PL-ILL
        'And he arrived in their flat.'

There are special constructions that also realize entity referents, although the surface linguistic expressions are not referential. Such expressions are, for example, elliptical noun phrases, where the head-noun is missing in the adjective+noun combination, and the adjective carries the case marking (10b). Note that similar constructions appear in other languages as well.

(10)   a.   *a   vidám béká-t*   b.   *a   vidám-at*
              the happy frog-ACC   the happy-ACC
              'the happy frog'   'the happy one'

Adverbials with person/number marking, e.g., *mellé-m* (next-1SG) 'next to me' also occur regularly. Such adverbials express the same as the referential phrases like *én mellém* (I next-1SG), and thus we analyze them as realizing entity referents. We also list event referents but reduced to the minimum necessary. Referents to events/states are naturally evoked within each utterance. Nevertheless, we do not add all of these event referents to our analysis; rather we only consider them when they are (also) realized by an anaphoric device in a subsequent clause. In case the text contains such linguistic expressions, we consider the event referent they evoke, as well as their antecedents. In all other cases, we leave out event referents.

A large number of plural referents (e.g., *they, the dog and the boy*) occur in our data, and the way they are captured can affect our analysis. We take

---

[7]Although, in most cases, the zero marked argument is the subject.

such plural referents as sets of referents, following the treatment of plural reference in DRT (Kamp & Reyle 1993). We assign them to plural pronouns and to zero elements next to plural verb forms (e.g., 3PL).

(11) u_1:   *fogt-a   a   vödr-é-t*
              took-3SG.D the bucket-PS.3SG-ACC
              'he$_{=boy}$ took his bucket'   CF-list: $r_{boy} > r_{bucket}$

        u_2:   *és   meg-fenyegett-e   a   béká-t*
              and PRT-threatened-3SG.D the frog-ACC
              'and he$_{=boy}$ threatened the frog'   CF-list: $r_{boy} > r_{frog}$

        u_3:   *és   ez-zel   el-ment-ek*
              and this-INS PRT-went-3PL
              'and then they$_{=boy,dog}$ went away'   CF-list: $\{r_{boy}, r_{dog}\}$

In u_1 above, the verb is inflected for 3rd person singular, the zero pronoun of the subject and the zero pronoun of the possessor of the object both realize the center/referent r_{boy} (referring to the boy). In u_2, there are two referents realized: r_{boy} by the zero pronoun as before and r_{frog} by the noun phrase *a békát* 'the frog.ACC'. In u_3, the subject is zero and the verb is inflected for third person plural, which realizes the plural referent (a set of referents), including the referents of the boy (r_{boy}) and the dog (r_{dog}). In the first analysis, we take each plural referents as a distinct one. An alternative approach will be discussed in Section 4.

3.2.3 Ranking in Hungarian

One of the most important parameters in a centering analysis 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, and the features that determine the ranking vary across languages (see also Section 2.1). Hungarian is a flexible word order language, where grammatical functions are not marked in the syntactic structure, but primarily marked by a rich system of case morphemes (there are approximately 20 cases in Hungarian). Word order variations are rather related to information structure, i.e., to the discourse-semantic functions of *topic* and *focus*. We argue that grammatical function as the major determining feature for ranking is not the most appropriate one for Hungarian. Given the similarities in the use of zero coding and overt topicalization, we could

consider a similar ranking as proposed for Japanese. However, we do not follow this suggestion. The main aim of our study is to investigate the discourse functions of our target morphosyntactic strategies. Since ranking is a crucial aspect in determining the transition type, and eventually the discourse function, considering structural topic marking and zero coding as a determining feature for ranking would interfere with the purpose of the analysis. We argue that the best ranking for our study on Hungarian is the one based on lexical 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.

3.3 Methodology

The 12 texts were segmented into 602 utterances, following the instructions presented in Section 3.2.1. During the analysis, we registered the set of forward-looking centers (CF_1, ..., CF_n), ranked according to their LCS position determined by the given predicate. The first element of the ordered CF-set is also the preferred center ($CF_1 = CP$) of the utterance. Following the basic rules of Centering Theory (see Section 2), at each utterance, we determined the backward looking center (CB) and the transition type. For the CB, we also registered the following: (i) the type of expression (zero, overt NP, overt personal pronoun etc.), (ii) its surface position: whether it is in a postverbal position, in the clause-initial 'topic position', in the preverbal 'focus position' and so on, (iii) its LCS position determined by the predicate and (iv) its case. Surface position and case are only relevant in cases where the CB is overtly expressed. For an illustration, consider Figure 2 below. The transition type is indicated at the second utterance, e.g., the transition from utterance #3 to #4 is CON.

All 12 stories were examined in the way showed in Figure 2. In the analysis, we were looking at the correspondence of transitions and the ways the CB is expressed. In our evaluation, we focused on the use of structural topic marking and zero coding, and on the characterization of their behaviour in terms of local coherence in discourse.

#	predicate	properties of the CB				CB	Ranked CF-list			Trans
		type	pos	LCS	case		CP	CF2	CF3	Type
1	go						r_{boy}	r_{dog}	e_{fish}	NULL
2	believe	0		1		r_{boy}	r_{boy}	e_{fish}		CON
3	captured	ppro	top	1	nom	r_{boy}	r_{boy}	r_{frog}		CON
4	fall in	0		1		r_{boy}	r_{boy}	r_{wat}		CON
5	jump out						r_{frog}	r_{bra}		NULL
6	watch	0		1		r_{frog}	r_{frog}	r_{boy}	r_{dog}	CON

Figure 2 Centering analysis example

3.4 Findings

We investigated 602 utterances, hence 602 transitions. In total we identified 1295 centers, of which 353 (27,3%) are realized by a zero element, and 240 (18,5%) are expressed using structural topic. The rest of the centers are expressed overtly in a non-topic position. There are 448 backward-looking centers (CBs). First we have looked at the following: (i) the distribution of the transitions, (ii) What does the structural topic express: CB, CP or neither of the two? and (iii) What does the zero element express: CB, CP or neither of the two? Then, we looked at the correspondences between the expression of the CB and the transition from two directions: (i) In case the CB is expressed by { structural topic / zero }, what is the transition? and (ii) For a given transition type how is the CB expressed?

As for the distribution of the transitions, out of the 602 transitions there were 194 Continue, 105 Retain, 105 Smooth-shift, 44 Rough-shift and 154 Null. The first two transition types reflect a continuation of the CB, which occurred 299 times (49,7%, thus almost the half of it). The second two show some kind of shift of the CB, which happened 149 times (24,8%) and the latter one corresponds to the beginning of a new discourse segment, which was the case 154 times (25,6%). These counts meet the expectation of a coherent discourse, where the preferred local transition is continuation, but there are significant number of changes, either at the local level (ssh/rsh) or on a higher level (null).

3.4.1 The use of the 'topic position'

With respect to structural topic and the related transitions, we have two leading issues: (i) What does the structural topic realize, i.e., whether it is the CB, the CP or neither of the two? and (ii) In case the CB is encoded by a

structural topic, what is the transition?

As for the first question, we have found that the element in the 'topic position' (i.e., structural topic) tends to express a center that is not the CB. From all cases where a constituent occurred in the topic position (240 in total), 176 times (73,3%) it realizes a center that is not the CB, and merely 64 times (26,7%) the CB. This strongly indicates that structural topic marking generally does not correspond to the CB of the utterance. This finding is crucial in our discussion, as it contradicts our first prediction (Section 3.1). Both the CB and the Hungarian clause-initial 'topic position' are characterized in terms of 'aboutness' and a classical understanding of 'aboutness topic' (e.g., WJP (1998) and É. Kiss (2004); Section 3.1). Our findings, however, suggest that structural topic marking in the clause-initial position and the CB of the utterance do not lead to the same (or similar) topic notion. We propose that the function of the structural topic in Hungarian must further be investigated, and it should be given a discourse-based characterization reflecting aspects of the local or global coherence beyond the 'classical' topic function in terms of aboutness. É. Kiss' (2004: p. 8) definition touches upon discourse related functions but essentially this characterization does not go beyond the sentence-level and beyond determining the subject of the predication. The discourse related aspects in her description are basically used to explain the referentiality and specificity requirements of the referent expressed by the element in the structural topic position.

Here we concentrate on determining the discourse function of the structural topic. As we saw above, the referent of the structural topic cannot simply be equated with the CB, i.e., the most salient or foregrounded center. Considering local coherence and the transitions, we need to investigate what relation we can infer between structural topic marking and transition types. The second question at the beginning of this section need to be revised. We have shown that structural topic tends to express a non-CB, hence the question of what transition is related to a CB realized by structural topic is not the appropriate one to begin with. If we look at the distribution of CPs and non-CPs realized by a structural topic, we see that there is an even stronger tendency that structural topic is related to the CP (= predicted next CB) of the utterance. 204 times (85%) the element in the topic position realizes the CP, while only 36 (15%) times a non-CP.

There are two ways to target the second question, i.e., which transition

is related to structural topic marking. We can look at which transition we have in all cases where the utterance contains a filled topic position, and we can look at the transition where the CP is realized by a structural topic. These two counts are very similar in our data, given that 85% (204 out of 240) of the topic expressions realize the CP, and from the remaining 36, 13 are iterated topics, co-occurring with a CP expressed by a structural topic.
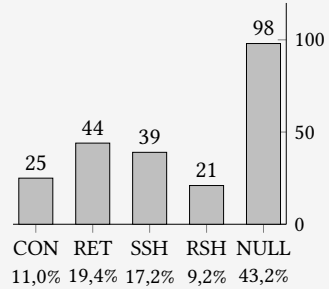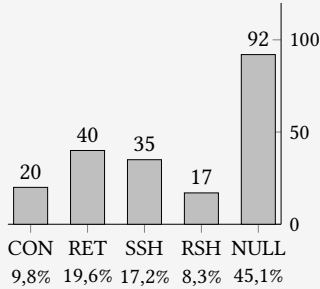

Figure 3 Distribution of transitions for sentences where the CP is realized with a structural topic

Figure 4 Distribution of transitions for sentences that contain a structural topic

Both counts show that structural topic marking in an utterance tends to correspond with some kind of shift ('topic shift') in the discourse. Looking at the second count in Figure 4, we see that in 43,2% the transition is NULL, while in 26,4% the transition is a SHIFT. These transitions both reflect some kind of change/shift in the discourse, SSH/RSH at the local-level, while NULL at the global-level. The latter transition is related to the beginning of a new discourse segment. Hence, in total, 69,6% of the structural topics are related to a shift-like transition, and only 30,4% are related to some kind of continuation. For the latter we must note that there are almost twice as many RET transitions than CON transitions. In the RET transition the predicted next CB ($\approx$ topic) is different, hence at least there is a predicted shift (whether or not the shift actually takes place in the next utterance).

This finding, i.e., that structural topic tends to correspond to a shift, fits to the generalizations by Comrie (1999) and Van Valin (2005). A relatively large number of NULL transitions associated with the topic position further suggests that the syntactic 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.

3.4.2 The use of zero coding

In our data, 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. As for the first aspect, we see a strong tendency to express the CB by a zero element. There are 448 CBs in our data, of which 287 (64,1%) are encoded by a zero element, while only 62 (13,8%) are encoded with a structural topic. From the rest, 94 (21,0%) are expressed by an overt element outside of the topic position, and 5 cases were unclear. The cases we classified as 'unclear' are the ones where the CB is a set referent, of which the elements, i.e., the individual referents, are expressed in different ways. See (12) for an illustration. In u_3, the CB is the set referent $\{r_{dog}, r_{turtle}\}$, while the referent of the turtle (r_{turtle}) is expressed using structural topic and the referent of the dog (r_{dog}) is expressed by a zero possessor.

(12) u_1:   *és el-kezdt-e   ugat-ni   a   teknőst*
                and PRT-began-3SG.D bark-INF the turtle-ACC
                'and he$_{=dog}$ began to bark at the turtle'

        u_2:   *hát komoly vitá-juk   volt*
                well serious argument-PS.3PL was
                'well, they$_{=dog+turtle}$ had a serious argument'

        u_3:   *hiszen a   teknős végülis be-rántott-a   a   gazdá-já-t*
                since the turtle after.all PRT-pulled-3SG.D the owner-PS.3SG-ACC
                a   víz-be
                the   water-ILL
                'since after all, the turtle pulled his owner into the water'

	CB	CP	ranked CF list	transition
u_1	r_{dog}	r_{dog}	$r_{dog} > r_{turtle}$	CON
u_2	r_{dog}	r_{arg}	$r_{arg} > \{r_{dog}, r_{turtle}\}$	RET
u_3	$\{r_{dog}, r_{turtle}\}$	r_{turtle}	$r_{turtle} > r_{boy} > r_{dog} > r_{water}$	RSH

Looking at the issue of the relation between zero coding and the CB from the other direction (from zero coding to CB/non-CB), we can conclude that out of all zero elements (353), 289 (81,9%) realize the CB and only 64 (18,1%) realize a center that is not the CB. These findings point to the direction

that the semantic notion of 'aboutness topic' or "*the discourse entity that the utterance U_i, most centrally concerns*" (WJP 1998: 3) is associated with zero coding and not with the structural topic position in Hungarian.

To determine the function of zero coding in discourse, we look at the relation between a zero CB and the transition type. Out of the 287 cases, where the CB is realized by a zero element, we have found 215 times (74,9%) some kind of continuation (CON/RET) and only 72 times (25,1%) a shift of the CB (SSH/RSH). A zero-marked CB is never associated with the NULL transition.
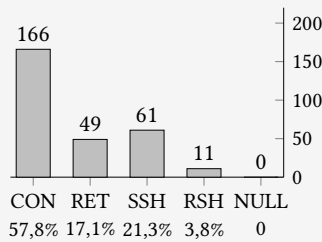
Figure 5 Distribution of transitions for sentences with a zero-marked CB

The exact distribution of the transitions is shown in Figure 5. 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.

4 Set reference and RETAIN/SMOOTH-SHIFT sub-types

Considering plural referents as 'different' centers needs some elaboration regarding the transition types that require a difference between the CB and the CP of the current utterance and/or between the CBs of the current and the previous utterance. Respectively, $CB(U_i) \neq CP(U_i)$ holds for RETAIN and ROUGH-SHIFT and $CB(U_i) \neq CB(U_{i-1})$ holds for SMOOTH-SHIFT and ROUGH-SHIFT. The issue is how to apply this 'difference-requirement'. Given that plural referents occur in large number in a text in any language, this issue is significant for any centering analysis in general. A strict interpretation of the 'difference-requirement' takes plural referents (i.e., sets of referents) as strictly different ones, without considering which individual referents are included in the given set. A more concessive view makes a distinction between cases where the 'difference-requirement' is considered between a

plural referent (a set) and a singular referent or between two plural referents, where the former includes the latter. For example, when the $CB(U_i) = r_1$ and $CP(U_i) = \{r_1, r_2\}$ or $CB(U_i) = \{r_1, r_2\}$ and $CP(U_i) = \{r_1, r_2, r_3\}$. This latter view leads to a more fine-grained distinction of transition types.

The second condition of RETAIN (see Figure 1) requires that the CB and the CP of the current utterance are different: $CB(U_i) \neq CP(U_i)$. Both the CB and the CP can either be a single referent or a set of referents. The difference between the two can be strict, $CB(U_i) \neq CP(U_i)$, meaning that they are either two different single referents, or one is a single referent that is not the element of the other plural referent, or they are two disjoint sets of referents. This strict difference means that the prediction is that in the next utterance, the most prominent center will be entirely new. The $CB(U_i)$ and the $CP(U_i)$ can also be overlapping, hence partly different (or partly the same). There are more possibilities:

(a) if one is a single referent and the other is a set, then either $CB(U_i) \in CP(U_i)$, or $CP(U_i) \in CB(U_i)$); and

(b) if they are both plural referents, then either $CB(U_i) \subset CP(U_i)$, or $CP(U_i) \subset CB(U_i)$.

These reduce to two relevant cases. When $CB(U_i)$ includes $CP(U_i)$, it predicts that the next topic is part of the current topic, i.e., it predicts that the speaker will partly go on with the same topic. Or, when the $CP(U_i)$ includes $CB(U_i)$, it indicates a prediction that the next topic includes the current topic, hence the prediction is that the speaker will go on with the same topic, but 'extends' it. Accordingly, we propose to make a distinction between different sub-types of RETAIN (RET) as: RET, RET$^-$ and RET$^+$.

RET	CB(U_i) = CB(U_{i-1}) or no CB(U_{i-1})	and	CB(U_i) ≠ CP(U_i)
RET$^-$			CP(U_i) ∈ CB(U_i) or CP(U_i) ⊂ CB(U_i)
RET$^+$			CB(U_i) ∈ CP(U_i) or CB(U_i) ⊂ CP(U_i)

Figure 6 Sub-types of RETAIN

The following two examples from our data illustrate the proposed transitions, RET^- (13) and RET^+ (14):

(13) u_1: *ahol   valami-t   hall-anak*
              where something-ACC hear-3PL
              'where$_{=lake}$ they$_{=boy+dog}$ hear something'

        u_2: *fel-figyel-nek ar-ra,   hogy valami-t   hall-anak*
              PRT-notice-3PL that-SUB that something-ACC hear-3PL
              'they$_{=boy+dog}$ notice that they$_{=boy+dog}$ hear something'

        u_3: *a kisfiú csend-re   int-i   társ-á-t*
              the boy   silence-SUB wave-3SG.D companion-PS.3SG-ACC
              'the boy silence his$_{=boy}$ companion'

	CB	CP	ranked CF list	transition
u_1	$\{r_{boy}, r_{dog}\}$	$\{r_{boy}, r_{dog}\}$	$\{r_{boy}, r_{dog}\} > r_{lake}$	SSH
u_2	$\{r_{boy}, r_{dog}\}$	$\{r_{boy}, r_{dog}\}$	$\{r_{boy}, r_{dog}\} > e_{hear}$	CON
u_3	$\{r_{boy}, r_{dog}\}$	r_{boy}	$r_{boy} > r_{dog}$	RET^-

In (13), the transition from u_2 to u_3 is RET^-, since the CP of u_3 is a single referent, r_{boy}, which is an element of the plural CB of u_3: = $\{r_{boy}, r_{dog}\}$.

(14) u_1: *Van egy kisfiú, aki halász-ni megy   a   kutyá-já-val.*
              is   a   boy   who fish-INF go[3SG] the dog-PS.3SG-INS
              'There is a boy who goes fishing with his$_{=boy}$ dog.'

        u_2: *Éppen meg-néz-i,   hogy milyen állapot-ban van a   tó.*
              just   PRT-look-3SG.D that what state-ILL is the lake
              'He$_{=boy}$ is just looking in what state the lake is.'

        u_3: *Aztán lát-ják,   hogy a   tav-on   van egy béka.*
              then see-3PL.D that the lake-SUP is   a   frog
              'Then they$_{=boy+dog}$ see that there is a frog on the lake.'

	CB	CP	ranked CF list	transition
$u1$		r_{boy}	$r_{boy} > e_{fish} > r_{dog}$	NULL
$u2$	r_{boy}	r_{boy}	$r_{boy} > r_{lake}$	CON
$u3$	r_{boy}	$\{r_{boy}, r_{dog}\}$	$\{r_{boy}, r_{dog}\} > r_{frog} > r_{lake}$	RET^+

In example (14), the transition from the second utterance to the third is

RET$^+$, since the CB of u_3 is a single referent, r_{boy}, which is an element of the plural CP = $\{r_{boy}, r_{dog}\}$.

Similar to the sub-types of RETAIN, we define sub-types of the SMOOTH-SHIFT (SSH) transition regarding its first condition that the CB of the current utterance and the CB of the previous utterance must be different: CB(U_i) ≠ CB(U_{i-1}). When these two are strictly different, it means that an entirely new topic is chosen (topic is shifted). If the current CB includes the previous CB, then the current 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 one, hence the topic is partially the same.

SSH	CB(U_i) ≠ CB(U_{i-1})	and	
SSH$^-$	CB(U_i) ∈ CB(U_{i-1}) or CB(U_i) ⊂ CB(U_{i-1})	and	CP(U_i) = CB(U_i)
SSH$^+$	CB(U_{i-1}) ∈ CB(U_i) or CB(U_{i-1}) ⊂ CB(U_i)	and	

Figure 7 Sub-types of SMOOTH-SHIFT

We find examples for both SSH$^-$ and SSH$^+$ transitions; see example (15).

(15) u_1:   *az ablak-on   ordibálva is   kerest-ék   a   béká-t*
             the window-SUP shouting also searched-3PL the frog-ACC
             'they$_{=boy+dog}$ were also looking for the frog shouting in the window'

         u_3:   *de nem talált-ák*
             but not found-3PL.D
             'but they$_{=boy+dog}$ could not find him$_{=frog}$'

         u_3:   *a   kiskutyus ki-esett   az ablakpárkány-ról*
             the doggy   PRT-fell[3SG] the window.sill-DEL
             'the doggy fell from the window sill'

	CB	CP	ranked CF list	transition
u_1	r_{dog}	$\{r_{boy}, r_{dog}\}$	$\{r_{boy}, r_{dog}\} > r_{frog} > r_{win}$	RET$^+$
u_2	$\{r_{boy}, r_{dog}\}$	$\{r_{boy}, r_{dog}\}$	$\{r_{boy}, r_{dog}\} > r_{frog}$	SSH$^+$
u_3	r_{dog}	r_{dog}	$r_{dog} > r_{win.sill}$	SSH$^-$

In the second utterance, the CB is "extended" from the single referent r_{dog} to the set referent $\{r_{boy}, r_{dog}\}$, hence the transition from u_1 to u_2 is SSH$^+$. In the third utterance, the CB is "reduced" from the set containing two

referents, the boy and the dog, to the single referent of the dog. Hence, the transition from u_2 to u_3 is SSH$^-$.

In the following, we examine the effects of considering the above sub-types of RETAIN and SMOOTH-SHIFT in our analysis. Recall the sub-types we proposed before. The RET$^+$ transition predicts that the speaker will extend the current topic, and the RET$^-$ transition predicts that the speaker will reduce the topic in the subsequent utterance. Similarly, with the SSH$^+$ transition the topic is kept but extended, and with the SSH$^-$ transition the topic is reduced or partially maintained.

In our data, we identified 99 RETAIN transitions, of which only 2 are of sub-type RET$^-$, while there are 26 of sub-type RET$^+$ and 71 of sub-type RET. We observed that there is a clear difference in morphosyntactic coding between the RET$^+$ transition and the other two (RET/RET$^-$). The relevant aspect here is the morphosyntactic coding of the CP. In case we have a RET$^+$ transition, the CP is in 81% coded by zero, while in case of the RET transition, the CP is in 91,5% overt (50,7% realized by structural topic, 40,8% realized outside of the topic position). This difference in morphosyntactic coding of the sub-types supports the distinction between RET$^+$ and RET/RET$^-$. As for the distribution of the "special" SMOOTH-SHIFT transitions, SSH$^+$ and SSH$^-$, we see that both are significantly represented in our data, and their relative distribution is similar. Out of the 105 occurrences of a SMOOTH-SHIFT, 17 are SSH$^+$ (16,2%) and 21 are SSH$^-$ (20%). This is different from the distribution of RET$^+$ versus RET$^-$ above, where RET$^-$ is marginal. Based on the related morphosyntactic coding, we can argue again that distinguishing the sub-types of the SMOOTH-SHIFT transition is meaningful. Similar to the sub-types of RETAIN, there is a clear difference in corresponding morphosyntactic coding. For the SMOOTH-SHIFT transition type from U_{i-1} to U_i, the expression of the CB in U_i is the relevant factor. We see that in case of the SSH$^+$ transition the CB is almost exclusively zero marked: in the 17 occurrences of SSH$^+$, the CB was zero marked in 16 cases, while it was encoded overtly outside of the topic position only once. With SSH$^+$, the CB was never encoded with structural topic. For the SSH$^-$ transition, on the other hand, there is a tendency to mark the CB with structural topic. The picture is less clear in this case, as we see zero coded CBs as well. In the 21 occurrences of SSH$^-$, the CB was encoded by a structural topic in 13 cases, by a zero in 7 cases, and by an overt element in a non-topic position once. Nevertheless, the tendency for structural topic

is clear. Given that in case of the "plain" SSH transition we also encounter various encodings of the CB, we suggest that a distinction between SSH$^+$ and SSH/SSH$^-$ can be of relevance.

We conclude from the above findings that the proposed sub-types, RET$^+$ and SSH$^+$, are indeed special. These are the transitions, where the predicted or actual next topic is continued, with something merely added to the set. Without considering the corresponding morphosyntactic coding, we argue that in these cases, the topic or the predicted topic is kept. In SSH$^+$, the CB(U_{i-1}) is fully preserved in CB(U_i), and in RET$^+$ the predicted topic CP(U_i) is fully preserved with respect to CB(U_i). Based on this argumentation, we propose to take RET$^+$ and SSH$^+$ as a continuation and count these cases under the CON transition. See below, in Figure 8, our proposed revision of determining the transition types accordingly.

Trans type	relation between CB(U_i) and CB(U_{i-1})		relation between CB(U_i) and CP(U_i)
CON	CB(U_i) = CB(U_{i-1}) or no CB(U_{i-1})	and	CB(U_i) = CP(U_i) or CB(U_i) ∈ CP(U_i) or CB(U_i) ⊆ CP(U_i)
	CB(U_{i-1}) ∈ CB(U_i) or CB(U_{i-1}) ⊂ CB(U_i)	and	CB(U_i) = CP(U_i)
RET	CB(U_i) = CB(U_{i-1}) or no CB(U_{i-1})	and	CB(U_i) ≠ CP(U_i) or CP(U_i) ∈ CB(U_i) or CP(U_i) ⊆ CB(U_i)
SSH	CB(U_i) ≠ CB(U_{i-1}) or CP(U_i) ∈ CB(U_{i-1}) or CP(U_i) ⊂ CB(U_{i-1})	and	CB(U_i) = CP(U_i)
RSH	CB(U_i) ≠ CB(U_{i-1})	and	CB(U_i) ≠ CP(U_i)

Figure 8 Revised transition types

This proposal is also supported by the corresponding morphosyntactic encodings as discussed before. Following our proposal, the distribution of transitions for sentences with a zero-marked CB is changed as shown in Figure 9, which shows the counts if we consider all RET$^+$ and SSH$^+$ transitions as CON. This revised distribution further strengthens our claims on the relation of zero marking and the transition types (see Section 3.4).
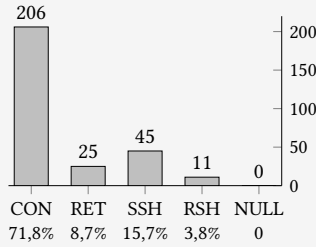
Figure 9 Revised distribution of transitions for sentences with a zero-marked CB (if RET+ = CON and SSH+ = CON)

Despite the convincing evidence from morpho-syntax, the above analysis is merely a first proposal. A more detailed investigation of the transitions and the corresponding morphosyntactic encoding is needed, and the validity of the proposed sub-types and their special treatment must be justified in a larger cross-linguistic setting. Furthermore, along the same line of argumentation, we can theoretically distinguish 9 sub-types of the Rough-shift transition. However, because of space limitations, we skip the discussion of these options in this paper and leave it for further investigation.

5 Conclusions

Both of our target morphosyntactic strategies are related to topicality. Structural topic marking is used to indicate what the sentence is about (É. Kiss 2004) and a zero argument is considered the most unmarked topic (Van Valin 2005). In Centering Theory, the backward-looking center (CB) is taken as an equivalent of 'aboutness topic' in other theories, while the preferred center (CP) is taken as the predicted next topic. Therefore, the goals of our study call for examining the expression of the CB and CP, as well as the corresponding transitions in a larger amount of text. We investigated 12 naturally occurring Hungarian narratives (see Section 3), containing 602 utterances. After analyzing our data within the framework of Centering Theory, we looked at the distribution and relation of given morphosyntactic codings and salient referents, as well as at the correspondences between the expression of the CB and the transition from different directions.

We found a clear correspondence between zero coding of the CB and the Continue transition, as well as an overtly topicalized CP and Shift transitions (Sections 3.4.1 and 3.4.2). Regarding structural topics, we have also found that they are very often associated with a null transition, which

indicates the beginning of a new discourse segment. We conclude from this that there is a strong tendency to mark larger discourse units (discourse segments) by structural topics; hence this morphosyntactic strategy should also be considered at the level of global coherence. Our findings and final conclusions do not go entirely against the claims of, for example, É. Kiss (2004), on the analysis of structural topic in Hungarian. Rather, we offer a more elaborate characterization of the function of structural topic marking in discourse, which was merely given as "*the topic foregrounds an individual*" (É. Kiss 2004: 8), without specifying what 'foregrounding' means, i.e., what the exact process is behind it. In this paper, we provided an extension by investigating such processes within discourse. We also offered and explanation for zero elements along the same lines, and a comparison of the two different strategies. Regarding zero coding in Hungarian, we argue that it is crucial to look beyond the morphosyntactic licensing conditions, and we investigated the question: what is the function of the available zero coding within discourse and information exchange?

Our findings suggest that the view that 'topicality is marked structurally in Hungarian' is only partly accurate, and it should rather state that if there is a structural topic in Hungarian, it marks some kind of topicality, namely a *shifted topic*. This is seen as singling out a constituent for being the topic of the subsequent utterances. Topicality can also be marked by zero elements, which signals a *continuing topic*. Hence, we conclude that on the level of discourse coherence the two morphosyntactic strategies reflect the two-dimensional aspect of salience (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 can operate both on the local or the global level of discourse coherence.

The findings in this study contradict our first prediction/hypothesis, which lead to a more elaborate characterization of topicality in Hungarian. On the other hand, our findings meet the second hypothesis based on Comrie (1999) and Van Valin (2005), of which we provided an experimental evidence. We claim that such experimantal studies, based on naturally occurring texts, are of great importance for any theoretical work. The formal analysis of any linguistic phenomenon requires that theoretical claims are verified by empirically valid studies.

Beyond our primary goal as explained before, we also discussed the issue of set referents and their proper treatment within Centering Theory. We proposed an extension to the rules of determining transition types by adding cases where there is an inclusion relation between two centers of which we need to determine whether they are different or the same. We introduced our first proposal on this issue in Section 4, that need further investigation, preferably in a cross-linguistic setting.

Despite the outcome and results of the study presented here, there are several further issues raised. First of all, our proposal needs to be evaluated by a comparative experiment in different languages that manifest some overt morphosyntactic topicalization strategy, as well as zero (argument) coding. Such a language is, for example, Japanese, in which a similar study is reported by Shimojo (2016). The results point in the same direction, however, the parameters in the two analyses are slightly different, which needs further investigation. Regarding the outcome for Hungarian, there are important questions we need to explore further. We need to look at cases that deviate from the tendencies we identified and find out the possible reasons behind it. With respect to this, we need to investigate the cases where a zero CB is associated with some kind of shift transition, and the cases where a structural topic cooccurs with a kind of continuation. The question is whether these cases are due to some 'mistake' or 'noise' (that is expected in a naturally occurring narrative), or whether there is another communicative or discourse function behind them, which we have not identified yet. These issues are left for further research and will be discussed in subsequent papers.

Abbreviations and glosses

1, 2, 3	1st, 2nd, 3rd person	FOC	focus	PS	possessive
ACC	accusative	ILL	illative	SG	singular
CT	contrastive topic	INF	infinitive	SUB	sublative
D	definite conjugation	INS	instrumental	SUP	superessive
DAT	dative	PL	plural	TOP	topic
DEL	delative	PRT	particle		

References

Barwise, Jon & Perry, John. 1983. *Situations and attitudes*. Cambridge, MA: MIT Press.

Brennan, Susan E. & Friedman, Marilyn W. & Pollard, Carl J. 1987. A centering approach to pronouns. In Sidner, Candy (ed.), *Proceedings of 25th annual meeting of the Association for Computational Linguistics*, 155–162. Stanford, CA: Association for Computational Linguistics.

Chiarcos, Christian & Claus, Berry & Grabski, Michael. 2011. Introduction: Salience in linguistics and beyond. In Chiarcos, Christian & Claus, Berry & Grabski, Michael (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 Walker, Marilyn A. & Joshi, Aravind K. & Prince, Ellen F. (eds.), *Centering Theory in Discourse*, 55–69. Oxford: Clarendon Press.

É. Kiss, Katalin (ed.). 1995. *Discourse Configurational Languages*. Oxford University Press.

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

Fretheim, Thorstein & Vaskó, Ildikó. 1996. Lexical properties and pragmatic implications of some markers of temporal succession and simultaneity: A contrastive analysis of Hungarian, Norwegian and English. *Language Sciences* 18(3-4). 791–810.

Gécseg, Zsuzsanna & Kiefer, Ferenc. 2009. A new look at information structure in Hungarian. *Natural Language & Linguistic Theory* 27(3). 583–622.

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

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

Grosz, Barbara J. 1977. *The Representation and Use of Focus in Dialogue Understanding*. Stanford, CA: Stanford University dissertation.

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

Grosz, Barbara J. & Joshi, Aravind K. & Weinstein, Scott. 1986. Centering: A Framework for Modeling the Local Coherence of Discourse. Manuscript.

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

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

Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15(4). 531–574.

Jackendoff, Ray. 1990. *Semantic Structure.* Cambridge, MA: MIT Press.

Jaeggli, Osvaldo A. & Safir, Kenneth J. 1989. The null subject parameter and parametric theory. In Jaeggli, Osvaldo A. & Safir, Kenneth J. (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.

Kameyama, Megumi. 1985. *Zero Anaphora: The Case of Japanese.* Stanford, CA: Stanford University dissertation.

Kameyama, Megumi. 1986. A property-sharing constraint in centering. In Biermann, Alan W. (ed.), *Proceedings of the 24st Annual Meeting of the Association for Computational Linguistics*, 200–206. New York, NY: Columbia University/Association of Computational Linguistics.

Kameyama, Megumi. 1998. Intrasentential Centering: A Case Study. In Walker, Marilyn A. & Joshi, Aravind K. & Prince, Ellen F. (eds.), *Centering Theory in Discourse*, 89–112. Oxford: Clarendon Press.

Kamp, Hans & Reyle, Uwe. 1993. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory.* Dordrecht: Kluwer Academic Publishers.

Koeneman, Olaf & Zeijlstra, Hedde. 2019. Morphology and pro drop. In Lieber, Rochelle (ed.), *Oxford Encyclopedia of Morphology*. Oxford University Press.

Lejtovicz, Katalin E. & Kardkovács, Zsolt T. 2007. Anaphora resolution. In Rudas, Imre J. (ed.), *Proceedings of the 8th International Symposium of Hungarian Researchers (CINTI 2007)*, 591–599. Budapest Tech.

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

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

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

Rambow, Owen. 1993. Pragmatic Aspects of Scrambling and Topicalization in German: A Centering Approach. Paper presented at the Workshop on Centering, Institute for Research and Cognitive Science (IRCS), University of Pennsylvania, May 1993.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1). 53–94.

Roberts, Craige. 2011. Topics. In von Heusinger, Klaus & Maienborn, Claudia & Portner, Paul (eds.), *Semantics* (Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK), Volume 33.2), 1908–1934. Berlin: de Gruyter Mouton.

Shimojo, Mitsuaki. 2016. Saliency in discourse and sentence form: zero anaphora and topicalization in Japanese. In Fernandez-Vest, M. M. Jocelyne & Van Valin Jr., Robert D. (eds.), *Information structure and spoken language in a cross-linguistic perspective*, 55–76. Tokio: de Gruyter Mouton.

Sidner, Candace L. 1979. *Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse*. Cambridge, MA: MIT dissertation.

Surányi, Balázs. 2015. Discourse-configurationality. In Féry, Caroline & Ishihara, Shinichiro (eds.), *The Oxford Handbook of Information Structure*, 422–440. Oxford University Press.

Tomioka, Satoshi. 2020. Topic. In Gutzmann, Daniel & Matthewson, Lisa & Meier, Cécile & Rullmann, Hotze & Zimmermann, Thomas Ede (eds.), *The Wiley Blackwell Companion to Semantics*. Wiley Blackwell.

Turan, Ümit D. 1998. Ranking Forward-Looking Centers in Turkish: Universal and Language-Specific Properties. In Walker, Marilyn A. & Joshi, Aravind K. & Prince, Ellen F. (eds.), *Centering Theory in Discourse*, 139–160. Oxford: Clarendon Press.

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

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

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