CENTERING-THEORY-BASED-ANNOTATION OF CZECH

Alena Poncarová

Charles University in Prague, Faculty of Arts, nam. Jana Palacha 2, 116 38 Prague, Czech Republic

Abstract

My paper aims to present the principles of partially automatic annotation based on Centering theory I used within my project of syntactic and text linguistics analysis of Czech. I intend to present basics of Centering theory (background, tools, methodology, and principles) and focus on rules based on Centering theory and consequences for the analysis. Moreover, I would like to present some problematic features of Centering theory from the point of view of applying Centering theory to praxis.

Key words: *Annotation, Centering theory, Czech, Syntactic and text linguistics analysis*

1. MOTIVATION

Present paper represents the part of larger research of Czech, which includes syntactic and text linguistics analysis, and attends to describe basic principles of text development with respect to information structure, constituent structure and coreferential chaining. Dominant (methodological) background of the research is based on Centering theory approach (Brennan 1995), one intermediate step of this research (which this paper is dedicated to) is to apply the Centering theory approach to authentic Czech texts. This applying Centering into practice shows, which aspects of this theory are plausible with Czech and which are not. Centering theory is originally based exclusively on English language data, so the application of this theory to Czech material can offer interesting comparison of Czech representing Slavic languages and English representing Germanic languages. From another perspective it could offer valuable feedback for the theory itself – it will be verified if Centering theory as it is designed is able to capture all aspects of structurally different language such as Czech, in other words the aspiration to be an universal theory will be verified.

2. CENTERING THEORY

Centering theory approach was developed in 70's, but the first paper was published in early 80's and then modified (Grosz, Joshi & Weinstein 1995). Centering is a complexity, noun-based theory (Ariel 2013), its' aim is to relate "focus of attention, choice of referring expression, and perceived coherence of utterances within a discourse segment"[1]. Centering aims to model "local attentional state"[2] – that is why Centering focuses only on two immediately adjacent utterances – the subsequent one and the given one (preceding and particular one, etc.). Strict locality principle is one of the most restrictive aspects of the whole theory. It clearly defines the basic level at which the analysis can be controlled (number of possible uncontrolled factors is reduced), on another hand it reduces the possibilities of the analysis (the most important restriction is that the relations between parts of text which would be further apart stay outside the range of Centering theory). As follows the previous, in the focus of Centering theory is the utterance, not a sentence, because "the same sentence uttered in different

[1] Grosz, B, Joshi, A & Weinstein, S 1995, 'Centering – A Framework for Modeling the Local Coherence of Discourse', *Computational linguistics*, vol. 21, no. 2, pp. 204.
[2] Grosz, B, Joshi, A & Weinstein, S 1995, 'Centering – A Framework for Modeling the Local Coherence of Discourse', *Computational linguistics*, vol. 21, no. 2, pp. 207.

discourse situations may have different centers, because determination of centers is related to reference resolution"[3].

To model text coherence and relations between particular parts of the text Centering theory operates with a few tools. First of them is the concept of Centers of Attention, the second basic pillar of Centering theory are four types of Transition. To be able to use basic concept of Centers of Attention it is necessary to define the relations on which the realization of particular types of Centers of Attention is evaluated and to declare rules and claims, which form the instructions how to treat them. The concept of Transition is directly dependent on the Centers of Attention classification, because the particular type of Transition, which applies in the particular part of the text, is defined on the basis of Centers of Attention classification within the immediately preceding utterance compared with the Centers of Attention classification within the given utterance.

2.1 Centers of Attention

Concept of Centers

As a Center of Attention is classified every entity present in the discourse regardless on the formal expression it is captured by: "centers are semantic objects, not words, phrases, or syntactic forms"[4]. When determining the Centers of Attention in a particular utterance, it is necessary to take into account the combination of discourse factors: "center determination is not solely a syntactic, semantic, or pragmatic process"[5]. According to the Centering theory literature (Grosz, Joshi & Weinstein 1995; Walker, Joshi & Prince 1998, etc.) there are three types of Centers of Attention.

First of them is called FORWARD-LOOKING CENTER (Cf) and it is the center which "represents discourse entities evoked by the utterance U_i in a discourse segment D"[6] and which has to be explicitly mentioned in the utterance. Its main role in the discourse is to open the potential coreferential chain, to be the antecedent of such coreferential chain. The set of forward-looking centers is ranked reflecting thematic role (Agent vs. Patient); grammatical role; and surface position.

"Perhaps the most important single element of Centering theory is the idea that entities realized in an utterance can be ranked according to their likelihood to be the center of the following utterance. The intuition is that certain features of an utterance serve as guides to the attentional structure of the discourse segment of which the utterance is a part. In particular, a speaker will design her utterance (U_i) not only to include a point of continuity with the previous utterance (U_{i-1}) but also to indicate how likely it is that each entity represented in U_i will be the main source of continuity […] in the next utterance (U_{i+1})."[7, 8]

Opposite kind of Centers of Attention is the BACKWARD-LOOKING CENTER (Cb), which is the entity which connects the particular utterance with the previous one. Connection between two utterances is based on coreferential relation between two (direct) realizations of the same entity in immediately adjacent utterances. This type of Center of Attention creates from separate utterances the complex discourse. To be able to determine which entity is in the center of attention of particular discourse segment Centering theory states the unique backward-looking center rule: "Each U_i has exactly one

[3] Kruijff-Korbayová, I & Hajičová, E 1997, 'Topics and Centers: A comparison of the salience-based approach and the Centering theory', *The Prague Bulletin of Mathematical Linguistics*, vol. 67, no. 1, pp. 37.
[4] Grosz, B, Joshi, A & Weinstein, S 1995, 'Centering - A Framework for Modeling the Local Coherence of Discourse', *Computational linguistics*, vol. 21, no. 2, pp. 208.
[5] Grosz, B, Joshi, A & Weinstein, S 1995, 'Centering - A Framework for Modeling the Local Coherence of Discourse', *Computational linguistics*, vol. 21, no. 2, pp. 210.
[6] Walker, M, Joshi, A. & Prince, E 1998, 'Centering in Naturally-Occurring Discourse: An Overview', in M. Walker, A. Joshi, E. Prince (eds.), *Centering theory in Discourse*, Clarendon Press, Oxford, pp. 2.
[7] Cote, S 1998, 'Ranking Forward-Looking Centers', in M. Walker, A. Joshi, E. Prince (eds.), *Centering theory in Discourse*, Clarendon Press, Oxford, pp. 55.
[8] Notation was modified according to the Walker, Joshi & Prince (1998) style. Original uses the U, U – 1, U + 1 notation.

backward-looking center."[9, 10] It is very important to keep in mind that the backward-looking center is strictly local, i.e., it interacts with the set of forward-looking centers of the previous utterance. From this principle it follows that it is absolutely necessary to reexamin the status of backward-looking center with each change of the utterance: next utterance in the line could have completely different backward-looking center. This principle holds in case of each type of Centers of Attention.

Third type of Centers of Attention is called the PREFERRED CENTER (Cp) and it is "preferred", because it is the highest ranked element of the list of forward-looking centers. This relation between preferred center and the set of forward-looking centers is important for the backward-looking center disambiguation: "The ranking of the forward-looking centers affects the choice of the backward-looking center of the subsequent utterance."[11] Centering theory (Walker, Joshi & Prince 1998) claims that the preferred center of the particular utterance becomes a backward-looking center of the following utterance. In most cases the preferred center is the subject of particular sentence (from the perspective of grammatical roles), of course, there are other factors which have to be taken into account – word order, clausal subordination, lexical semantics, etc. (Kruijff-Korbayová & Hajičová 1997).

Concept of Realization relations

Entities within the discourse can be present both directly and indirectly. In case of direct representation, the expression creates a node of coreferential chain: typical example is the pronominalisation in such cases like *Tom likes Alice. He invited her for a coffee*. Expressions *Tom* and *he* are members of the same coreferential chain, pronoun *he* preserves *Tom* present even in the second utterance. The coreferential relation in case of *Alice* and *her* exhibits the same characteristics. This situation is called DIRECT REALIZATION.

Second type how the entity can be present in a discourse is so-called REALIZATION, which is based on the principle of indirect representation, i.e. bridging (associative anaphor). Such relations are for example between *library* and *books* in cases like *John went to the library. He had read all own books. Books* are not completely new in the discourse, because they are closely related to the concept of *library*.

2.2 Types of Transition

Second basic pillar of Centering theory is the classification of Transition. As I mentioned above there is a close connection between the Centers of Attention classification and the Transition typology, because the Transition directly depends on the Centers classification. Centering theory distinguishes four types of Transition.

	$Cb(U_i) = Cb(U_{i-1})$ $Cb(U_{i-1}) = ?$	$Cb(U_i) \neq Cb(U_{i-1})$
$Cb(U_i) = Cp(U_i)$	Continue	Smooth-shift
$Cb(U_i) \neq Cp(U_i)$	Retain	Rough-shift

Table 1: Types of Transition

Source: Walker, M, Joshi, A. & Prince, E 1998, *Centering theory in Discourse*, Clarendon Press, Oxford, pp. 5.

[9] Grosz, B, Joshi, A & Weinstein, S 1995, 'Centering – A Framework for Modeling the Local Coherence of Discourse', *Computational linguistics*, vol. 21, no. 2, pp. 210.
[10] Notation was modified according to the Walker, Joshi & Prince (1998) style. Original uses the U_n notation.
[11] Kruijff-Korbayová, I & Hajičová, E 1997, 'Topics and Centers: A comparison of the salience-based approach and the Centering theory', *The Prague Bulletin of Mathematical Linguistics*, vol. 67, no. 1, pp. 25.

The relations between backward-looking and preferred centers in the previous and particular utterances are crucial. Types of Transition that preserve the backward-looking center from the previous utterance are called Continue and Retain. Two types that do not are Shifts (Smooth- and Rough-shift). Cross criterion is based on the relations between backward-looking center and preferred center of particular utterance: if they are represented by identical entity, the Transition is classified as Continue; if they are not, Transition is classified as Retain.

In other words: If the backward-looking center of the particular utterance is represented by the same entity as preferred center and is preserved from the preceding utterance, it is the Continue type of Transition. If the backward-looking center of the particular utterance is represented by different entity than its preferred center but is still preserved from the preceding utterance, it is the Retain type of Transition. If the backward-looking center of the particular utterance is not the same entity as the backward-looking center of the preceding utterance, Centering theory distinguishes two types of Shifts depending on whether the backward-looking center has a form of subject (Smooth-shift), or it does not (Rough-shift).

2.3 Rules

Centering declares two center management rules governing Centers of Attention classification and the Transition classification.

Pronoun rule

Kruijff-Korbayová and Hajičová (1997) summarized the rule Centering considers very important for the Centers of Attention classification, especially for the backward-looking center determination, as follows: "If any element of $Cf(U_i)$ is realized by a pronoun in U_{i+1}, then the $Cb(U_{i+1})$ must be realized by a pronoun also."[12, 13] This rule specifies the position of pronominalised entities: it puts them to the top of the list of backward-looking center candidates. Application of this rule is not problematic in the situation there is only one pronoun in the utterance, but becomes problematic when there is more than single pronoun within the utterance (for more information consult Poncarová 2014).

Center movement rule

Second rule formulated by Centering treats the Center movement, i.e., type of Transition: "Sequences of [C]ontinuation are preferred over sequences of [R]etaining, and sequences of [R]etaining are to be preferred over sequences of [S]hifting." [14] With respect to the coherence, naturalness and the smoothness of the discourse based on these types of Transition Walker, Joshi & Prince (1998) declare a Continue type to be the preferred and most neutral one. Second most natural one is Retain followed by Smooth-shift and Rough-shift. For the natural discourse is typical to combine particular types of Transition, the discourse based on a single type would be very unnatural and "boring".

3. CENTERING-THEORY-BASED ANNOTATION

Centering theory serves to me as the basic approach to the text relations' development analysis. However, it is necessary to use another approach in addition to be able to construct Centering-theory-based analysis of Czech, because Centering theory operates with such input data as subject function, coreferential-chain-membership etc. that have to be classified first, after that Centering annotation can be applied, the pre-annotation is the necessary condition. I am going to present a) some general principles of the analysis, b) data I used, c) pre-annotation and annotation concepts, d) annotation schema, and e) example, which illustrate problematic aspects of the annotation.

[12] Kruijff-Korbayová, I & Hajičová, E 1997, 'Topics and Centers: A comparison of the salience-based approach and the Centering theory', *The Prague Bulletin of Mathematical Linguistics*, vol. 67, no. 1, pp. 28.
[13] Notation was modified according to the Walker, Joshi & Prince (1998) style. Original uses the U_n, U_{n+1} notation.
[14] Kruijff-Korbayová, I & Hajičová, E 1997, 'Topics and Centers: A comparison of the salience-based approach and the Centering theory', *The Prague Bulletin of Mathematical Linguistics*, vol. 67, no. 1, pp. 28.

3.1 Principles of annotation

To apply Centering theory into practice, there are few methodological restrictions, which have to be taken into account. First of them is that the annotation applies only to the noun-like parts of speech (nouns, pronouns, adjectives, numerals). That is because Centering theory operates with the concept of entities and Centers based on the principle of noun-like parts of the utterance. Second methodological feature is that each utterance should be classified according to the Centers of Attention separately, because the Centers classification of particular utterance is independent of the previous one. There is only one exception – that is the backward-looking center determined by the classification of Centers in the previous utterance (forward-looking centers ranking). Third aspect considers the Centering theory concept of Transitions – it is based on the boundaries between two utterances. This point is not reflected in the part of project that I am presenting, because this paper focuses on the Centering-based-annotation specifically, other parts of project are not in the center of interest of this paper.

3.2 Data

For syntactic and text linguistics research of Czech language material it is possible to use the only syntactically annotated corpus of Czech, Prague Dependency Treebank (hereinafter referred to as PDT) administrated at Charles University in Prague. PDT is based on the functional approach of Prague School (Prague School of Functional and Structural Linguistics) and the annotation follows the principle of Functional Generative Description (Sgall 1964, 1967), every utterance is coded in the form of dependency tree with atributes.

This corpus captures information usable for the Centering-theory-based analysis: a) information about constituent function of each node (entity) at the analytical layer, and b) information about information structure status of each node (entity) at the tectogrammatical layer. In addition, it is also possible to use the information about role in coreferential chains, which is captured at tectogrammatical layer too. Another great advantage of choosing this corpus and its theoretical background is a) that this corpus operates with automatic tools for the annotation (constituent function at the level Centering-theory-based annotation requires can be performed automatically, even the information structure analysis can be partially automatically annotated), and b) that the strategy in PDT is to generate omitted nodes at tectogrammatical layer (Czech is typical subject-omitting language, so for the purpose of the analysis it is quite important to be able to reconstruct it).

Constituent structure is within the PDT analytical layer captured as afun attribute. In accordance with the Centering theory approach (dominant position of subject within Centering theory) the analysis distinguishes only subject from other constituents even though the automatic annotation assigns all kinds of constituent functions available within the PDT classification. Information structure is captured by the attribute tfa, which acquires three possible values (t for contextually bound entity, c for contrastively contextually bound entity and f for contextually non-bound entity).

3.3 Annotation

Annotation exhibits two phases: because of the dependency of the Centering-theory-based annotation on other metalinguistic information, it is necessary to annotate data according to the constituent and information structure first (first phase), then it is possible to perform the Centering-theory-based annotation (second phase). Annotation of constituent and information structure was performed automatically and followed the principles applied within PDT annotation (rules, attributes tfa and afun, values). This pre-annotation serves as the basic level for the Centering-theory-based annotation.

Centering-theory-based annotation had two basic levels itself – first of them is the automatic analysis following the rules I designed. Second step was to correct the annotation manually, if necessary.

Rules for the automatic annotation

I designed mutually independent rules considering constituent and information structure. In case of constituent structure, there is the Subject rule: Any entity, which fulfills the function of the subject within the particular sentence, is classified as preferred center. Implicit implication of this rule combined with the Centering theory principles (specificity of backward-looking centers, residual

nature of forward-looking centers) is that any other constituent function leads to the classification as forward-looking center.

The procedure was pretty much the same within the information structure, the entity judged as contextually bound (does not matter whether just bound or contrastively bound, i.e., tfa attribute can have both possible values t and c) was classified as backward-looking center. This classification follows the principle of Centering theory pronoun rule, because if there was any pronoun in the sentence, it would be judged as contextually bound (tfa = t), or contrastively contextually bound (tfa = c) already within the pre-annotation. Implicit implication of this rule is that the rest (i.e., tfa attribute has a value f and encodes the contextually non-bound entity) is classified as forward-looking center.

There is one additional rule with special status, because it has less serious value than previous two: if there is any entity, the expression of which creates the part of coreferential chain, than this entity is declared to be the backward-looking center of particular utterance. There are two significant restrictions in case of this additional rule: the coreferential relation is restricted to the previous utterance (because of the strict locality principle of Centering theory) and the judged entity (its representation in the particular utterance) has to form the anaphoric part of particular coreferential chain, not its antecedent (because the backward-looking center has to be contextually anchored).

Problem is that PDT annotation assigns the tfa attribute to every single node in the tree structure, therefore there are many backward-looking center candidates in the utterance (all nodes with the tfa attribute value t or c). Some of them can be eliminated by the nature of the entity (noun-like parts of speech), but unfortunately some of them do not. The situation when the automatic annotation predicts more backward-looking centers in the single utterance violates the unique backward-looking center restriction, but it is not possible to eliminate this phenomenon automatically, it has to be provided by manual correction.

3.4 Annotation schema

Each entity was first automatically classified as preferred center, backward-looking center, forward-looking center, or combination of previous values (e.g., extreme case would be: subject rule classifies the entity as preferred center, information structure rule can classify the entity as forward-looking center, and coreferential chain rule can predict the backward-looking center value to the entity). The possibility of more than one value assigned to the entity is the reflection of the fact that rules operate separately and that it is not possible (at the level I work on) to treat the disambiguation automatically.

Category	Clasification	Values
t_lemma	entity	[lemma]
afun	constituent function	Sb, Obj, Adv, Atr
tfa	contextual boundness	t, c, f
C_afun	Center of Attention based on constituent structure	Cp
C_tfa	Center of Attention based on information structure	Cb, Cf
C_cor	Center of Attention based on coreferential relation	Cb
pos	part of speach	C.*, N.*
coref_tfa	contextual boundness of the antecedent	t, c, f
coref_afun	constituent function of the antecedent	Sb, Obj, Adv, Atr
deepord	order in the deep structure	[number]
lex_anode_ord	order in the surface structure	[number]

Table 2: Annotation schema

Source: Alena Poncarová

For further analysis, it is very important to be able to reconstruct the reason why particular entity was classified as particular type of Centers of Attention (which rule and value is this classification based on, etc.) That is why the annotation schema captures except lemma, values of attributes coding constituent function (afun) and information structure (tfa) and the part of speech (pos) even the antecedents' constituent and information structure, if coreference occurred (coref_afun, coref_tfa) and the information about both deep and surface word order (deepord, lex_anode_ord). But the most important are the C_tfa, C_afun, and C_cor categories, which capture the Center of Attention classification based on the information structure (C_tfa), constituent structure (C_afun), and coreferential relations (C_cor).

3.5 Case study

Let me now introduce the example, which clarifies the annotation process. I will focus on the second utterance (utterance number 1), the first one (utterance number 0) serves only as referential, preceding, utterance.[15]

(0) *Potential errors can be eliminated even faster.*

(1) *Further improvement is related to the forthcoming amendment of the Securities Act.*

In Czech there are four entities in the utterance: *improvement, amendment, Securities* and *Act*. Automatic pre-annotation classified the *improvement* as subject of the sentence and therefore it was marked as the preferred center. This result of the subject rule is in accordance with Centering theory approach. However there are two candidates for the backward-looking center, i.e., *improvement* and *Securities*, and Centering theory does not offer any clue which of them is the proper one, both are plausible, neither of them is pronominalised, neither of them is the member of coreferential chain, etc.

How to solve this problem with unique backward-looking center restriction? There are three possible logical solutions. First of them is to develop "super-rule", which could be used when nothing else works, such as "If the individual rules lead to contradictory conclusions, follow the principle of surface word order: The closer the expression is to the previous utterance, the more likely it becomes a backward-looking center." But this solution is not completely plausible with Centering theory, because it would be unfounded, arbitrary selection of one of the principles at the expense of the other.

The second possibility is to follow the previous utterance structure. This solution is plausible with Centering theory approach, because it respects the locality principle, preserves the same type of analysis, etc., but its weakness is that rules are mixed across utterances. Each utterance should be classified separately first and after that the comparison takes place.

The third solution is to consult larger part of discourse arrangement and to map the structure of existing context to the particular utterance (e.g., if there is a larger part of text considering *improvement strategies and tools*, it would be preferable to classify *improvement* to be the backward-looking center than to classify *Securities* that way). However, this possibility violates the basic principle of Centering theory, i.e., the locality principle.

The concept of "super-rule" seems to be unsystematic; the larger discourse parallel is based on principles, which are completely strange to the Centering theory. That is why the acceptable solution seems to be the second option, to consult the previous utterance, because the relation based on neighborhood is applied at many opportunities within Centering theory and is inherently anchored in the methodology.

Applying the previous-utterance-consultation to the example presented above, it is necessary to consult the utterance number 0: *Potential errors can be eliminated even faster*. Neither *improvement* nor *Securities* are present in the previous utterance, not even in the omitted form. It means, that neither of them has the right to be classified as backward-looking center and both are classified as forward-looking centers. The situation any utterance has no backward-looking center should be reserved for the

[15] Original version in Czech: *Případné chyby lze odstranit ještě rychleji. Další zlepšení souvisí s chystanou novelou zákona o cenných papírech.*

first utterance in the discourse. But it could happen that there is no backward-looking center in the utterance even if this particular utterance is not the initial one. In any case, it is methodologically smoother to leave the utterance without backward-looking center than to unjustifiably classify any entity as the backward-looking center. Manual correction like this one has to be applied on every single utterance in the sample, or at least it has to be checked whether it would be needed.

4. EVALUATION

As I tried to demonstrate, Centering theory offers in many ways useful tools and methodological background for the analysis of Czech. However, there are few aspects, which have to be revised, most problematic one is the application of unique backward-looking center restriction. I tried to point to the weaknesses of Centering theory and to propose the modifications, which could offer improvements.

With respect to the annotation schema, there were no problems with such operations as a) entities identification, because noun-like parts of speech are identified with no problems; b) preferred center classification, because subject identification works with no problems; c) forward-looking centers classification, because forward-looking center type has no restrictions (except the requirement to be explicitly expressed on the surface) and it is the most general category.

The other way round, I had to face problems associated with Centering theory. For example there are problems with a) backward-looking center classification, because Centering theory does not offer clues for the disambiguation in the situation with multiple candidates; b) multipronoun utterances, i.e., pronoun rule application, because there is no clue to determine, which of the pronouns is the proper one to become the backward-looking center; c) coreferential relations, where the larger part of text take part, because Centering operates at the level of utterances exclusively and does not offer any approach to apply in such a situation.

4.1 Centering theory modification

These problematic parts of the annotation led me to try to modify parts of Centering theory to make it consistent with Czech data. It is important to emphasize, that Centering theory pronoun rule and claims and restrictions dealing with Centers of Attention as presented above are rather recommendations than strict rules, so they lead to different results very often. That causes many problems and that is why I try to add more aspects into the applied perspective to make them less contradictory. First, I consider it necessary to try to solve the multiple backward-looking center candidates situation.

Multiple pronoun situation

If there is only one entity – it does not matter whether pronominalised once or twice within single utterance – it becomes the backward-looking center automatically. It is the prototypical application of pronoun rule. If there are more entities pronominalised, there should be secondary criterion, which would help to decide, which entity is the "more backward-looking-center-like". As a subordinate criterion, I suggest to use the subject function of the previous utterance: If one of the entities, which candidate to become a backward-looking center of particular utterance, is the subject of the previous utterance, it is the argument for this entity to be classified as backward-looking center of particular utterance. This approach applies the same principle as was applied in the example above (there was additional perspective of previous utterance applied, in this case additional perspective of another already existing principle applies).

This rule considering subject of the previous utterance should apply only in case of multiple pronouns situation, otherwise the pronoun rule should apply regardless of which entity is the subject of the previous utterance. Proposed modification is not hundred percent solid (if neither of the backward-looking center candidates is subject of the previous utterance, this modification does not improve the approach), but it could help to solve at least some problematic cases. If this approach is not helpful, it would be necessary to check the coreferential-chain-membership: any entity involved in coreferential

chain has a better chance to become a backward-looking center than the entity, which is not involved in such a text relation.

Multiple backward-looking center candidates

In case there is no pronoun but still there are multiple backward-looking center candidates in the utterance (in the form of full expressions), relational criterion should apply: the previous utterance arrangement should be consulted and the entity, which is contextually more bound, should become the backward-looking center. This subsidiary criterion is possible to apply only when at least one entity of the backward-looking candidates is present in the previous utterance too. If the level of contextual boundness is not helpful (i.e., two or more entities have approximately the same level of boundness; or the level is not clear), the criterion of coreferential-chain-membership should apply.

5. CONCLUSION

My research considers the analysis of authentic Czech texts with respect to the syntactic and text linguistics analysis of text development. It is based on the methodological background of Centering theory. This paper aims to present small part of that research: Centering-theory-based annotation of Czech.

In my paper I have focused on the applicability of Centering theory concepts onto Czech language material and tried to evaluate Centering theory with respect to its usability for such an analysis. The essential feature of Centering theory which seemed to be the most difficult problem, i.e., the one hundred percent dependency of Centering-theory-based annotation, turned out to be uncomplicated, because Prague Dependency Treebank has a quite developed system of annotation.

I have presented Centering theory as a concept, which could offer useful perspective while used to analyze authentic language data (such as Centers of Attention perspective, realization and direct realization principles, etc.). But Centering theory brings problematic aspects too (the most complicated one is the unique backward-looking center restriction and other elements connected with it such as pronoun rule application, multiple members of coreferential chains within the same utterance, etc.).

That is why I proposed modifications of certain aspects of Centering theory:

(1) in the situation of multiple pronouns representing different entities within the same utterance, the arrangement of Centers of Attention of the previous utterance should be consulted: subject of the previous utterance has more chance to become a backward-looking center;

(2) in the situation of multiple entities, which could potentially become the backward-looking center of particular utterance, information structure of the previous utterance should be consulted: the more contextually bound the entity is, the more likely it becomes a backward-looking center;

(3) in case that arrangement of constituent structure of the previous utterance or the level of contextual boundness of previous utterance do not help to disambiguate the single backward-looking center, the coreferential-chain-membership criterion should apply: the more an entity is involved in any coreferential relation the more likely it becomes a backward-looking center.

To conclude Centering theory proved to be a methodological approach that, with several adaptations, may well serve in analyzing authentic language material – even so different from the language Centering theory is primary based on. Of course, there are still many aspects that have to be examined and tested, but for now I postpone them.

6. LIST OF USED ABBREVIATIONS

Cb – backward-looking center

Cf – forward-looking center

Cp – preferred center

D – discourse

PDT – Prague Dependency Treebank

U_i – particular utterance

U_{i-1} – immediately preceding utterance

U_{i+1} – immediately following utterance

7. ACKNOWLEDGEMENT

I have to thank Martin Popel for his very willing assistance with the automatic part of annotation and Šárka Zikánová for her guidance and valuable advice. Last I have to mention Pavlína Jínová – without her I would never be interested in Centering theory.

This output was created within the project *Interpersonal communication in interdisciplinary perspective*, subproject *Centering and Czech – syntactic analysis* (No. 26910102) solved at Charles University in Prague, Faculty of Arts from the Specific university research funds in 2015.

8. REFERENCES

Ariel, M 2013, 'Centering, accessibility and the next mention', *Theoretical linguistics*, vol. 39, no. 1, 2, pp. 39-58.

Brennan, S 1995, 'Centering Attention in Discourse', *Language and Cognitive Processes*, vol. 10, no. 2, pp. 137-167.

Cote, S 1998, 'Ranking Forward-Looking Centers', in M. Walker, A. Joshi, E. Prince (eds.), *Centering theory in Discourse*, Clarendon Press, Oxford, pp. 55.

Grosz, B, Joshi, A & Weinstein, S 1995, 'Centering - A Framework for Modeling the Local Coherence of Discourse', *Computational linguistics*, vol. 21, no. 2

Kruijff-Korbayová, I & Hajičová, E 1997, 'Topics and Centers: A comparison of the salience-based approach and the Centering theory', *The Prague Bulletin of Mathematical Linguistics*, vol. 67, no. 1, pp. 25-50.

Poncarová, A 2014. 'Centra pozornosti v teorii centeringu', in J. Křivan et al. (eds.), *Lingvistika Praha 2014*, Univerzita Karlova v Praze, Praha. Available at: http://lingvistikapraha.ff.cuni.cz/node/206 [cit. 2014-12-24]

Rysová, K 2014, 'O slovosledu z komunikačního pohledu', Ústav formální a aplikované lingvistiky, Praha. Available at: https://ufal.mff.cuni.cz/books/preview/2014-rysova_preview.pdf [cit. 2015-03-02].

Sgall, P 1964, 'Generativní systémy v lingvistice', *Slovo a slovesnost*, vol. 25, no. 4, pp. 274-282. Available at: http://sas.ujc.cas.cz/archiv.php?art=1315 [cit. 2015-03-02].

Sgall, P 1967, '*Generativní popis jazyka a česká deklinace*', Academia, Praha.

Walker, M, Joshi, A. & Prince, E 1998, 'Centering in Naturally-Occurring Discourse: An Overview', in M. Walker, A. Joshi, E. Prince (eds.), *Centering theory in Discourse*, Clarendon Press, Oxford, pp. 2.