Dynamics in the meaning of sentence and of discourse

Petr Sgall
sgall@ufal.mff.cuni.cz

From a linguistic viewpoint it may be stated that most different elements or parts of the system of language itself and of its use in communication are dynamic. The aim of the present contribution is (1) to discuss the impact of dynamics on the sentence structure, pointing out that, nevertheless, its core can be described as having a substantially simple structure, so that the relative easiness of the child's acquisition of language may be described as based on a pattern not far more complex than systems that are understood as innate on independent reasons, and (2) to illustrate how the dynamics of the discourse patterns can be seen as connected with a finite mechanism enabling the herarer to identify the coreferential antecedents.

1. Dynamics within the sentence

1.1. The intrinsic dynamics of communication, present in the utterance (sentence occurrence), is also reflected in the structure of a sentence as a type, since sentence is anchored in the context by the **topic-focus** articulation (TFA, information structure) of the sentence. TFA reflects the 'given – new' strategy, but differs from it in belonging to the systems of individual language, rather than to the domain of cognition.

In the linguistic descriptive **framework** of Functional Generative Description (see Sgall et al. 1986, Hajičová et al. 1998), which represents a continuation of important insights gained by the classical Prague School of structural and functional linguistics and uses a dependency based syntax, we attempt at an appropriate handling of most different combinations of sentence parts belonging either to topic or to focus. TFA is understood as one of the aspects of underlying sentence structure, expressed by an interplay of word order and of specific features of sentence prosody, which correspond to the hierarchy of 'communicative dynamism' (underlying word order) as its means of expression. The underlying (tectogrammatical) representation (TR) starts with topic proper and proceeds to focus proper (the most dynamic item).

1.2. TFA is **semantically relevant**; it can be interpreted on the basis of the relation of **aboutness**: each of the elements of the underlying structure of the sentence can be understood as 'contextually bound' (cb) or 'non-bound' (nb), i.e. as the linguistically patterned counterpart of what in the layer of cognition appears as 'given' and 'new' piece of information, respectively. In the prototypical case, the nb items are included in the focus, and the cb items belong to the topic. Thus, "contextually bound" should not be understood in a straightforward etymological way (see Sgall, Hajičová and Panevová 1986, Ch. 3): a nb item may well be "known" in a cognitive sense (and thus e.g. expressed by an anaphoric pronoun), and a cb item may refer to an entity known from the situation, rather than from the preceding co-text, cf. the cb pronoun *we* (an indexical, the reference of which is – although with indistct boundaries – determined by the utterance itself) and the nb *her* and *him* (i.e. the foci of the two

coordinated clauses) in (1):

(1) When the two young people entered, we recognized only her, but not him.

The semantic(-pragmatic) interpretation of the sentence may be based not just on its predicate-argument pattern, but rather on a scheme of the shape F(T), with the negative counterpart ~F(T) (a formal treatment of this aspect of the interpretation was presented by Peregrin 1994; 1996, see also B. H. Partee in Hajičová et al. 1998, pp. 48-53 and Partee 1998).

As the following examples show (with capitals indicating a marked position of sentence stress), TFA is relevant for the truth conditions, so that it cannot be relegated to a "wastebasket" of pragmatics or of style.

(2)(a) I work on my dissertation on Sundays.
   (b) On Sundays, I work on my dissertation.

(3)(a) English is spoken in the Shetlands.
   (b) In the Shetlands, ENGLISH is spoken.

(4)(a) Dogs must be carried.
   (b) DOGS must be carried.

1.3. Furthermore, within F even the underlying **word order** (which differs from the surface order e.g. in what concerns such shallow rules like "adjective before noun", or in placing F as the carrier of the (typically falling) sentence stress in a marked (not clause-final) position), is fixed, corresponding to the 'systemic ordering'. In this ordering, as documented by (5) and (6), Means (and also Directional.from) precede Directional.where-to, cf. the fact that in the marked (b) variants the *to*-groups can only belong to T; in other words, such a question like *What did they do?* can only be answered by (5)(a) or (6)(a), see Sgall et al. (1995):

(5)(a) We went by car to a lake.
   (b) We went to a lake by car.
(6)(a) They moved from Chicago to Boston.
   (b) They moved to Boston from Chicago.

The set of underlying representations, TRs, which in the prototypical case (if no coordinated structure is present, etc.) can be described as having the shape of projective dependency trees (see Hajičová et al. 1998), may be specified by a restricted set of very general rules, either in a declarative way (using unification and checking the conditions specified in the valency frames of the head words – the presence of each obligatory dependent, the just mentioned order of dependents in F, the saturation of every argument, which can be present just once as depending on an occurrence of its head), or in the form of a generative procedure not stronger than a pushdown-store generator, deriving the tree from top to bottom and left-to-right, with every node generated more to the right than its sister nodes have been. In either case, every CB node precedes its mother node and its NB sisters.

The TRs are disambiguated (just certain peripheral cases, like the marked patterns of quantifier scopes, being left underspecified) can be understood as 'the meaning of the sentence' in the sense that their set is appropriate as input for a procedure of semantic(-pragmatic) interpretation (cf. Sgall 1994). Thus, TRs may be understood as 'linguistic meaning' (LM), which, if the reference of the referring

expressions included in them is specified, determine propositions (sentence intensions).

1.4. Let us then assume that it is possible, in principle, to construct a grammar that recursively enumerates the sentences of the language described and assigns each of them a structural description determining also its LM (see Hajičová and Sgall 1980, Sgall et al. 1986). The next step concerns the relationship between LM and the **cognitive content** of the sentence. If we identify the underlying structure and LM, we are committed to claim that every syntactically well-formed sentence has a meaning. This is often denied with sentences such as the following:

(7) Your ashtray was shooting.
(8) The cat was eating grass.
(9) Jim causes the sky to be blue today.
(10) John annoyed me by the sky not being blue.

However, the grammatical well-formedness of such sentences cannot be denied, since their counterparts such as (11) - (14), which differ from them just in individual morphological (esp. modal) oppositions, clearly are well-formed and can even be used as expressing true assertions:

(11) Ashtrays cannot shoot.
(12) Do cats eat grass?
(13) Jim cannot cause the sky to be blue today.
(14) John could not annoy me by the sky not being blue.

It can be maintained that the sentences (7) - (10), being well-formed, do carry meaning in the sense of LM. The reader may understand them as sentences, i.e. as strings which s/he can translate into another language, or the occurrences of which in discourses s/he is able to check as for their truth value. Under normal conditions (as referring to a subcollection of possible worlds including the actual one) none of them can convey a true statement (without exaggerating or using them in a figurative sense). This impossibility concerns the relationship between LM and cognitive content, rather than sentence structure; the linguist's task thus does not include the identification of conditions under which a sentence can carry a true assertion. Questions of this kind belong to the interdisciplinary domain of the sematico-pragmatic interpretation, rather than to linguistics in the narrow sense. We prefer the term 'semantico-pragmatic' interpretation, since it takes into account that knowledge-based inferencing and reference specification are relevant.

1.5. To abstract LM from the truth conditions of sentences, the concept of **synonymy** (identity of LM) is useful; it may be specified on the basis of substitutability 'salva veritate' in all contexts except the quotational ones. A caveat concerns contexts such as The ancients didn't know that Hesperus was Phosphorus, in which the two names are mentioned, rather than used, i.e. their 'content' is spoken of; cf. the critical remarks on Carnap's 'intensional isomorphism' in Sgall et al. 1986:35-40). The language system can then be viewed as including LM, structured primarily by valency relations (cf. 'deep cases' or 'theta roles') and containing disambiguated values of number, tense, modality, etc. LM (as a level of the system of language) can be distinguished from what is conditioned by inferencing, by resolution of indistinctness

and by other aspects of the way from LM to pragmatico-semantic representation, be it based on truth conditions or perhaps on another framework.

1.6. If LM is accounted for by a dependency based sentence structure with complex node labels (in which function words are represented by indices of lexemes), then the **core** of language can be captured as based on a **simple pattern**, coming close to systems that may be understood as innate on independent reasons (proposition calculus). Even if the topic-focus articulation (based on the 'given-new' strategy) is seen as one of the aspects of LM, the LM patterning of sentences can then be described as tree-like objects that may be univocally represented just by bracketted strings of symbols. Each of these symbols comprizes a lexical component and syntactic and morphological indices, cf. e.g. the (simplified) representation (16) for the sentence (15), in which along with different relations of syntactic dependency also coordination (Conjunction) is present:

(15) Jim and Jane's son, who were present there, belong to the best specialists,
(16) ((Jim ((Jane).Appurt son.Sing.Def).Conj (.Restr (who.Plur).Act (present).Obj be.Pret.Decl.Imperf (.Loc there))).Act belong.Pres.Decl.Imperf (.Dir specialist.Plur.Def (.Restr good.Superl))

Note: Every dependent item or collocation is enclosed in its pair of parentheses, which denote (i) either a dependency relation with its index attached to that parenthesis that is oriented towards its head: Appurt(enance, broader than Possession), Restr(ictive Adjunct), Act(or), Obj(ective), Loc(ative) Dir(ectional-where-to), etc., (ii) or a coordination construction with its symbol attached to the right parenthesis: Conj(unction), Disj(unction), etc. The indices attached to the lexical item (here indicated only by its orthographic form, which has to be substituted by a symbol for lexical meaning) correspond to the morphological values: Sing(ular), Def(inite), Pret(erite), Decl(arative), Imperf(ective), Superl(ative), etc. The items written to the left of their heads are contextually bound (in topic, in the prototypical case), those to the right of their heads are non-bound (in focus or local focus).

   The transition between LM and the surface forms of sentences can be handled by a set of rules (including movements) that does not surpass the generative power of one or two (subsequent) pushdown transducers, so that the whole description of language is not much stronger than context-free (cf. Plátek and Sgall 1978).

1.7. The impact of the topic-focus articulation on the semantico-pragmatic interpretation can be described in the frame of this or that version of post-Montaguian intensional semantics. It is then possible to account also for the difference between a **presupposition** and an **allegation**, i.e. specific kinds of entailment, fulfilled or not by subcollections of possible worlds for individual occurrences of sentences (an allegation is an assertion A entailed by an assertion carried by a sentence S, with which the negative counterpart of S entails neither A nor its negation, see Hajičová 1993, Partee 1996). Often a definite noun group triggers a presupposition if it occurs in the topic of a sentence, but only an allegation if it belongs to the focus (as is known from the discussions on sentence pairs such as (17) and (18), see Sgall et al. 1986:82-86).

(17) The King of France is (not) bald
(18) The exhibition was (not) visited by the King of France

Issues of the transition from LM to truth-conditional semantics have been discussed by Peregrin (1994;1996). The possibility of such a transition supports the view that truth conditions of sentences can well serve as a reliable starting point for a theory of semantics/pragmatics, i.e. that the legacy of what was presented by A. Tarski, R. Carnap and others still offers certain advantages, even though the truth conditions of sentences have to be relativized in some way to individual subclasses of contexts, if the content of utterances is to be described systematically.

1.8. In Hajičová, Partee and Sgall (1998) a transduction of LMs into **Tripartite Structures** (Operator - Restrictor - Nuclear Scope) is characterized, as well as the primary and secondary positions and scopes of focus sensitive particles (operators) such as *only, even*, etc. Let us briefly recall some of the relevant sentences discussed there (with their chosen LMs) and specify (with a maximally simplified notation) what parts of their individual readings belong to the Operator (O), Restrictor (R) and Nuclear Scope (N) of the corresponding tripartite structures.

(19) John only sits by the TV.
(19') O only, R John, N sits by the TV
(19") O only, R John sits, N by the TV

In ex. (19) the particle occupies its prototypical position in LM, so that the focus of the particle is identical with the focus of the sentence on either reading, i.e. with the verb included in the focus in (19'), and in the topic in (19"). Example (20) is more complex, displaying several versions (the different sentences (a) - (d)); also here the focus of the particle is identical to the focus of the sentence on most of the readings, although in (c) and on one reading of (b) the focus is limited to a part of the embedded clause (to the dancer, cf. the tripartite structures (b") and (c')); in (d) the focus of the sentence includes Mary, while the position of said is ambiguous.

(20)(a) He only said that Mary liked the DANCER.
   (b) He said that Mary only liked the DANCER.
   (c) He said that Mary liked only the DANCER.
   (d) He only said that MARY liked the dancer.

(20')(a') O only, R he, N said that Mary liked the dancer
    (a") O only, R he said, N that Mary liked the dancer
    (b') O only, R he said [about] Mary N [that she] liked the dancer
    (b") O only, R he said that Mary liked, N the dancer
    (c') O only, R he said that Mary liked, N the dancer
    (d') O only, R he [about who] liked the dancer, N said [that this was] Mary
    (d") O only, R he said [about who] liked the dancer, N [that this was] Mary

In other examples the particle is in the topic, so that its focus does not cross the boundary between the topic and the focus of the sentence; we use the ASSERT operator of Jacobs (1984) in the interpretation of these sentences, in which the position of the main operator is not occupied by any overt particle:

(21) (What did even PAUL realize?) Even Paul realized that
    Jim only admired MARY.

(21') O ASSERT, R (O even, R realized, N Paul), N (O only, R Jim admired, N Mary)

(22) (A: All of us have seen that Jim only eats vegetables.) B: If even Paul realized that Jim only ate vegetables, we would choose another RESTAURANT.
(22') O if, R (O even, R (O ASSERT, R (O only, R Jim ate N vegetables), N realized), N Paul), N we would choose another restaurant

This approach allows for a description of more complex examples, such as the German sentence brought into discussion by J. Jacobs; we understand here *Peter* to constitute the bearer of the intonation center of the sentence (falling stress) and *Roman* ['novel'] to bear a rising stress that marks a contrastive (part of) topic:

(23) Sogar PETER kennt nur einen Roman von Goethe.
   [Even PETER knows only a novel by Goethe].
(23') O even, R (O only, R know [what] by Goethe, N a novel), N Peter

In other cases the particle itself carries the intonation center, and thus constitutes the Nuclear Scope:

(24) She did it TOO.
(24') O ASSERT, R she did it, N R holds too [besides other assertions about her and it]

(25) They ALSO were there.
(25') O ASSERT, R they were there, N R also holds [besides other assertions about them and the given place]

1.9. It may be claimed that **two reliable starting points** for a semantico-pragmatic theory can be used if the classical approaches of European structural linguistics and of truth-conditional semantics are not abolished. Linguistics then would not be weakened by losing its cumulative character, and the substantial impact of the interactive features of language would be duly reflected both in the analysis of a sentence (with its topic-focus articulation as an aspect of its syntactic structure) and in that of an utterance (with the degrees of salience as a background for coreferential links in the discourse).

1.10. The work on a procedure of morphological and syntactic annotations of sentences from the large Czech National Corpus in the **Prague Dependency Treebank** (PDT) gives us an occasion of a useful and effective checking and enrichment of the descriptive framework, see Böhmová and Hajičová (1999), Hajič (1998; in press). In the annotations, we use three values of a specific TFA attribute, assigning every lexical (autosemantic) occurrence one of them:

   cb or t - for 'contextually bound' (prototypically in Topic, T),
   c - for 'contrastive (part of) Topic',
   nb or f - for 'non-bound' (typically in Focus, F)

The main verb and any of its direct dependents following it belong to F in the unmarked case, they carry index f.

Our analysis covers not only simple sentences, but also complex ones. It is not restricted to cases in which either T or F would correspond to a single constitutent, as example (26) illustrates:

(26) České radiokomunikace   musí v tomto roce rychle splatit dluh televizním divákům.
  Lit: Czech Radiocommunications must in this   year quickly pay   depth to-TV viewers.
     This year, Czech Radiocommunications have quickly to pay their debt to the TV viewers.

Fig. 1 gives a simplified variant of the tectogrammatical representation of (26), and (26') shows that it is possible to render such a dependency tree with a (one-to-one) linearization, in which every dependent item is closed in its pair of parentheses and the functor (denoting the kind of dependency relation: ACTor, OBJective, MANNer, BENefactive, ReSTRictive adjunct, etc.) is written as a subscript at the parenthesis oriented towards the head. It should be noted that function words, such as *muset* 'must' or *v* 'in' display no corresponding nodes in the represetation, but only indices as parts of the complex node labels. The indices Pres(ent), Necess(ity) and Plur(al) are just examples of one set of these markers, i.e. of the grammatemes, or values of morphological categories such as tense, modality and number. The counterpart of the preposition is the functor for a Temporal adjunct.
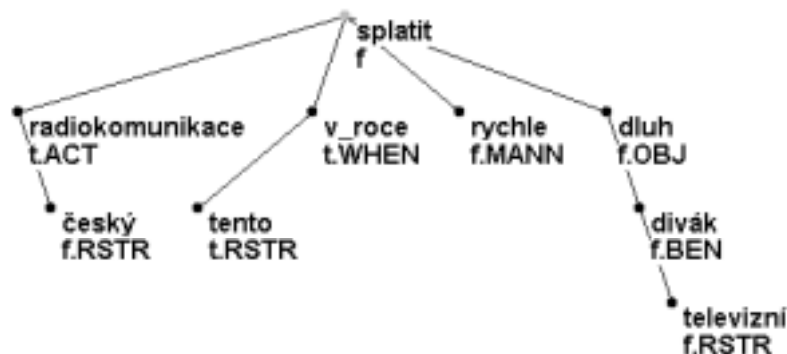

Fig. 1.

(26') ((České.f)$_{RSTR}$ radiokomunikace.Plur.t)$_{ACT}$ ((tomto.t)$_{RSTR}$ roce.t)$_{TEMP}$
     Czech   Radiocommunications   this   year
splatit.Pres.Necess.f ($_{MANN}$ rychle.f) ($_{OBJ}$ dluh.f ($_{BEN}$ (televizním.f)$_{RSTR}$ divákům.f))
pay   quickly   debt   TV   viewers

The basic features of underlying sentence structure, including the combinations of dependency with coordination and also the primary positions of focusing operators can be rendered in such a simple way (see Hajičová et al. 1998). The possibility of such an unambiguous linearization, which exhibits a pattern not substantially surpassing that of propositional calculus, shows that the core of the structure of language can be handled in a way that supports a hypothesis according to which the innate dispositions for language acquisition might not be much more complex than

the general human mental capacities. The non-prototypical layers of language (including the asymmetries between underlying structure and morphemics, phonemics, etc., such as synonymy, ambiguity, irregular word forms, idioms, as well as specific points in underlying structure itself, e.g. marked positions of focusing operators) may then be understood as being mastered by the learners step by step in the interactive communication. Thus, the acquisition of the mother tongue by children may be viewed as conditioned much more by communictive interaction (in the sense of Schnelle 1992) than by a complex system specific for the language faculty.

2. Dynamics in discourse

2.1. The views of the dynamics of discourse patterns as discussed nowadays in the approaches to discourse theory (starting with H. Kamp's Discourse Representation Theory, DRT) understand language as an interactive system serving human communication, with sentence occurrences anchored in context and representing an update operation on the hearers' memory states. Let us present just a few remarks to this highly significant step in the development of linguistics, as well as of semantics and pragmatics.

(i) It is useful consistently to distinguish between a **sentence** (as a type, as a unit of the system of language) and an **utterance** (an occurrence of a sentence within a specific discourse). While the latter can be regarded as displaying a content of its own, i.e. as expressing an update operation and connected with a truth value (if used appropriately, i.e. without a presupposition failure), the former is only connected with truth conditions. The sentence as such may be used in different contexts with regard to different possible worlds or situations and with different specifications of reference; without this, the occurrences of a declarative sentence (with the exception of sentences carrying analytic statements) can, in the general case, correspond to different truth values when used in different contexts.

(ii) The sentence as a type is characterized, as for its contextual possibilities, by its topic-focus articulation, which restricts the possibilities of its use.

(iii) The identification of the **discourse referents** by the hearer is based on the view discussed by D. Lewis, concerning the relevance of the degrees of **salience** of the elements contained (according to the speaker's assumptions) in the hearer's stock of information. The development of these degrees during the discourse (see Section. 2.3 below) can be captured to a certain degree on the basis of psycholinguistic tests and of an analysis of different kinds of discourse. Krahmer and Theune (1999) discuss a way to capture the degrees of salience in a formal framework. If the description of discourse structure is enriched in this sense, it is possible to specify the main factors of the finite mechanism a hearer uses to identify the reference of the referring expressions in an utterance. It may be stated that whenever the image of a referent is significantly more salient than the images of all other entities which may be referred to by the same expression (a pronoun, a noun or a noun group) then this expression may be used without blurring the reference.

(iv) The fact that this mechanism is based on degrees, i.e. on a partial ordering, makes it possible to specify the entity referred to more adequately than with an operator based on iota inversum. With an occurrence of a sentence such as (27) we do not need to claim that there is a single book and a single table in the universe of discourse.

(27) Jim put the book on the table.

It is sufficient to say that the hearers will be able to identify which book (and which table) is being referred to if the image of one book (table) is much more salient at the given time point than those of all the others.

(v) This can also help to avoid the difficulties connected with the fact that in specific cases, the hierarchy of the degrees of salience may include non-prototypical factors. It is not always so that the condition of a 'much higher' salience is met if only one of the objects of the given kind (books, tables) was mentioned in the just preceding utterance. Such a marked case can be seen in the following segment of a discourse:

(28) I was not able to imagine that Tom would pay such a sum. But the situation changed when Jim emerged with his charisma and his arguments. Now he HAS paid.

In such a specific case (with the connection between the action of paying and Tom as its Actor), the weak pronoun *he* clearly can be referentially identical to an antecedent that is not present in the preceding utterance. The relationship between a sentence boundary (full stop) and a coordinating conjunction can then be illustrated by (29)(a) and (b).

(29)(a) I could not imagine that Tom would pay such a sum. But the situation changed when Jim emerged with his charisma and his arguments, and now he HAS paid.

(b) I could not imagine that Tom would pay such a sum. But the situation changed when Jim emerged with his charisma and his arguments, and HAS paid.

Only in (29)(b) the (deleted) subject of the last clause refers in any case to Jim. In (29)(a) the possibility of its referring to Tom is present, as it is in (28). Thus, the use of a coordinating conjunction may differ semantically from an occurrence of an intersentential boundary.

As we have seen, there are several aspects in which the framework of a description of the discourse structure can be enriched if due attention is paid to the topic-focus articulation and to the degrees of salience.

2.2. The older approaches to the **identification of the referent**, working with *iota inversum,* certainly are not acceptable any more, since it is clear that the uniqueness of a possible referent is far from granted. It is not typical that a sentence such as (27) would be used in a context (and situation) in which just a single book can be mentioned, and the just discussed competition between Tom and Jim as possible referents of the pronouns in (28) - (29) points in the same direction. Moreover, it is not always so that the antecedent occurs in the preceding context; it may well be determined by the situation of the discourse, in other words, the term 'context' may be used as including both 'co-text' and 'consituation'.

Let us observe the following text segment (with the referring expressions relevant for our discussion printed in bold) as an example:

(30) In **the library he** entered **the reading room**, took **some books** from **the shelf** and put **them** on **the single desk that was free**.

The expressions *he* and *the library* have their referential antecedents in the preceding **co-text** (as we have seen, this is not necessarily the preceding sentence token);

the antecedent of *the reading room* is determined by an **associative** link, i.e. by a certain type of accommodation anchored in *library;* perhaps it is possible to work with an assumption of the speaker (or entailment) that the library has a single reading room, although this entailment is triggered by the positive utterance only, not by its negative counterpart, i.e. it is an allegation in Hajičová's (1993) sense, rather than a presupposition;

the expression *some books* is **indefinite** (specifying), exhibiting a new referent, not yet identified, although serving as a starting point for further coreference: we face the establishment of a new member of the scene (i.e. of the easily accessible part of the universe of discourse);

the group *the shelf* again finds its referent on the basis of accommodation, i.e. of an **associative** link; it is impossible to work with an entailment according to which a reading room of a library would contain a single shelf, but perhaps an association between *shelf* and *book* is relevant here (...*the shelf on which the book was placed*);

the referent of *them* is determined by co-text, and *books* is the single noun in plural that can be referred to by the pronoun (if no longer segment of the preceding co-text is seen as relevant);

the expression *the single desk that was free* is an explicit individual description, the definiteness of which is conditioned both by its content (with the restrictive adjunct clause) and by an associative link to *the reading room.*

2.3. It may be assumed, as we already mentioned, that there is a **finite mechanism** the addressee can use to identify the referents in cases such as those involved in Section 2.2. If the backbone of such a mechanism is seen in the hierarchy (partial ordering) of **salience**, then it can be understood that this hierarchy typically is modified by the **flow of discourse** in a way that was specified by Hajičová et al. 1982; 1995, Hajičová 1993). Among the attributes of the contextual anchoring of word tokens, there are their relationships to their coreferential antecedents, which can be described having in view that individual lexical occurrences with their discourse referents display different degrees of salience. Thus, in (30), the referents of the expressions *he, the library, them* can be understood to be more salient than other 'competing' referents, and this concerns also the anchoring antecedents of the accommodation of *the reading room* and *the shelf*.

In the flow of a discourse, prototypically, a new **discourse referent** emerges as corresponding to a lexical occurrence that carries f; further occurrences carry t or c, their referents being determined by their degrees of salience (although the difference between 0 and 1, i.e. between the lowest degrees of salience reduction, is not decisive).

It appears to be possible to capture at least certain aspects of this hierarchy by the following rules, formulated here in a tentative form, with x(r) to be read as 'the referent r has the salience degree x', ➔ denoting the change of salience connected with the given occurrence of sentence S, and $1 \leq m,n$:

(i) if r is expressed by a noun group (or a pronoun) carrying f, then n(r) ➔ 0(r);

(ii) if r is expressed by a noun (group) or pronoun carrying t or c, then n(r) ➔1(r);

(iii) if n(r) ➔ m(r) in S, then m+2(q) obtains for every referent q that is not itself

referred to in S, but is immediately associated with the item r present here;[1]

(iv) if r neither is included in S, nor refers to an associated object, then n(r) ➜ n+2(r).

2.4. The functioning of these rules[2] may be **illustrated** by a section of text from PDT starting with (26), given above. For the sentences following (26) in the corpus, we give here first their wording, adding indices that indicate the degrees of activation reduction of the individual ocurrences of referring examples in the subsequent utterances and then a simplified TR of each sentence.

(31) Jejich.1 vysílače.1   dosud pokrývají signálem.0 programu.0 ČT.1 2.0 méně
než polovinu.0 území.0   republiky.0.
     Their   transmitters hitherto cover   by-signal of-program ČT   2   less
than half   of-territory   of-Republic.

(31') ((jejich.t) vysílače.t) (dosud.t) pokrývají.f (signálem.f (programu.f (ČT.t (2.f))))
((méně.f (než-polovinu.f)) území.f (republiky.t))

(32) Na moravsko-slovenském pomezí.1   je   řada   míst.0,   kde   (oni.1)
nezachytí ani   první program.0 České   televize.1.
     On Moravian-Slovakian   borderline there-is number of-places where (they)
do-not-get even first   program   of-Czech Television.

(32') ((na-moravsko-slovenském.t) pomezí.t) je.f (řada.f (míst.f ((kde.t) (oni.t) (ne.f)
zachytí.f ((ani.f) (první.f) program.t ((České.t) televize.t)))))

(33) Do   rozdělení.1 federace.1   totiž   signál.1   zajišťovaly
vysílače.0   v SR.0.
     Until division   of-federation as-a-matter-of-fact signal.Accus provided
transmitters.Nomin in S(lovac)R(epublic).

(33') (do-rozdělení.t (federace.t)) (totiž.t) (signál.t) zajišťovaly.t (vysílače.f (v-SR.f)).

(34) Česká televize.1 žádá   urychlenou výstavbu.0   nových vysílačů.0.
      Czech Television requires quick   construction of-new transmitters.

(34') ((Česká.t) televize.t) žádá.f (((nových.f) vysílačů.t) výstavbu.f (urychlenou.f))

The development of salience reduction of the referents most frequently mentioned in (31) - (34) is characterized in Tab. 1, a preliminary form of which was presented in Hajičová and Sgall (2001); it includes numbers of salience reduction degrees and of those rules from Section 3 that are the main sources of the degrees.
     Two further remarks may be added, concerning details of our analysis that have not been discussed above and may not be directly found in the previous publications we refer to: (a) a noun group consisting of a head with t or c and of one or more adjuncts with f constitutes a referring expression as a whole, in the prototypical case, and gets degree 0, if it occurs in F; this concerns e.g. the group *vysílače v SR* ('transmitters in the Slovac Republic') in sentence (33), or *ČT 2* (*CTV 2*) in (31); here *2* is treated as an adjunct of *CT*; (b) the difference between the degrees 0 and 1 is not sufficient for a safe choice of reference, so that, e.g., the reference of the pronoun *jejich* (*their*) after (26) by itself is indistinct, and only inferencing helps to establish that *České radiokomunikace* (*Czech Radiocommunications*) are referred to

(viewers normally do not have transmitters at their diposal).

after	(26)	(31)	(32)	(33)	(34)
CRC	1 (ii)	1 (ii)	3 (iii)	3 (iii)	(3) (iii)
CTV	3 (iii)	1 (ii)	1 (ii)	2 (iii)	1 (ii)
CTV 1	2 (iii)	2 (iii)	0 (i)	2 (iii)	3 (iii)
CTV 2	2 (iii)	0 (i)	2 (iii)	2 (iii)	3 (iii)
viewer	0 (i)	2 (iii)	2 (iii)	3 (iii)	3 (iii)
sig.	3 (iii)	0 (i)	2 (iii)	1 (ii)	3 (iii)
CR	3 (iii)	1 (ii)	3 (iii)	3 (iii)	3 (iii)
CSF	-	-	3 (iii)	1 (ii)	3 (iv)
terr.	3 (iii)	0 (i)	2 (iii)	2 (iii)	4 (iv)
tr.	-	1 (ii)	2 (iii)	0 (i)	0 (i)

Tab. 1.

Abbreviations:

CRC - Czech Radio(tele)communications   CTV - Czech TV
CR - Czech Republic   CSF - (CS) Federation
CTV1(2) - 1st (2nd) program of CTV   tr. - transmitter
terr. - territory of CR   sig. - signal of CTV

Even with this short piece of discourse, its segmentation is reflected, if its first subsegment, discussed up to now (sentences (26) and (31) - (34)), is compared with its continuation, i.e. sentences (35) - (38), given below. While the first segment deals primarily with CTV and its signal (cf. the relatively high salience of *CTV, CTV1, CTV2, RC, signal* and *viewer* in most parts of the segment), sentences (35) - (38) are devoted to financial issues, as can be seen from the following facts: (a) *money* gets

degree 0 after (35), in which it functions as its focus proper (the most dynamic item), (b) *Czech crown* gets degree 1 after (36), in which it is an embedded part of the focus, and (c) the group *financial coverage* gets degree 1 in sentence (37).

We present the continuation without the TGTSs:

(35) Naše společnost může úkol splnit, ale chybějí nám peníze.
Our company can the-task.Accusative fulfil, but is-lacking us.Dative the-money.Nominative.

(36) Letos by výstavba technického zařízení v sedmi lokalitách stála 120 miliónů korun, ale můžeme uvolnit jen 80 miliónů.
This-year, would the-construction of-technical equipment in seven localities cost 120 million crowns, but we-can spend only 80 million.

 (37) Proto o finančním zabezpečení jednáme s Českou televizí, uvádí ekonomický ředitel Českých radiotelekomunikací Miroslav Cuřín.
Therefore about (its) financial coverage we-discuss with Czech Television, states the-economic director of-Czech Radiotelcommunications M. C.

(38) Dalších 62 miliónů korun si vyžádá výstavba vysílačů a převaděčů signálu v pohraničí.
Further 62 million crowns.Accusative Refl. will-require the-construction.Nominative of-transmitters and transferrers of-the-signal in the-border-area.

6. Conclusions

We are aware that, along with the rules characterized above, there are other factors that have to be investigated, which are important for different kinds of discourses. This concerns various aspects of the discourse situation, of domain knowledge, of specific textual patterns (with episodes, poetic effects, and so on). Factors of these and further kinds can be studied on the basis of the salience degrees, which are typical for basic discourse situations.

   In any case, we may conclude that it is useful for a theory of discourse semantics to reflect the degrees of salience. This makes it possible to distinguish the reference potential of referring expressions and thus the connectedness of the discourse. Discourse analysis of this kind may also be useful for application domains such as text segmentation (in accordance with topics of individual segments), or information retrieval (specifying texts in which a given topic is actually treated, rather than being just occasionally mentioned).

Footnotes

1 Only immediate associative links are taken into account for the time being, such as those between (*Czech*) *crown* and *money*, or between *TV* or (*its*) *signal* and (*its*) *viewer*.

2 These tentative rules, which have been presented at several occasions (starting with Hajičová and Vrbová 1982) for the aims of a further discussion, still wait for a systematic testing and evaluation, as well as for enrichments and more precise formulations. These issues may find new opportunities now, when e.g. a comparison with the centering theory gets possible and when a large set of annotated examples from continuous texts in PDT is available. An automatic derivation of such features can only be looked for after the lexical units included get a very complex and subtle semantic classification.

References

Böhmová A. and E. Hajičová (1999). The Prague Dependency Tree Bank I: How much of the underlying syntactic structure can be tagged automatically? The Prague Bulletin of Mathematical Linguistics 71, 5-12.

Hajič J. (1998). Building a syntactically annotated corpus: The Prague Dependency Treebank. In: Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová, ed. by E. Hajičová, 106-132. Prague: Karolinum.

Hajič J. (in press). Disambiguation of rich inflection (Computational morphology of Czech). Prague:Karolinum.

Hajičová, Eva, 1993. Issues of sentence structure and discourse patterns. Prague: Charles University.

Hajičová Eva, Hoskovec Tomáš and Petr Sgall (1995). Discourse modelling based on hierarchy of salience. Prague Bulletin of Mathematical Linguistics 64, 5-24.

Hajičová E., Partee B. and P. Sgall (1998): Topic-focus articulation, tripartite structures, and semantic content. Amsterdam:Kluwer

Hajičová E. and P. Sgall, 1980. A dependency based specification of topic and focus. SMIL (Stockholm) 1/2:93-140.

Hajičová E. and P. Sgall (2001): Topic-focus and salience. In: Proceedings of the 39th Annual Meeting of the association for Computational Linguistics and the 10th Conference of the European Chapter, Toulouse:CNRS, 268-273.

Hajičová E. and J. Vrbová (1982). On the role of the hierarchy of activation in the process of natural language understanding. In: *COLING* 82. Ed. by J. Horecký. Amsterdam: North Holland, 107-113.

Jacobs J., 1984. Funktionale Satzperspektive und Illokutionssemantik. *Linguistische Berichte* 91:25-58.

Jakobson R., 1964. Efforts towards a means-ends model of language in interwar Continental linguistics. In: Vachek J., ed., *A Prague School reader in linguistics.* Bloomington:Indiana University Press, 481-485.

Krahmer E. and M. Theune (1999), Efficient generation of descriptions in context. In:

R. Kibble and K. van Deemter (eds.), Proceedings of the workshop The Generation of Nominal Expression, associated with the 11th European Summer School in Logic, Language and Information.

Kruijff G.-J. M. and I. Kruijffová (2001): A hybrid logic formalization of information structure sensitive discourse interpretation. In: V. Matoušek, P. Mautner, R. Mouček and K. Taušer (eds.): *Text, speech and dialogue* 4, Berlin/Heidelberg:Springer, 31-38.

Luelsdorff Philip, ed., 1994. *The Prague School of structural and functional linguistics - A short introduction.* Amsterdam/Philadelphia:Benjamnis.

Partee B. H., 1996. Allegation and local accommodation. In: Partee and Sgall (1996), 65-86.

Partee B. H. (1998): Focus, quantification, and semantics-pragmatics issues. In: Bosch P. and R. van der Sandt, eds. : *Focus. Linguistic, cognitive and computational perspectives.* Cambridge University Press, 213-231.

Partee, B. H. and P. Sgall, eds., 1996. Discourse and meaning. Amsterdam/Philadelphia:Benjamins.

Peregrin J.. 1994. Topic-focus articulation as generalized quantification. In: Bosch P. and R. van der Sandt, eds. (1994) *Focus and natural language processing.* IBM Working Paper 7. Heidelberg: IBM Deutschland, 379-388.

Peregrin J., 1996. Topic and focus in a formal framework. In: Partee and Sgall (1996), 235-254.

Petkevič, V., 1995. A new formal specification of underlying structures. *Theoretical Linguistics* 21:7-61.

Plátek M. and P. Sgall, 1978. A scale of context sensitive languages: Applications to natural language. *Information and Control* 38:1-20.

Schnelle H., 1991. *Natur der Sprache.* Berlin: W. de Gruyter.

Sgall P. 1994. "Meaning, reference and discourse patterns". In: Ph. Luelsdorff (ed.). *The Prague School of Structural and Functional Linguistics.* Amsterdam/Philadelphia: J. Benjamins, 277-309.

Sgall P., Hajičová E. and J. Panevová (1986): The Meaning of the Sentence in Its Semantic and Pragmatic Aspects, ed. by J. L. Mey, Dordrecht:Reidel - Prague: Academia.

Sgall P., O. Pfeiffer, W. U. Dressler, and M. Půček (1995): Experimental research on Systemic Ordering. *Theoretical Linguistics* 21:197-239.