TOPIC, FOCUS AND ANAPHORA

Eva Hajičová

Charles University, Prague, Czech Republic

Abstract: A dynamic approach to discourse structure based on the notion of the hierarchy and changes of activation (salience) of items in the stock of knowledge assumed by the speaker to be shared by him and the hearer, and taking as a starting point the topic-focus articulation of the sentences the discourse is composed of is characterized and illustrated on an analysis of a sample of narrative text.

Keywords: topic-focus articulation, discourse, stock of shared knowledge, salience, anaphora, coreference

1. INTRODUCTION

The Praguian notion of the information structure of the sentence (functional sentence perspective, topic-focus articulation, henceforth TFA) as analyzed in the writings of V. Mathesius, J. Firbas, F. Daneš, and, within an explicit theoretical framework, in Sgall *et al.* (1973; 1986) has found its counterparts in most different present-day theoretical frameworks and is viewed as an important aspect of the analysis of discourse, contributing to discourse cohesion and to the study of the use of anaphoric pronouns.

The dichotomy of topic and focus may be understood as the linguistic patterning basically corresponding to the cognitive opposition between 'given' ('known') and 'new' information, with the caveat that a 'known ' item can be referred to as not 'given', esp. if put into contrast. A weak anaphoric pronoun can only be used as contextually bound, i.e. in the topic (unless embedded within the focus); in the prototypical case, it is coreferential with an item occurring in one of the closely preceding utterances, and it always refers to an entity that in the given time-point is above a certain threshold of salience in the stock of shared knowledge (assumed

ISBN: 0 08 043 438X

by the speaker to belong to the hearer's short-term memory; the original model we refer to in our paper was presented first by Hajičová and Vrbová in 1981, and is further developed in Hajičová, 1987; 1993, Hajičová *et al.*, 1995).

2. THE HIERARCHY OF SALIENCE IN THE STOCK OF SHARED KNOWLEDGE

The stock of shared knowledge (SSK in the sequel) has a dynamic character: the repertoire of elements included in it as well as their degrees of activation (salience) change as the discourse progresses. The difference in these degrees reflects the contention that some elements are relatively easier to access then some others. The decisions on the possibility/impossibility of pronominal reference on one side, and the resolution of anaphora on the other side thus can be based (among other issues, of course) on the difference of the degrees of salience.

An element is considered to be activated to a higher or lower degree depending on various factors. In the papers quoted above, based on an analysis of discourses of a monologue character, the following heuristics based basically on the topic-focus articulation of sentences were introduced:

(i) an element referred to in the focus part of the utterance has the highest degree of activation;

(ii) an element referred to in the topic part of the utterance by a nominal group obtains the degree of activation one degree lower than the highest one;

(iii) an element referred to by a (weak, unstressed) pronoun preserves its activation;

(iv) the activation of elements introduced once into the discourse but not mentioned in the subsequent utterances fades away; the decrease is higher if this element was introduced in the focus and has no longer been mentioned afterwards, and lower if this element was (re)mentioned in the topic part;

(v) elements not explicitly introduced into the discourse but standing in an associative link with an explicitly mentioned element "follow" the latter element in the rise/fall of activation.

The above heuristics can then be formulated in a shape of the following 'rules' (it should be noticed that the degrees of activation impose a partial ordering on elements in the stock of shared knowledge: there can be elements having the same degree of activation):

(1) If $E(x_a)$ is in the focus of S, then $a^n \rightarrow a^0$.

(2) If $P(x_a)$, then $a^n \rightarrow a^n$.

(3) If $NP_d(x_a)$ is in the topic of S, then $a^n \rightarrow a^1$.

(4) If $a^n \rightarrow a^m$, then b^{m+2} obtains for every object b that is not itself referred to in (the underlying representation of) S, but is immediately associated with an item present there.

ISBN: 0 08 043 438X

(5) If x_a neither is included in S, nor refers to an associated object (see Rule (4) above) and was referred to in the topic of the immediately preceding utterance, then $a^n \rightarrow a^{n+1}$.

(6) If x_a neither is included in S, nor refers to an associated object (see Rule (4) above) and was referred to (only) in the focus of the immediately preceding utterance, then $a^n \rightarrow a^{n+2}$.

Notation: x_a denotes an expression x referring to an object a; a^n denotes that this object is salient to the degree n in SSK (the maximum of salience is denoted by n = 0). To the left (right) of the arrow we indicate the state immediately preceding (following) the utterance of a sentence S in which x occurs; $P(x_a)$ denotes that x is expressed by a weak (unstressed) anaphoric pronoun or is deleted in S (albeit present in the underlying representation concerned); $NP_d(x_a)$ denotes that x is a definite NP rather than a weak pronoun; $E(x_a)$ denotes that x_a is any expression other than a weak pronoun.

Appendix 2 brings a sample of a representation of discourse flow in this way (the texts are in Appendix 1): for this illustration we have chosen a slightly shortened part of Paul Wilson's English translation (Dvorak in Love, pp. 251-253) of Josef Škvorecký's Czech book Scherzo capriccioso (Odeon, Praha, 1991, pp. 418-422; first edition in 1983 by Sixty-Eight Publishers, Toronto). The rows represent the state of the activation of the selected items after the sentence the number of which is on the left has been uttered. The numbers of the columns denote the degree of activation; we would like to emphasize that the numerical values are of no substantial importance; what is relevant is the relative degree, with zero being the highest, one being higher than two, etc. For the sake of simplicity, only the degrees of activation of eight items are recorded, namely Kovarik (K), Magda Dvorak (M), the lady L), the black man (A), the buggy or the song (banjo, baritone) (B), torch, waterlily (T), rowboat (R), Antonin Dvorak (A).

3. HIERARCHY OF SALIENCE AND SOME ISSUES OF DISCOURSE STRUCTURE

As we have shown in our previous papers quoted above, such an analysis of discourse makes it possible to throw some light on several issues of discourse structure:

(a) Certain patterning can be readily observed: e.g. a more or less regular change of groupings of items on the "top of the stock" if the discourse fluently passes from one group of items talked about to another group, or a cluster of items staying on the top with other items just entering the stage and leaving it very quickly. Different types of text exhibit different shapes of the flow charts.

(b) The proposed representation of the flow of discourse offers one way of segmenting the discourse more or less distinctly into smaller units according to which items are the most activated ones in these stretches.

(c) An interesting issue for further investigation is that of the identification of possible thresholds (i.e. the placement of vertical lines in the flow charts): thus one can investigate whether and under which conditions such a threshold adds to other prerequisites for the possibility/impossibility of pronominal reference, for the use of a full definite NP in topic, for the necessity to use stronger means for a reintroduction of some already mentioned item, and for the necessity of such a reintroduction to occur in the focus). The presence of

ISBN: 0 08 043 438X

'competitors', of course, is highly relevant for such investigations. For an illustration, see Sect. 4 below.

(d) Last but not least, the proposed representation of the flow of discourse can serve as a basis for the identification of 'topics' of the discourse. It is often disputable to determine 'the' topic of a given discourse; however, the discourse topic(s) occur (or at least the items associated with these topic(s), whatever the notion of association may be understood to stand for) most probably among the items staying longer (or more frequently) among the most activated items, i.e. on the top of the stock.

(e) A comparison of the flow schemes for parallel texts (i.e. an original and its translation) provides an interesting material for the study of the structure of text as influenced by the means of expression available in the languages compared. As we can see, the two schemes in Appendix 1 (one for Czech, and one for English) are almost identical; the small deviations are very local and mostly concern only adjacent sentences. In one place - at the very beginning, sentences 3 and 4 - the difference is given by the fact that while in English the only object to which the (plural) pronoun 'them' may refer are the two figures (from sentence 1) or the two horses (from sentence 3), the singular form of the corresponding Czech pronoun 'jej' may refer cataphorically only to the blue buggy. This difference is evidently due either to the translator's negligence or to the fact that the reference to the buggy might include (by the way of association) also a reference to the horses. The other place we want to comment upon is the segment including the sentences 35 through 39: in the English text, the two youngsters, Magda and Kovarik, are reintroduced already in 35 (by the pronoun 'they'), while in the Czech text, they appear as late as in the sentence 39 (referred to by a possessive pronoun in the noun group 'u jejich břehu', 'on their side of the river'); one of the factors that underly this difference can be seen in the fact that 'banjo' in 35 is newly introduced item that has to appear in the focus. In Czech the rightmost position of the subject meets this condition, which is not that straightforward in English; the translator's choice was to add a pronominal subject (referring to the listeners) to be able to put 'banjo' as the Object into the focus position.

4. ANAPHORIC MEANS

Let us turn now in more detail to the relationships between the choice of coreferring expressions (weak pronoun, noun, noun group with simple or complex adjuncts) and the degrees of salience, as illustrated by our sample text in the Appendix. We refer here to the serial numbers of sentences (more precisely: clauses) in the text (given as the numbers of the lines in the scheme); let us note that in the first sentence of our segment, the pronoun *they* refers to Dvořák's small daughter Magda and to his guest Kovařík, who take a walk in the surroundings of Dvořák's house in Iowa (during his stay in the US). The following observations can be made:

(i) A weak (zero) pronoun refers to a highly salient item (see sent. 11, 34; 9); this pronoun expresses a contextually bound item; the use of a strong (long) pronoun is limited to cases when the reference is made to a contextually non-bound item (in the focus of the sentence) or to a contrastive item in the topic (expressed by an (optional) phrasal stress; in the latest version of our analysis of the topic-focus articulation, this item is marked by a special superscript in the underlying representation of the sentence, see Hajičová *et al.*, in press).

(ii) If two items are close to each other in their degrees of salience, the use of a weak pronoun is limited to (i) cases with relevant grammatical oppositions (gender, number, see 'him' in 4) and (ii) with clear pragmatic basis for inferencing (see e.g. 'they' in 2 and 'them' in 4).

(iii) Otherwise, the coreference to one of the competitors is to be made clear by the use of a noun (see 16), or even of a noun group with simple or complex adjuncts (see 30).

(iv) Such stronger means have to be used also if the salience has faded away (see 19, 24, 26); if the salience goes beyond a certain threshold of comprehensiveness, the item needs to be reintroduced by a reference in the focus part of the sentence.

5. CONCLUSION

Discourse coherence, one of whose aspects is displayed in this way, also is the central concern of the centering theory (Grosz, Joshi and Weinstein 1995). A comparison of the two approaches shows that the notion of backward-looking center corresponds well to that of topic proper, while the occurrence of forward-looking centers raises the degrees of salience of their referents (for a more detailed discussion, see Kruijffová and Hajičová, 1997). The ranking of the forward-looking centers, however, cannot be properly stated only in terms of syntactic relations, and it is not sufficient to determine the degrees on the basis of the single immediately preceding utterance. Thus, two layers should be distinguished: that of language structure (with the backward looking center or the topic), and that of a cognitively based hierarchy (underlying the scale of forward looking centers).

REFERENCES:

Grosz Barbara J., Aravind K.Joshi and Scott Weinstein (1995). Centering: A Framework for modeling the local coherence of discourse. *Computational Linguistics* **21**, 203-225.

Hajičová Eva (1987). Focussing: a meeting point of linguistics and Artificial Intelligence. In: *Artificial Intelligence II - methodology, systems, application* (Ph. Jorrand and V. Sgurev, (Eds.)), 311-322. North Holland, Amsterdam. Hajičová Eva (1993). *Issues of sentence structure and discourse patterns.* Charles University, Prague.

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

Hajičová Eva, Partee Barbara H. and Petr Sgall (in press). **Topic-focus articulation, tripartite structures, and semantic content**. To be published by Kluwer, Amsterdam.

Hajičová Eva and Jarka Vrbová (1981). On the salience of the elements of the stock of shared knowledge. *Folia linguistica* **15**, 291-3O3.

Kruijffová Ivana and Eva Hajičová (1997). Remarks on the notion of 'centers' and on some tenets of the 'centering theory'. *Prague Bulletin of Mathematical Linguistics* **67**, 25-50.

Sgall Petr, Eva Hajičová and Benešová Eva (1973). **Topic-Focus Articulation and Generative Semantics.** Scriptor, Kronberg/Taunus.

Sgall Petr, Eva Hajičová and Jarmila Panevová (1986). **The meaning of the sentence in its semantic and pragmatic aspects**. Academia, Prague and Reidel, Dordrecht.

APPENDIX 1: SAMPLE TEXT

A slightly shortened part (pp. 251-253) of the English translation (Dvorak in Love, Lester and Orpen Dennys Limited, Toronto, 1986, translated by Paul Wilson) of Josef Škvorecký's Czech book Scherzo capriccioso, Odeon, Praha, 1991 (pp. 418-422); first edition in 1983 by Sixty-Eight Publishers, Toronto.

The English text:

1. Across the river they could now see a fire with two figures beside it.

2. When they moved closer,

3. they could make out two white horses against the background of the dark bushes.

4. Then he recognized them.

5. The pale blue buggy.

6. Two hours ago, the beauty from Chicago had sat on the seat

7. while the black man in livery had gone into Kapinos's for beer.

8. They stopped ...

9. and looked across the river.

10. The young lady in the white dress was biting into a chicken leg. ...

11. He looked at Magda.

12. The child's eyes, wide in amazement, stared across the river at this fairytale banquet....

13. He looked at the straw hat.

14. Yes, beside it in the grass a pair of white shoes had been casually tossed

15. and beside them lay a crumpled white pile. ...

16. The beauty stood up

17. and threw the half-eaten leg into the fire.

18. She stretched,

19. said something to the man ...

20. She lifted up her skirts

21. and, stepping gingerly through the grass,

ISBN: 0 08 043 438X

22. she began walking upstream.

23. Her head became a cooly glowing torch.

24. Intoxinated, Kovarik stepped forward

25. and silently followed the beautiful phantom's pilgimage. ...

26. The child padded silently behind him. ...

27. The child whispered.

28. "She's a Rusalka! A water nymph!"

29. He caught his breath.

30. The girl across the river unlaced her bodice

31. and ...she had lifted the skirt over her head,

32. slipped out of it

33. and stood there in nothing but white knee-length knickers ...

34. He couldn't take his eyes off her. ...

35. From downstream they could hear a banjo playing.

36. A pleasant baritone voice sang, "I dream of Jeannie with the bright golden hair ..."

37. The girl let her hands drop ...

38. Cautiously, she stepped into the water.

39. On their side of the river, beyond the low bushes below them, something creaked.

40. Looking towards the sound, he could barely distinguish the outline of a small rowboat

41. and, in it, someone's dark silhuette.

42. The moonlight fell on the head, the white whiskers, the hair in disarray.

43. The Master!

44. He looked quickly across the stream

45. and saw the Rusalka up to her waist in the water. "Borne like a vapour on the summer air; I see her tripping where the bright streams play ..."

46. The Master's head turned in profile towards the velvet baritone.

47. He doesn't see;

ISBN: 0 08 043 438X

48. he only hears,

49. he thought.

50. He himself saw. ...

51. The Rusalka was slowly lowering herself into the water, ...

52. Finally, all that remained on the water was a burning waterlily.

53. Suddenly the child saw too

54. and shrieked,

55. "Papa!"

56. The Master started,

57. looked around

58. and then saw.

The Czech text:

1. Na druhém břehu se objevil oheň a u něho dvě postavy.

2. Když přišli blíž,

3. zabělala se na pozadí temných keřů dvě bílá koňská těla.

4. Potom jej poznal.

5. Bleděmodrý kočárek.

6. Před dvěma hodinami v něm seděla kráska z Chicaga

7. a černoch v livreji ji šel pro pivo ke Kapinosům.

8. Zastavili se,

9. a ..., hleděli na druhý břeh.

10. Mladá dáma v bílých šatech okusovala kuřecí stehýnko. ...

11. Pohlédl na Magdu.

12. Dětské oči, vyvalené úžasem, zíraly k druhému břehu,

12' kde se konala hostina z pohádky. ... (13)

ISBN: 0 08 043 438X

13. Pohlédl na slamák. (14)

14. Ano, vedle něho se v trávě povalovaly překocené bílé střevíčky a zmuchlaná bílá hromádka. ... (15)

16. Kráska vstala

17. a odhodila na půl okousané stehýnko do ohně.

18. Protáhla se,

19. řekla něco černochovi ...

20. Kráska si vyhrnula sukně,

21. a vysoko našlapujíc v trávě,

22. vydala se proti proudu říčky.

23. *Hlava* se *jí proměnila* v *chladně zářící* pochodeň.

24. Omámeně vykročil

25. a jal se tiše sledovat pouť krásného přeludu. ...

26. Dítě ťapkalo mlčky vzadu. ...

27. Dítě zašeptalo.

28. "Vona je Rusálka!"

29. Zatajil se mu dech.

30. Dívka na druhém břehu si rozepjala živůtek

31. a ..., zvedla sukně nad hlavu

32. a soukala se z nich.

33. Za okamžik zůstala jen v bělostných kalhotkách po kolena ...

34. Nemohl od ohýnku odtrhnout oči. ...

35. Zdola, po proudu řeky, zaznělo banjo.

36. Příjemný baryton zpíval: "I dream of Jeannie with the bright golden hair - "

37. Kráska spustila ruce, ...

38. Dívka opatrně vstoupila do vody.

39. Za nízkým křovím u jejich břehu cosi zapraskalo.

40. Pohlédl tam

40' a teprve nyní rozeznal mezi listím obrysy loďky. (41)

41. V ní se právě vztyčoval nějaký člověk (42)

42. a na hlavu mu dopadlo měsíční světlo. (43)

42' Divoký vous, zježené vlasy. (44)

43. Mistr! (45)

44. Rychle pohlédl přes řeku (46)

45. a spatřil Rusálku už po pás ve vodě. "Born like a vapor on the summer air - I see her tripping where the bright streams play - " (47)

46. Mistrova hlava v profilu natočeném směrem k sametovému barytonu. (48)

47. Nevidí, (49)

48. jenom slyší, (50)

49. napadlo ho. (51)

50. Sám viděl. ... (52)

51. Rusálka se pomalu nořila do vody, ... (53)

52. Konečně zůstal na hladině jen hořící leknín. (54)

54. A do toho dítě zaječelo: (55)

55. "Tati! ..." (56)

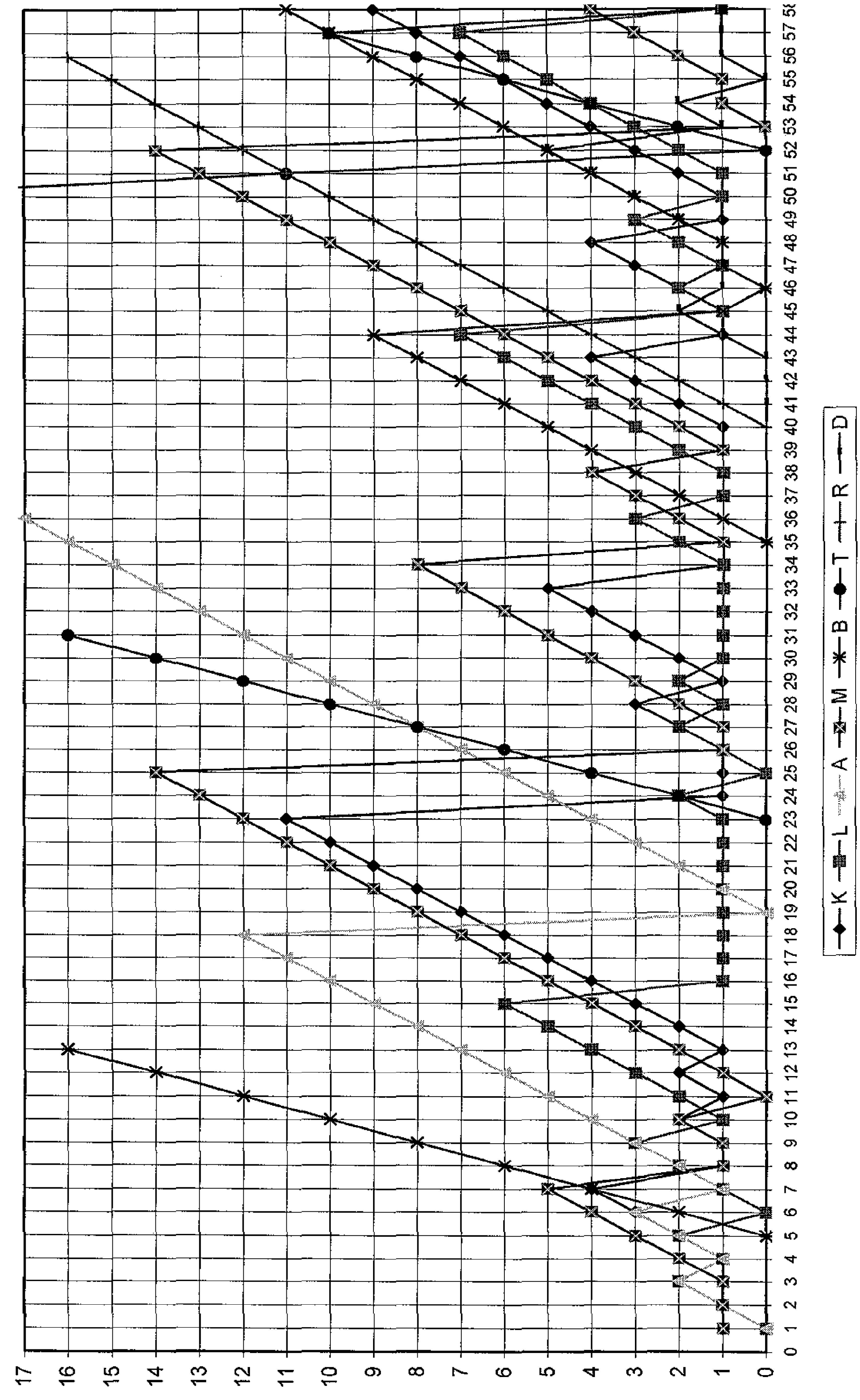
56. Mistr sebou škubl, (57)

57. rozhlédl se, (58)

58. teprve teď spatřil. (59)

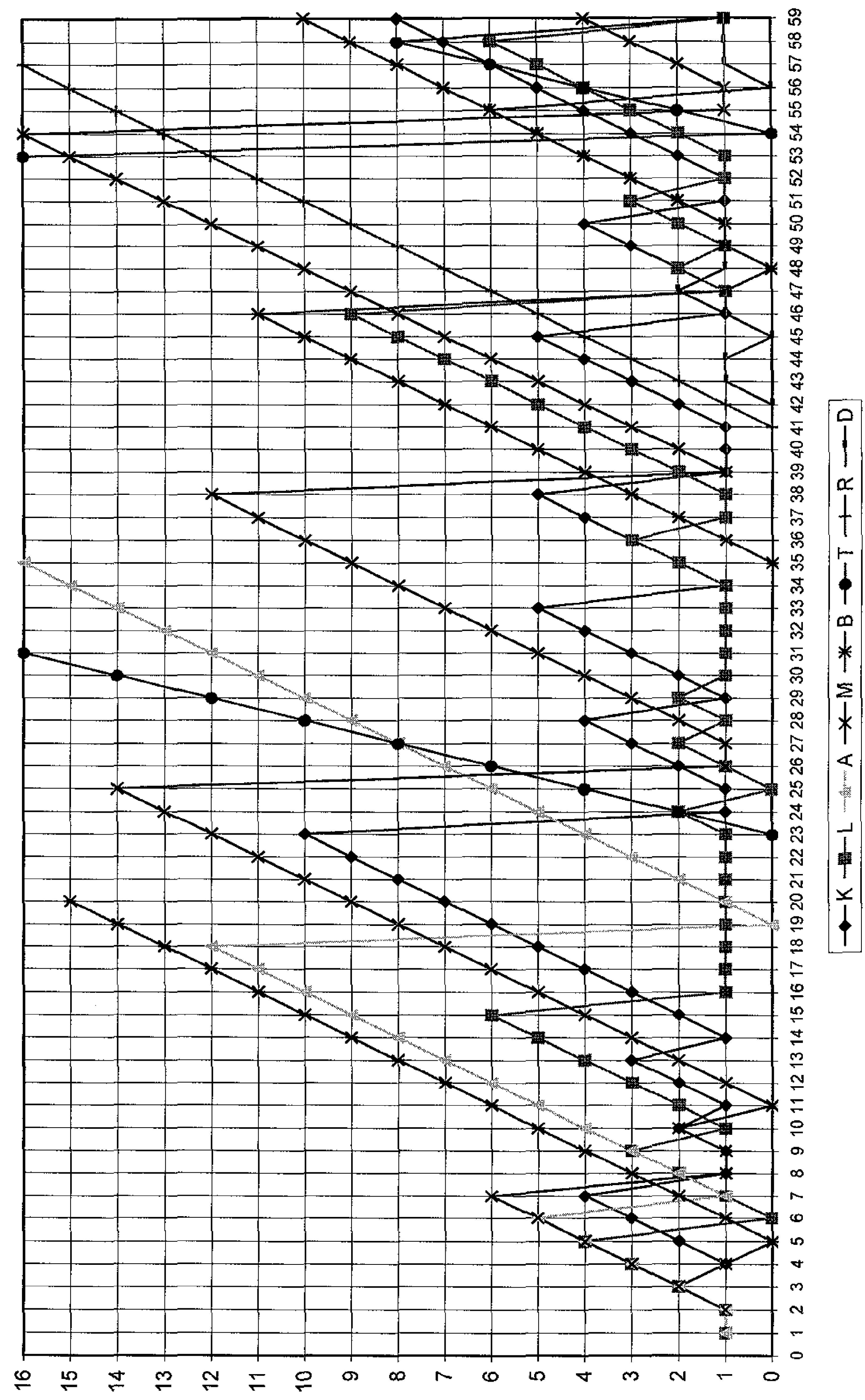
For technical reasons, the numbering of the sentences in the Czech scheme in Appendix 2 had to be modified in order not to include numbers with primes. Therefore, while the numbering on the left-hand side corresponds to the English sentences, the numbers in parentheses correspond to the scheme.

Appendix 2: Scheme of Activation (English)



ISBN: 0 08 043 438X

Appendix 2: Scheme of Activation (Czech)



ISBN: 0 08 043 438X