László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

Tanulmány

László Hunyadi

Grouping, the cognitive basis of recursion in language

Abstract

This paper presents the results of a number of production experiments aimed at identifying the underlying principles of grouping. The detailed analysis of grouping of various degrees of abstractness demonstrates that the same principles underlie visual, abstract prosodic and linguistic grouping, thus questioning the autonomy of language hypothesis in its strict sense.

The initial task was to account for the fact that the rhythmic segmentation of the utterances of identical sentences by a number of speakers is perceived as virtually identical. Three kinds of experiments were carried out to specify such abstract underlying properties: experiments on the grouping of abstract visual elements, experiments on the grouping of abstract prosodic (pitch-related) elements and finally, experiments on grouping in real natural language utterances.

First, subjects were presented with a sequence of abstract elements of the form "•" grouped together into structures with various dependencies and they were asked to assign the perceived groupings a corresponding temporal structure using mouse clicks. It was found that the resulting temporal structures were dependent on the type of grouping the elements represented. Importantly, recursive groupings and only those were assigned a recursive temporal structure. These experiments demonstrated that one has the capacity of assigning a recursive temporal structure to the recursive grouping of abstract elements.

Next, similar patterns were formed using a sequence of capital letters and subjects were asked to pronounce them according to the structure they perceived. Both temporal and tonal grouping were measured. It was confirmed that the grouping of such abstract prosodic elements shows the same temporal organisation as found in the case of the grouping of abstract visual elements. In addition, variation by tone was also found as a means of grouping so that recursive structures and only those were assigned a recursive tonal structure.

Finally, real utterances with and without recursive structure were tested with regard to their temporal and tonal organisation. It was found that both rhythm and tonal structure essentially depend on the principles found in the cases of the grouping of more abstract (visual or prosodic) elements, i.e inherent grouping and recursion. Accordingly, grouping in speech prosody was found to be essentially recursive.

We also considered the issue if prosodic and syntactic recursion apply to the same groups of elements. It was found that the two do not match completely but that they do not contradict either. Prosody matches syntactic grouping in its underlying phrasing, but it also has additional (semantic, pragmatic) functions which are expressed at a derived level of prosody.

The fact that the same principles were identified for various, linguistic and non-linguistic modalities made it possible to assume that recursion observed in syntax is not specific to language, instead, it takes its origin from the more abstract human cognitive faculty of grouping. Such a conclusion of these experimental results, contrary to the 'recursion only' argument of Hauser *et al.* 2002 and Fitch *et al.* in press support the view that the underlying principles of syntactic recursion did not evolve for language alone, instead, they can be traced back to more general cognitive functions, in particular, grouping. The results, we hope, may inspire further studies to identify possible evolutionary traits of grouping in non-humans as well.

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

0 Introduction: recursion and the debate on the faculty of language

We started the current research by wishing to find an account for the observation that the temporal structure (rhythm) of individual utterances of one and the same sentence by different speakers is perceived as being virtually identical. As it was observed in Hunyadi 2002, such utterances, regardless of the individual speech tempo, follow the very same rhythmic pattern so that the recordings of the utterances can be technically combined into a "virtual choir" allowing one to perceive them as happening at the same time. Our starting point was the following set of assumptions: first, however dependent speech rhythm may appear to be on the actual semantic and pragmatic content, speech rhythm is essentially determined by certain more general principles applying to each utterance of a given language. Second, these principles are general enough to underlie the prosody of human language in general. Third, due to the commonly observed relationship between speech and music regarding their rhythmic properties, we assumed that these principles are to be found beyond language (and, eventually, music) in general. Accordingly, what we were aiming to find were those principles which govern rhythm (and, to a lesser extent, pitch variation) without concrete linguistic content but which are also inherited by language (and, eventually, music – thus suggesting essential common grounds for the two without the implication that one can be derived from the other; cf. a similar view by Lehrdal & Jackendoff 1983). We designed a series of experiments to find out how time, the main component of rhythm appears in the representation of non-linguistic structures, structures of a highly abstract nature.

In order to find out if such properties of rhythm are abstract and general enough to apply to less abstract modalities, we tested them on abstract, pitch-related prosodic material without linguistic content. Finally, the results were tested in real speech prosody to see if those abstract properties of grouping also apply to language.

In fact, we were looking for some meaningful principles which might prove to be essential in associating grammatical structure (syntax in particular) with speech prosody and, beyond speech, music as well, whose association with speech is commonly assumed. In this connection, our initial question could easily be broadened in scope to be related to the current debate on the evolutionary properties of language. As it was suggested in Hauser *et al.* 2002, the faculty of language in its narrow sense (FLN) consists of a single important property: recursion. Recursion is thought to be unique to syntax in the sense that it did not evolve from similar properties of some other, more general cognitive systems.[1] Even though subsequent debate has raised important questions regarding the implications of this assumption for the evolutionary status of language (whether it evolved for something other than communication) and, especially, the 'recursion only' hypothesis (cf. Lappin *et al.* 2000, Pinker & Jackendoff 2005), it did not affect the empirical status of recursive structures in syntax. Accordingly, our attempt to identify how abstract temporal grouping and speech rhythm are related and to apply the principles of this relation to syntax as well, implied that, following the recursion of syntax, we should try to see if abstract timing, on the one hand, and speech rhythm (and speech prosody in general), on the other, are also recursive. Since it has been shown that music, with its important relations to speech is recursive (cf. Lerdahl and Jackendoff 1983), it appeared essential to find out if abstract timing and speech prosody are also based on the same property. The series of experiments ranging from quite abstract, non-linguistic ones to

[1] In their subsequent article they emphasise that when saying "recursion is unique to language" they consider it as a strong hypothesis that can be readily refuted by empirical study (cf. Fitch, Hauser & Chomsky 2005).

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

speech utterances were designed to get closer both to the original issue of the cognitive basis of speech rhythm and the broader issue of the relation of syntax to other cognitive systems.

1 Evidence for the temporal properties of abstract grouping: the experiments

The aim of the experiments was to find out the cognitive basis for prosodic segmentation. As it is well known, prosodic segmentation is denoted by several means including intonation, pitch level, accent, stress, pause and rhythm. In most cases these means appear in combination, but this combination may vary and preference may be given to one or another parameter. Accordingly, intonation in the form of specific patterns of the sequence of tones is considered to be the main formal means of denoting the intonation phrase (IP), the major prosodic unit of an utterance (cf. Pierrehumbert 1980, Pierrehumbert and Liberman 1982), but the individual types of IPs are essentially accompanied by appropriate types of accent or stress as well (cf. Pierrehumbert 1980, Ladd 1996). In addition, pauses between IPs can contribute to the identification of boundaries between these prosodic units. These parameters can show a wide range of variation both regarding their actual formal acoustic characteristics and their combination with other parameters depending on semantic, pragmatic and even paralinguistic factors. Due to such diverse reasons for variation it may seem to be rather difficult to unambiguously identify and characterise those parameters which guide us in perceiving and interpreting individual utterances of one and the same sentence as having essentially the same prosodic segmentation. Our task becomes even more difficult if we consider that what we wish the experiments to show is not *how identical descriptively* such utterances are but, taking it as given, *what cognitive principles* this perceived identity is based upon. Thus we need to generalise the problem and formalise our experiment to the extent that properties specific to language are minimised. Accordingly, we should design an experiment where the effects of word semantics, pragmatic and paralinguistic factors are not present. On the positive side of experiment design, however, we should include material which captures the essential property of an utterance of being composed of groups of elements with the elementary relations of co-ordination and subordination (embedding) between them.

Apparently, our experiments, at least at the outset, should be reduced to structures involving the least modality-specific parameters. Speech is generally characterised by a complex of intonation, intensity and duration. Intonation and pitch variation appear to be mostly associated with sounds: it is found in speech and music, but it is too specific to be found in some other modalities, such as dance. Characteristics of speech include accent and stress, the complex variation of pitch and intensity and are associated with intonation: units of intonation are organised around accented syllables. Accent and stress are, in this sense, also specific to speech and music: being associated with the overall temporal organisation of an utterance or phrase, they determine rhythmic structure. Rhythm, in its turn, however, represents the variation of temporal duration (including pause and boundary breaks), a phenomenon general enough to be found in a wide variety of cognitive systems. Accordingly, the study of the representation of structured groupings of abstract, semantically empty elements along the dimension of time can be expected to offer results which can be applied to less abstract, i.e. more specific systems of representation, including language.

Thus, our initial experiments are designed to study the relation between the perception of groupings and their temporal representation. Without an attempt to offer a comprehensive

analysis of the vast amount of valuable previous work on the topics of timing and grouping, we only wish to point out in what significant aspects our approach is different.

As for timing, with the recognition of the fact that cognitive processes cannot be understood without their temporal dynamics, temporal perception has become an important issue in the study of perceptual acts, cognitive processing and volitional motor control (cf. Pöppel and Wittmann 1998). These studies are basically aimed at identifying and describing timing mechanisms as an essential part of the neurocognitive machinery. By studying various temporal data we can learn via what mechanisms information is processed in our mind (such as, among others, sensorimotor behaviour (cf. Mates *et al*, 1994), or intrinsic time constants vs. event identification in the low-frequency and high-frequency mechanisms, cf. Fraisse 1982, 1984.

As for grouping, it was already observed in the *Gestalt* tradition (cf. Wertheimer 1923) that a major factor in the visual and auditory grouping of elements is their relative proximity. There is a continuous scale of the expression of relative distance along which the probability of assigning three or more elements into a single group depends. A similar grouping effect can be achieved by using other single classifying properties, such as similarity in size or shape, colour, material etc.

What we are concerned with in our experiments is not the conditions under which elements can be perceived as grouped together, instead, considering the perception of groupings as given, we want to find out how the various kinds of groupings (co-ordination, subordination) are represented as a variation in time. Since these relations are discrete, for proper perception we need to denote them by discrete representations, such as using parentheses: elements that are included in a pair of parentheses will be considered as a single group; elements outside these parentheses as separate from that single group. We will use this very simple form of denoting groupings because we expect that the use of parentheses will enable subjects to perceive them unambiguously.

As for timing, what we will focus on in our experiments is not the temporal aspects of neurocognitive processes as emphasised in recent research on timing related to speech production (i.e. how, under what temporal constraints these processes, such as event identification of variable length are performed; cf. Friederici *et al* 2004, Kraut *et al*. 2003, Palmer *et al*. 2003, Schiller *et al*. 2003, Pylkkanen *et al*. 2002, Sanders *et al*. 2002), instead, we will use the aspect of time to learn how temporal variation represents the various forms of grouping, probably one of the most basic cognitive relations. As we have mentioned, prosodic grouping is denoted by several means including intonation, pitch level, intensity (accent/stress) and time. Since it is variation by time (duration) that is multi-modal, i.e. less specific to the particular system of representation, what we will learn about grouping by its help can further be used to study the basic properties of grouping in more specific systems of representation, including speech prosody as well.

There were two sets of experiments: those aimed at determining the role of timing alone in abstract grouping and those determining the role of timing as well as pitch variation in a more specific modality, prosody.

The set of *experiments on timing alone* included four experiments. The first experiment was designed to find out how sequences of simple visual elements involving no grouping or just grouping of co-ordination are represented by some temporal structure. The second experiment was aimed at finding the temporal representation of the sequences of the same visual elements at different levels of embedding. The third experiment was aimed at finding out how the calculation of rhythm can play a role in determining the given temporal

representational structure. The fourth experiment included real sentences in a natural language (Hungarian) with the aim to see if the temporal structure of the corresponding prosody can be explained by that of the previously analysed abstract ones.

The set of *experiments on prosodic grouping* included four experiments. The first experiment was aimed at identifying the underlying role of timing and pitch variation in the simple prosodic grouping of abstract, pitch-related prosodic elements with no semantic content. The second experiment was to show how timing and pitch variation are represented in the embedded grouping of such elements. The third experiment was aimed at finding if prosodic grouping is available at an early (pre-school) age. The fourth experiment was conducted to give support for the hypothesis that speech prosody is in fact recursive.

1.1 Experiments on the abstract grouping properties of timing

1.1.1 Experiment 1: grouping of simple visual elements

Material and methods

The participants of these experiments were 50 university students aged 18-22. They were instructed to look at the patterns that appeared on the computer screen and represent the structure they saw by a sequence of mouse clicks. No further instructions were given (especially, they received no hint that would suggest them to use rhythmical variation for the differentiation of the individual patterns. The actual recording of response times was preceded by a practice phase and the subjects could decide when to start the real experiment. Patterns were given in a random order.

Here and in all other experiments the computer program *PsyScope* v. B28 was used on Apple G4 computers to present stimuli and record the response times. In order to find if structural differences produce significantly different temporal representations in terms of segment length the statistical analysis of distribution and MANOVA (number of segment length types x number of pattern types) was carried out using the statistical package JMP 5.1 by SAS Institute.

In this experiment the following patterns were included as visual stimuli[2]:

(1) ••••

(2) (••)(••)

(3) (•••)•

(4) •(•••)

(5) •(••)•

[2] In referring to structure, the term *pattern* will mean the arrangement of a sequence of the elements "•", including the presence or absence of parentheses. The term *group* will denote a set of "•" included in a pair of parentheses, alternatively, a single "•" at the beginning or end of a pattern bordered by a parenthesis on its other side. A s*egment* will refer to the distance between two adjacent "•"'s either inside a pair of parentheses or on both sides of a single parenthesis and measured in ms calculated from the difference of the onset times of the two "•"'s. They will be denoted by the letters *a*, *b*, *c* etc. starting from the left so that segment *a* will be the distance between the first two adjacent "•"'s, segment *b* the distance between the second and the third, etc. Consequently, a pattern of four "•"'s will consist of three segments.

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

Graphically, each of the patterns consisted of four instances of the character "•". Mouse clicks were expected to be produced one for each of these elements. The time of each mouse click was recorded. The interval between two mouse clicks was considered as the length of a segment. As such, the above patterns consist of three segments each, which, following their left-to-right order will be called *a*, *b* and *c*.

With the experiment we wished to support the hypothesis that, although the duration of segments will differ in their absolute values, (a) it is meaningful to classify them as belonging to two categories, i.e. 'long' and 'short' and (b) this categorisation is structure-dependent.

Results and discussion

The corresponding segments of the above patterns had the following distribution of means; cf. *Table 1*:

pattern\ segment	**(1) ••••**	**(2) (••)(••)**	**(3) (•••)•**	**(4) •(•••)**	**(5) •(••)•**
a	435,38	314,51	323,67	932,51	837,41
Std Dev	228,91	148,66	121,30	429,47	445,99
b	719,52	862,02	323,08	302,67	270,71
Std Dev	2015,24	333,46	90,47	113,04	109,64
c	445,12	308,77	937,60	321,96	711,20
Std Dev	273,71	151,28	367	139,93	291,19

Table 1

The graph below demonstrates that the relation between the structure of a given visual pattern and its temporal representation is in fact systematic; cf. *Figure 1*:

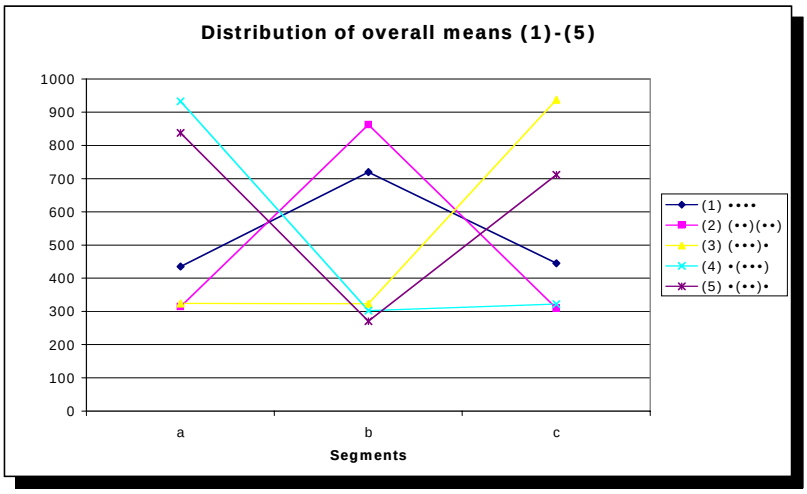


Figure 1

At a glance, these overall means (all subjects taken together) suggest the presence of a systematic relation: within each pattern, segments fall into two categories by length: long and short; and within each category the averages appear to be very similar. Moreover, the structural symmetry of patterns of (3) and (4) as well as (2) and (5) is also shown by the symmetry of the average values.

In particular, by directly comparing (1) and (2) we wanted to find out if the difference between segment length (long and short) was significant.[3] We found segment length to be significant ($F_{(2,46)}$=0.20901, p=0.0127) and pattern type (the presence/absence of parentheses) not ($F_{(1,47)}$=0.0035225, p=0.6859). There was no interaction between length and pattern type ($F_{(2,46)}$=0.0165387), p= 0.6857). It means that the significance of the internal structure of the patterns (whether they consisted of long and/or short segments) was not due to the linear sequence of the segments. Since, in contrast to (2), no grouping was graphically denoted in the visual stimulus of (1), we would expect that the two pattern types are contrasted in such a way that the resulting distinction by segment length is different for the two patterns. Indeed, the within-subject contrast of (1) and (2) confirms this: whereas the respective segments *a* differ significantly and so do the segments *c*, no significant difference was found between the segments *b*:

segment *a* of (1) vs. segment *a* of (2): $F_{(1,47)}$=0.7654035, p<0.0001
segment *b* of (1) vs. segment *b* of (2): $F_{(1,47)}$=0.0039612, p=0.6681
segment *c* of (1) vs. segment *c* of (2): $F_{(1,47)}$=0.4211608, p<0.0001

It means the following: as the overall means show, both patterns consist of segments of different length: segments *a* and *c* are short and the middle segment *b* is long. Although the patterns differ in the short segments, their middle segments are similarly long. The inside-subject comparison of segments in each of the patterns gives further detail of this relations: the difference between short and long segments (between *a* and *b* as well as *b* and *c*) is significant inside (2) (as expected by the denotation of grouping in the visual stimulus) and (not unexpectedly) not significant inside (1); cf.:

(1) segment *a* vs. segment *b*: $F_{(1,47)}$=0.022006, p=0.3144
segment *a* vs. segment *c*: $F_{(1,47)}$=0.0088071, p=0.5231
segment *b* vs. segment *c*: $F_{(1,47)}$=0.020094, p=0.3361

(2) segment *a* vs. segment *b*: $F_{(1,47)}$=4.0156175, p<0.0001
segment *a* vs. segment *c*: $F_{(1,47)}$=0.0116425, p=0.4631
segment *b* vs. segment *c*: $F_{(1,47)}$=3.4668718, p<0.0001

The fact that the difference in segment length is significant between the two patterns in their within-subject comparison, allows us to conclude that even within (1) there is a *relative* distinction of long and short segments.

[3] The applied analysis of MANOVA is capable of deciding if the difference between the means of two sets of values is significant but it cannot further specify to what degree they are different. Since, due to the linearity of the patterns of stimuli used, the onset times of responses were ever increasing, the analysis of absolute onset times could not have yielded relevant results. In contrast, the distance between two adjacent "•"'s increased or decreased regardless of their linear sequence, which made the study of the means of segments relevant to us.

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

The comparative graph of the overall least square means clearly demonstrates that the difference between means inside (1) is much less than between those inside (2); cf. *Figures 2* and *3,* respectively:

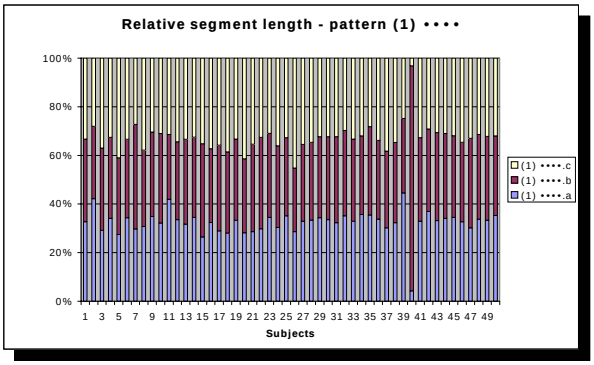


Figure 2

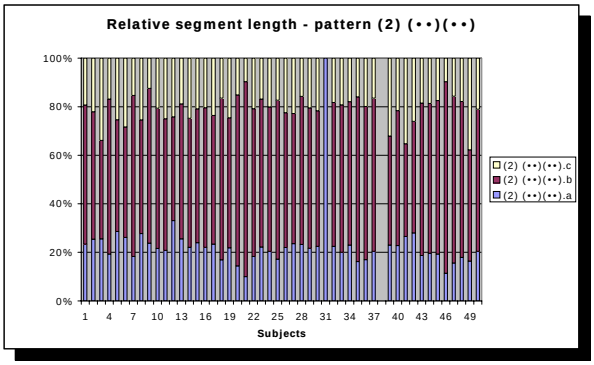


Figure 3

Looking at these two graphs we can notice the individual differences between the responses. We can also notice that these differences nevertheless fall within a clear range inside each pattern suggesting that there was a temporal grouping of long and short segments in both patterns. Whereas this is what we had expected for (2) with grouping marked by parentheses in the visual stimulus, we need to explain the relative distinction of long and short segments in (1) without any marking of grouping in the visual stimulus. We suggest that grouping is an elementary cognitive operation present in any, even unstructured sequence of elements. In other words, we tend to divide a sequence of four elements into two halves even if such division is not indicated. But this division is less prominent than if it is explicitely indicated in the stimulus: in (2) a much longer boundary break is produced between the two halves (between the second and the third element "•". The phenomenon observed in (1) suggests that grouping, the effect of rhythm in the broad sense is presumably inherent to the perception of any kind of structure[4]. Accordingly, the distinction of segments by length in (1) is an effect of this inherent rhythmic grouping only, whereas in (2) − of the temporal interpretation of the perceived grouping.

Next, we compared (3) and (4) to find out if the relative order of segments of similar length in a pattern is statistically significant for the distinction between long and short segments.

(3) (•••)•
(4) •(•••)

In order to make the proper calculation, averages of segments for each pattern were put in a decreasing order. We got the following distribution of overall means across responses; cf. *Figure 4*:

[4] This experimental observation gives support to the Anti-Lapse Provision of Selkirk's Principle of Rhythmic Alternation (Selkirk 1984) and also confirms Hayes' Disyllabic Rule (Hayes 1984).

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

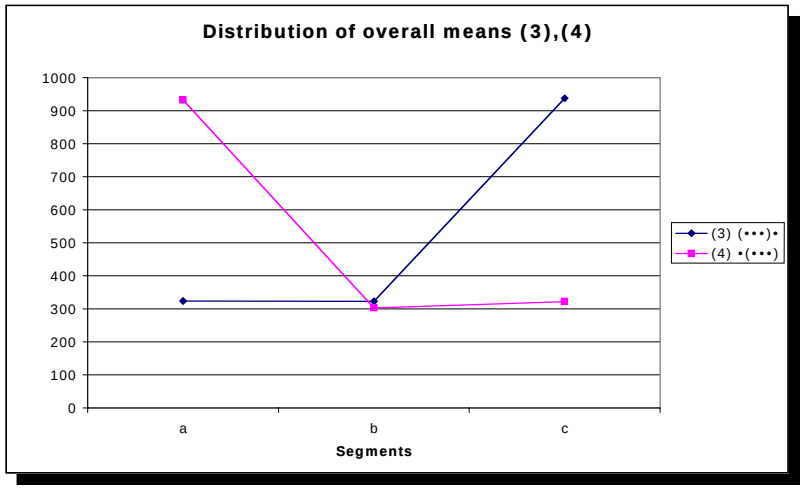


Figure 4

According to the data, segment length was significant ($F_{(2,46)}$=3.6576249, p<0.0001), whereas difference in pattern type (the linear sequence of long and short segments) was not ($F_{(1,47)}$=0.0056584, p=0.6085) with no interaction between the two ($F_{(2,46)}$=0.033688, p=0.4667).

Thus we could conclude that segment length is not influenced by the linear sequence of segments. Instead, it must be determined by some deeper, more abstract structural relation between the segments. In order to find out what the nature of this relation is, we carried out yet another comparison.

The nature of the effect of structure can be seen from the comparison of the corresponding short and long segments of (2) and (4); cf. *Figure 5*:

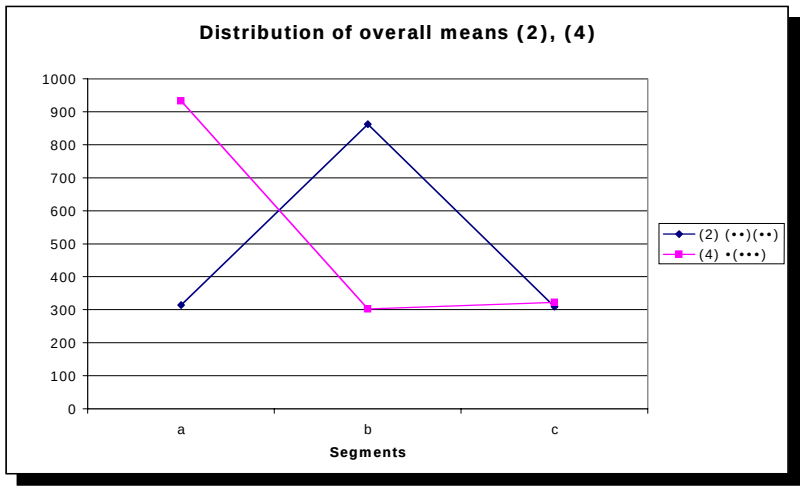


Figure 5

Segment length is statistically significant ($F_{(2,46)}$=3.3580993, p<0.0001) and pattern type is not ($F_{(1,47)}$=0.0865333, p=0.0495). There was interaction between segment length and pattern type ($F_{(2,46)}$=0.1130973, p<0.0851).

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

Although, as the graph above shows, the respective long segments (*b* of (2) and *a* of (4)) are of similar length, the short segment *b* of (4) is different from the rest of the short segments. What is suggested in this comparison is that it does indeed matter how many elements there are within one segment, so length is structure-dependent. Since both the respective long and short segments had similar timing values in the two patterns, what really determined the contrast of the two patterns was the number of elements within a group (within a pair of parentheses).

Actually, segment *a* of (2) only consists of these two elements and makes up a group of its own, whereas the two elements constituting segment *b* of (4) are joined by yet another element to make a three-element group. That is, contrast is not determined by length alone, it is essentially a structural difference. As the experiment showed, subjects perceived this structural contrast between the two patterns and assigned them significantly contrasting temporal representation accordingly.

As we could see, it is the internal structure of a pattern (the number of elements in a group) rather than the linear ordering of segments of different length that determines temporal representation. Consider now patterns consisting of segments of similar length but arranged in different pattern types and see if the inverse case is also true: the same kind of group (consisting of the same number of elements) has the same temporal representation even though the sequence of groups differs.

Considering (2) and (5), short and long segments are not only arranged in a different linear order but the long segments are found in a different structural arrangement, too. It is shown in their contrast: both segment length and pattern type are significant ($F_{(2,46)}$=3.0076634, p<0.0001 and $F_{(1,47)}$=0.5099821, p<0.0001, respectively), and there is also an interaction between the two effects ($F_{(1,47)}$=0.005073, p<0.0001).

(2) and (5) coincide in that both of them are built of short and long segments. They essentially differ in the order of the given short and long segments. Let us compare the first two segments only; cf. *Figure 6*:

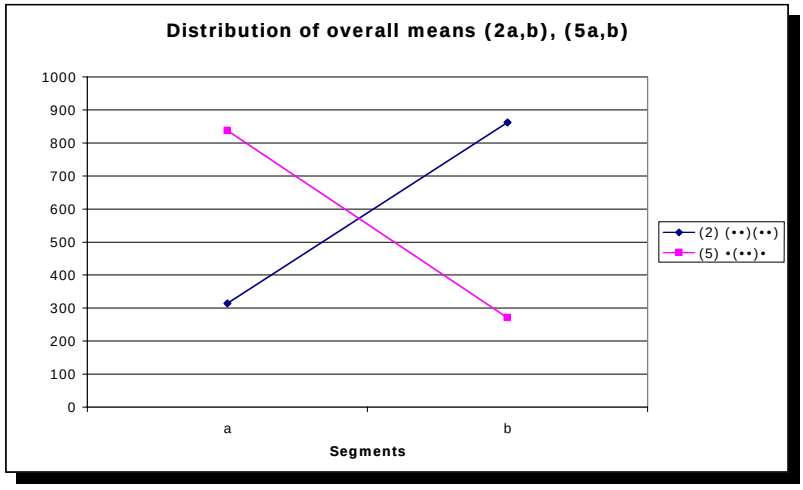


Figure 6

The figure shows that the individual segments clearly differ by length ($F_{(1,47)}$=3.2605702, p<0.0001), whereas the difference in pattern type has no effect ($F_{(1,47)}$=0.0400028, p<0.1768). Hence the symmetrical relation we notice between them. Accordingly, we can conclude that it

is indeed the internal composition of groups (the number of elements within a pair of parentheses) rather than the sequence of segments within the given pattern that matters in determining the actual temporal representation of the given segments of the visual stimulus.

1.1.2 Experiment 2: grouping of visual elements with embedding

Material and methods

The next series of experiments was conducted to find out the temporal representation of embedded structures, especially similar sequences (long segment followed by a short one) at different levels of a tree-like hierarchy. The participants of this experiment were the same as in the previous experiment described in section *1.1.1* and the conditions were also the same. We compared the following types (sequences of comparison highlighted):

(2) (••)(••)
(5) •(••)•
(6) •(•(••))•
(7) •(•(••))••

Results and discussion

The segments of these patterns had the following distribution of means; cf. *Table 2*:

pattern\ segment	**(2) (••)(••)**	**(5) •(••)•**	**(6) •(•(••))•**	**(7) •(•(••))••**
a		837.41		
Std Dev		445,99		
b	862.02	270.71	806,16	828,62
Std Dev	333,46	109,64	341,51	366,3
c	308.77		282,12	290,6
Std Dev	151,28		125,96	128,89

Table 2

Figure 7 demonstrates how similar the differences between short and long segments across pattern types are:

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

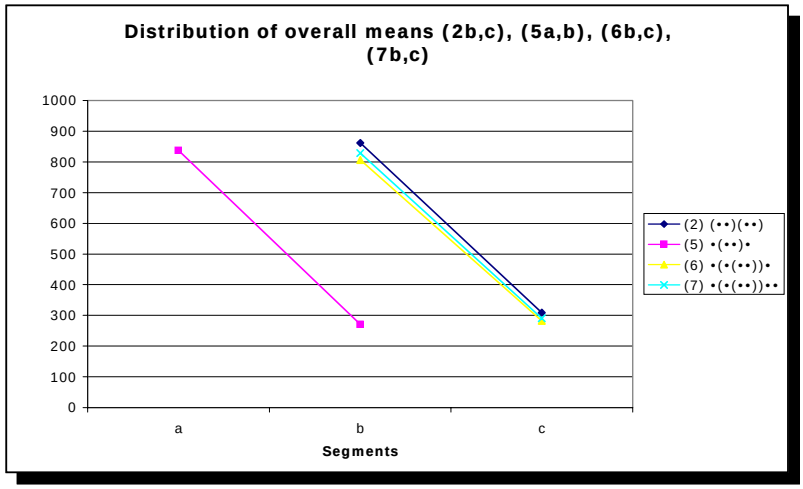


Figure 7

It was found that the difference of the respective segments •(••) by length was significant ($F_{(1,47)}$=3,6331922, p<0.0001). Also, important for this experiment, pattern type differences were not significant ($F_{(3,45)}$=0.098039, p=0.2352). No interaction was found between the two effects ($F_{(3,45)}$=0.0125188, p=0.9042). These data suggest that, again, it is the internal composition of the comparable group (the number of elements in a group) rather than the sequence of these groups that determines temporal representation. Accordingly, the sequence long + short in the form of stimulus •(••) has the very same long vs. short distinctions in timing regardless of its position in the overall structure. Namely, whether this structured sequence is part of a co-ordination as in (2) and (5), or part of a subordination (embedding) as in (6) and (7), the distinction between long and short remains significant. This result shows an essential property of grouping: it follows the principle of recursion: no matter what larger structure a given pattern is part of, it possesses the very same properties, in our case temporal ones.

The embedding of the pattern •(••) in (6) and (7) clearly shows a case of recursion: its temporal properties (distinction of long and short sequences and their relative duration) are preserved across pattern types from (2) through (7). (6) and (7), however, reveal a special property of embedding. Namely, although for a segment to be short or long depends on the number of elements in a given group (enclosed in parentheses); but how short or long the given segment is depends on the level of embedding in the hierarchy. Namely, the comparison of the first two segments, *a* and *b* of (6) and (7) confirms that segment length is significant ($F_{(1,49)}$=0.4294428, p<0.0001) and pattern type is not ($F_{(1,49)}$=0.0490959, p=0.1273) with no interaction between the two effects ($F_{(1,49)}$=0.0006232, p=0.862). Accordingly, although segment *a* and segment *b* of (6) and (7) (both being long segments, making a group of their own) share in common that they include a boundary of embedding (denoted by the first left parenthesis), the actual length of these adjacent segments significantly differs in both patterns. Their relative length is determined by the depth of embedding: as *Table 3* shows, segment *b* embedded into segment *a* becomes shorter than segment *a*:

pattern\ segment	**(6) •(•(••))•**	**(7) •(•(••))••**
a •(•(••))•	1003.14	1032,34
Std Dev	535,59	495,83
b •(•(••))•	806,16	828,62
Std Dev	341,51	366,3
d •(•(••))•	1056,22	1146,22
Std Dev	556,47	545,43

Table 3

Table 3 allows us to make yet another important observation: in addition to segment *b* being embedded into segment *a*, segment *d* (following the two right parentheses) represents a return to the original level of segment *a*, a kind of "de-embedding". We can notice that the segment introducing embedding is shorter than the one introducing de-embedding. The embedding of segment *b* into segment *a* starting from the second structural element "•" is denoted by its length being shorter than that of the segment it is embedded into (segment *a*); however, segment *d*, the segment representing de-embedding or return to the uppermost structural level (above the second structural element "•") is longer than segment *b*.

As we will show in further experiments, this observation regarding the difference between the timing of embedding and that of de-embedding will prove to be significant in determining the temporal properties of embedding in the grouping of other types of elements as well.

1.1.3 Experiment 3: the role of rhythm in temporal representation

Material and methods

The third kind of experiments was aimed at finding the rhythmic properties of segmentation. The participants and the conditions of the experiment coincided with the previous two experiments. The following types were compared (the segments to be compared are highlighted):

(8) •(•(••))
(9) •(•(•••))
(10) •(•(••••))

Results and discussion

The corresponding overall means and standard deviation are shown in *Table 4* and the difference between short and long segments in *Figure 8*:

pattern\ segment	**(8) •(•(••))**	**(9) •(•(•••))**	**(10) •(•(••••))**
a	1012,58	1071,06	1037,83
Std Dev	445,87	546,86	467,07
b	761,7	813,98	860,46
Std Dev	261,65	283,07	327,42
c	257,7	283,9	302,31
Std Dev	70,16	71,06	79,91

Table 4

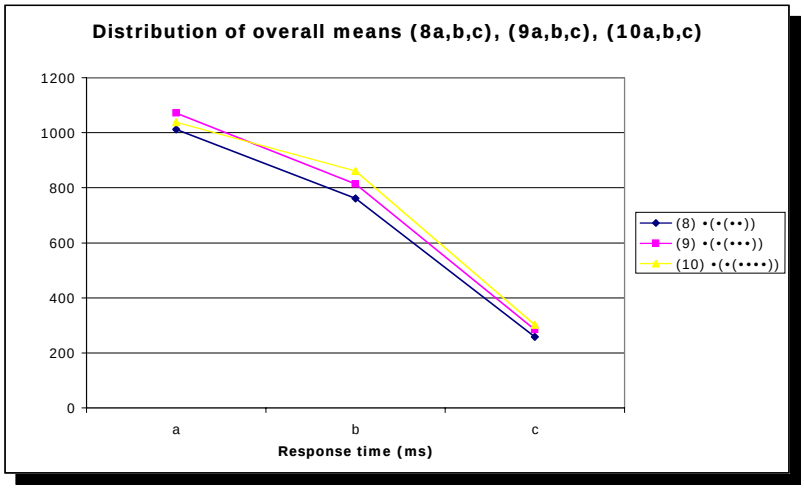


Figure 8

It was found that both segment length and pattern type had a significant effect ($F_{(2,45)}$=4.8150533, p<0.0001 and $F_{(2,45)}$=0.2939963, p=0.003, respectively). No interaction was found between the two effects ($F_{(4,43)}$=0.0628774, p=0.6123). In order to find out if the number of embedded elements had an effect, we made a more detailed comparison of the segments. Comparing segments *a* and *b* alone, we found that both segment length and segment type were significant ($F_{(1,46)}$=0.5825315, p<0.0001 and $F_{(2,45)}$=0.176827, p=0.0256, respectively) with no interaction between the two effects ($F_{(2,45)}$=0.0459871, p=0.3636). Then, the comparison of segments *b* and *c* alone similarly showed both segment length and pattern type to be significant effects ($F_{(1,46)}$=4.8117753, p<0.0001 and $F_{(2,45)}$=0.4274205, p=0.0003, respectively) with no interaction between them ($F_{(2,45)}$=0.052716, p<0.3148). These results allow us to conclude that the significance of segment length can be attributed to the level of embedding, whereas that of pattern type to the difference in the number of embedded elements, both being structurally determined differences.

Having a closer look at the means we found that the two adjacent groups in each pattern were represented by the subjects as having the same duration; accordingly, if the respective

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

groups did not contain the same number of elements, there was a tendency to substitute the missing elements in the shorter group. These missing elements were substituted by pauses so that the duration of the group on the left would match that of the group on the right. In addition, we also expected to have an additional pause between these two groups in order to mark the separation of the two distinct groups.

What we found in (8) confirmed these expectations: the pause between segment *b* and segment *c* was the duration of an actual element "•". However, (9) and (10) did not quite show this rhythmical regularity. Namely, in (9) segment *b* was indeed extended by pauses (each of the length of a "•") so that its duration could match that of segment *c*, but the additional pause between the two segments was somewhat shorter than the duration of a single "missing" element ("•"). At last, in (10) with 4 elements within segment *c*, segment *b* was not substituted with enough pauses to match the total duration of segment *c*.

We offer the following explanation for this observation: (8) was not really considered as a structure with embedding: the two segments *b* and *c* were perceived as structure (2) earlier, i.e. a pair of two co-ordinated segments of equal length. This is confirmed by the duration of the pause resulting in a fully symmetrical structure.

(9) and (10), on the other hand, were not perceived as sequences with symmetrical structure but as representing embeddings. We suggest that embedding is represented by the shortening of the additional pause (the length of a "•") between two adjacent segments: the fact that the pause in (9) is relatively longer than the one in (10) can be explained by the fact that shortening affects every single instance of pause as substitution of missing elements in a segment. Accordingly, since two pauses are to be inserted in segment *b* of (9) and three in (10), the shortening of each pause results in segment *b* of (10) being shorter than that of (9).

We can generalise the crucial difference in temporal structure between co-ordination and subordination to be as follows: in case of co-ordination we have a symmetrical structure with equal length of two neighbouring groups with a pause of the duration of a single element ("•"), whereas in the case of subordination embedding is marked by the shortening of every single instance of a pause. Thus embedding has the timing effect that it is marked by the shortening of boundary breaks between groups.

Considering the length of segments at the boundary of embeddings at different levels in (11), (12) and (13), we can also notice that embedding is expressed by the shortening of such boundary breaks (instances of the segment for comparison "•(• •" are highlighted):

(11) •(• •(• •))
(12) •(• • • (• • •))
(13) •(• • • • (• • • •))

The overall means and standard deviation of these patterns are shown in *Table 5*:

pattern\ segment	**(11) •(• •(• •))**	**(12) •(• • • (• • •))**	**(13) •(• • • • (• • • •))**
a	932,04	951,98	980,56
Std Dev	364,59	429,75	449,39
b	286,56	289,31	310,96
Std Dev	79,2	76,47	90,58
c	779,02		
Std Dev	285,43		
d	274,78	865,27	
Std Dev	129,16	316,75	
e		278,8	876,92
Std Dev		88,43	406,84
f			311,96
Std Dev			104,55

Table 5

The systematic difference between short and long segments is shown in *Figure 9*:

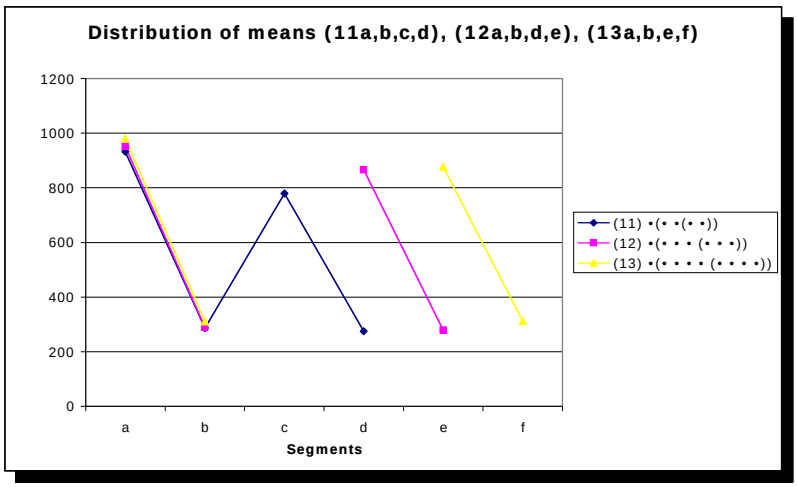


Figure 9

Again, we can conclude that segment length is a significant effect ($F_{(3,46)}$=4.2242506, p<0.0001) whereas pattern type is not ($F_{(2,47)}$=0.2156844, p=0.0102). There was no interaction between the two effects ($F_{(6,43)}$=0.1624642, p=0.343).

Let us note that patterns (8), (9) and (10) included a sequence of embedding varying in the number of elements in the rightmost, deepest embedded segment only. Note that the patterns (11), (12), (13) above also differ in the number of elements in this deepest embedded segment. Let us also recall that the comparison of the corresponding segments *b* and *c* alone in (8), (9) and (10) showed both segment length and pattern type to be significant effects. These results allowed us to conclude that the significance of pattern type could be attributed to the number of embedded elements, a structural difference between the three patterns. This generalisation is confirmed by a similar comparison of the deepest embedded groups of (11), (12) and (13) as well. The overall comparison of the means of segments (11c, d), (12d, e) and (13e, f) confirm that both segment length and pattern type are significant effects ($F_{(1,48)}$=3.5723377, p<0.0001 and $F_{(2,47)}$=0.1768302, p=0.0218, respectively) with no interaction between them ($F_{(2,47)}$=0.01354274, p=0.0506).

Looking at *Table 5* we can also notice that the means of the segment at the first boundary of embedding is systematically greater than the means of the segment at the second boundary of embedding in each type. These data confirm our earlier observation that the shortening of segments bordering an embedding is directly related to the increase of the depth of the embedding.

To sum up our findings as manifested in the experiment:

a. grouping is inherent to the perception of a set of elements;
b. temporal grouping is recursive;
c. segment length in pattern types is significant: the difference of segments according to length reflects a difference in pattern types;
d. linear order alone does not have an effect on segment length: there is no significant difference between types containing the same segments in different linear order;
e. segment length across group boundaries is dependent on the hierarchical structure of groups rather than their linear order;
f. non-adjacent elements can have a structural relation based on rhythm;
g. adjacent groups tend to have a symmetrical relation based on equal length: this is the basis for the rhythmical beat structure of timing.

1.1.4 Experiment 4: the temporal structure of speech prosody

Material and methods

Next, we conducted experiments with natural language sentences to find out if linguistic rhythm follows the observations made on the basis of abstract grouping. In particular, we wished to find out if recursive syntactic embedding has its temporal (rhythmic) equivalence based on grouping. The data were compared to a case of non-recursive syntactic insertion. Two kinds of sentences were involved in the experiment, (14a) and (14b) representing recursion and (15) being syntactically non-recursive. The participants were university students, 25 for pattern (14) and 36 for pattern (15), all aged 18-22.

Since in this experiment we were only interested in rhythm and not the accompanying intonation, in order to enhance the accuracy of measurement we did not use audio-recordings, instead, subjects were presented with written sentences which they were asked to read to themselves and, following the rhythm suggested by capitalised syllables apply mouse clicks at

those syllables, using the same technique and methodology as in the earlier experiments. In order to avoid mechanical, equal beat measures, they were also asked to read in an expressive manner.

The advantage of this methodology was that the markers (mouse clicks) were placed by the subjects themselves at the accented syllables exactly following their inner speech thus enhancing accuracy.

First, consider sentences with recursive embedding:

(14a) A macska, amit a kutya, ami megveszett, megharapott, elszaladt.
The $cat_{nom, i}$ $that_{acc, i}$ the $dog_{nom, j}$ $that_{nom, j}$ was rabid $bit_{(j)}$, $ran_{(i)}$ away.
'The cat, that was bit by the dog that was rabid, ran away.'

In order to facilitate the proper interpretation of the rhythmic representation of the sentence by the subjects, the sentence was visually presented in the following form:

A MAcska
amit a KUtya
ami MEGveszett
MEGharapott
ELszaladt.

The segmentation of this sentence was as follows[5]:

Segment *0*: (a)
Segment *a*: macska amit a
Segment *b*: kutya ami
Segment *c*: megveszett
Segment *d*: megharapott
Segment 00: (elszaladt)
According to the segmentation, segments *a* and *b* include an embedding, segment *d* a de-embedding.

(14b) A macska, amit a kutya, ami megveszett,
The $cat_{nom, i}$ $that_{acc, i}$ the $dog_{nom, j}$ $that_{nom, j}$ got rabid

amikor egy róka megharapta, megharapott, elszaladt.
when a $fox_{nom, k}$ $bit_{(k,j)}$, $bit_{(j)}$, $ran_{(i)}$ away.
'The cat, that was bit by the dog that got rabid when a fox bit it, ran away.'

In order to facilitate interpretation, this sentence was presented to the subjects in this form:

[5] The distance starting with an accented syllable (capitalised in the stimulus) until but not including the next accented syllable was considered a segment. Both in (14a) and (14b) the syllable including and preceding the first accented syllable, denoted as *Segment 0* and those including and following the last accented syllable and denoted as Segment 00 were, by the above definition, not considered.

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

A MAcska
amit a KUtya
ami MEGveszett
amikor egy RÓka MEGharapta
MEGharapott
ELszaladt.

This sentence had the following segmentation:

Segment *0*: (a)
Segment *a*: macska amit a
Segment *b*: kutya ami
Segment *c*: megveszett amikor egy
Segment *d*: róka
Segment *e*: megharapta
Segment *f*: megharapott
Segment *00*: (elszaladt)
According to the segmentation, segments *a*, *b* and *c* include an embedding, segment *f* a de-embedding.

Results and discussion

The overall means of duration of each segment is shown in *Table 6* and graphically represented in *Figure 10* below:

pattern\ segment	(14a) Macska1	(14b) Macska2
a	1339,08	1071,8
Std Dev	1427,76	552,61
b	955,24	1022,32
Std Dev	387,29	543,17
c	1005,6	1280,92
Std Dev	408,72	596,47
d	1124,92	861
Std Dev	465,41	571,81
e		1120,2
Std Dev		462,1
f		971,64
Std Dev		364,07

Table 6

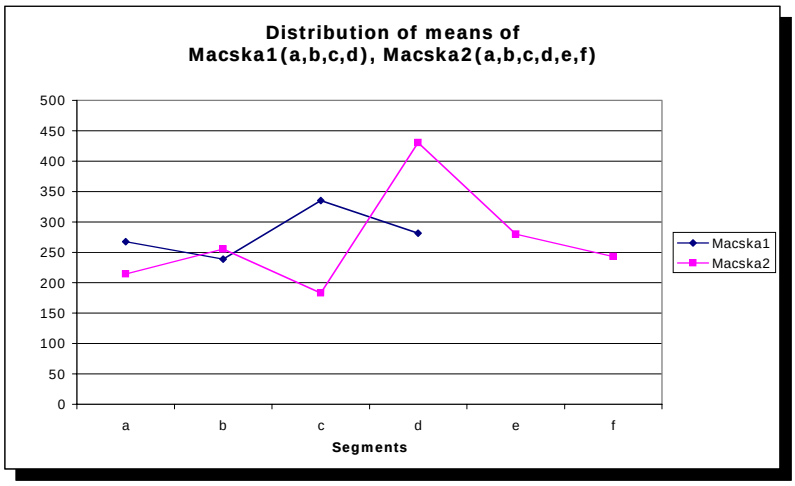


Figure 10

What we were essentially interested in was whether the temporal relation between embedding and de-embedding found in abstract grouping can be confirmed in concrete speech utterances as well; i.e., whether de-embedding can be found as longer (slower) than embedding. We wanted to measure it in terms of average syllable duration, based on segment length

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

(measured as the distance between two mouse clicks). In order to do so, segment length was divided by the number of syllables in each segment.

(14a) was found to have the following syllable duration for its segments:

(14a)
Segment *a*: 267.8 ms/syllable (5 syllables)
Segment *b*: 238.81 ms/syllable (4 syllables)
Segment *c*: 335.2 ms/syllable (3 syllables)
Segment *d*: 281.23 ms/syllable (4 syllables)

Everything between "macska" 'cat' and "elszaladt" 'ran away' is an embedded material, starting with an embedding contained in segment *a* and ending in a de-embedding contained in segment *d*. The above average values of syllable duration confirm that, for the same structural level, embedding is faster than de-embedding ($a < d$). Similarly, there is an embedded material between "kutya" 'dog' and "megharapott" 'bit', the embedding contained in segment *b* and the de-embedding in segment *c*. Since $b < c$, it shows again that embedding is faster than de-embedding.

The same systematic relation is confirmed by a similar analysis of (14b); cf.:

(14b)
Segment *a*: 214.36 ms/syllable (5 syllables)
Segment *b*: 255.58 ms/syllable (4 syllables)
Segment *c*: 182.99 ms/syllable (7 syllables)
Segment *d*: 430.5 ms/syllable (2 syllables)
Segment *e*: 280.05 ms/syllable (4 syllables)
Segment *f*: 242.91 ms/syllables (4 syllables)

Accordingly, everything between "macska" 'cat' and "elszaladt" 'ran away' is an embedded material, starting with an embedding contained in segment *a* and ending in a de-embedding contained in segment *f*. The above average values of syllable duration confirm that, for the same structural level, embedding is faster than de-embedding ($a < f$). Similarly, everything between "kutya" 'dog' and "megharapta" 'bit' is yet another embedded material, starting with an embedding contained in segment *b* and ending in a de-embedding contained in *e*. Since $b < e$, it confirms again that embedding is faster than de-embedding.

All this confirms that there is a systematic relation between embedding/de-embedding and their temporal expression: for the same structural level, i.e. returning to the same level from a lower level, embedding is faster than de-embedding It is important to note that here again it is not the immediate linear adjacency of segments that plays a role. Successive cases of embedding may not show a regular shortening of delimiting boundary brakes in absolute terms: structural complexity is more decisive than simple linearity. At the same time, what we can see here can be called the "bookmark-effect". Accordingly, when performing an embedding one leaves a bookmark behind with a record of the duration expressing the given embedding and when returning to the same level one recalls this value and makes the duration values (especially the duration of the delimiting pause of de-embedding) longer than that.

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

This requirement appears to be consistent with the observation we made in the previous section on the basis of the study of the temporal structure of abstract groupings. As we found in the analysis of grouping patterns (11), (12) and (13), embedding is expressed by the shortening of boundary breaks in general and, due to the fact that abstract grouping is less specific by nature than prosodic grouping, we may suggest that hierarchical prosodic grouping is based on the same timing requirement observed in the grouping of abstract, non-linguistic elements. Finally, similarly to the hierarchical grouping observed for abstract, non-linguistic elements, prosodic grouping is also recursive: at any level of the hierarchical tree, embedding is represented as shorter than de-embedding.

The second kind of sentences, this time with non-recursive syntactic insertion is represented by (15); syllables were capitalised to denote segment boundaries where response times of mouse clicks were supposed to be taken:

(15)
MEG tudnád mondani, hogy – az ÉN órám MEGállt – HÁNY óra van?
Preverb you could tell that the I my watch stopped how many hour is
'Could you tell me – my watch has stopped – what time it is?'

Segment *a*: megtudnád mondani hogy az
Segment *b*: én órám
Segment *c*: megállt
Segment *00*: hány óra van

(15) has the following values of syllable duration for its segments:

Segment *a*: 185.68 ms/syllable (8 syllables)
Segment *b*: 232.32 ms/syllable (3 syllables)
Segment *c*: 416.39 ms/syllable (2 syllables)

The two sides of insertion are represented by segments *a* and *c*. Clearly, the right side of the insertion is denoted by a longer pause (syllable duration is longer, cf. $c > a$), characteristic of embedding. Since (15) may, in a way, be similar to (5) •(••)• (the middle segment (••) of (5) cannot be an embedding either), we could expect that, similarly to (5) the first and the third segments are of nearly equal length in (15), too. To account for this longer pause found in (15) we will return in section *1.2.6*.

The overall means of duration of each segment is shown in *Table 7* and graphically represented in *Figure 11* below:

Segments	(15)
a	1485, 42
Std Dev	504,53
b	696,97
Std Dev	298,95
c	832,78
Std Dev	295,44

Table 7

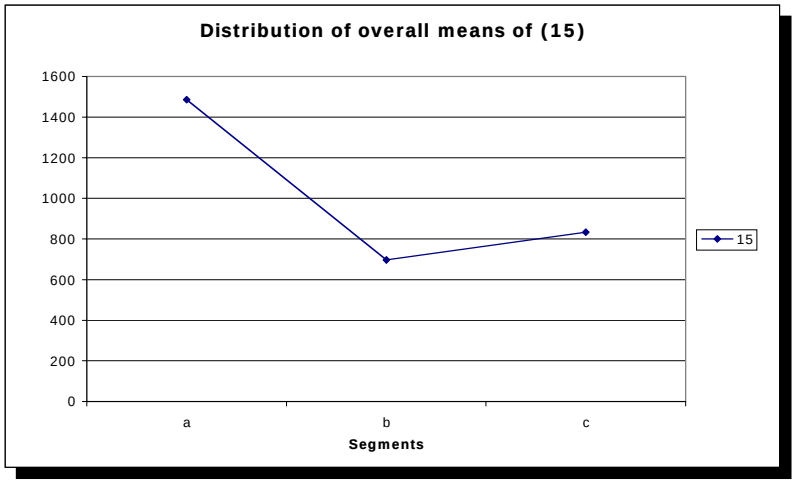


Figure 11

1.2 Experiments on the abstract grouping properties of prosody

1.2.1 Experiment 5: timing vs. simple grouping of abstract, pitch-related objects

As we saw in the previous section, temporal grouping of abstract visual objects is general enough to be applied to the temporal grouping of linguistic objects like words as well. In order to show this, we presented the rhythmic structure of sentences containing embeddings. Although rhythm is one of the aspects of prosody, it is not specific to it: what makes prosody specific is its being the system for the grouping of pitch-related objects. As such, we should expect that it uses its primary element, pitch to express the grouping of its pitch-related objects. Our next experiment was thus aimed at finding the ways pitch variation plays a role in grouping such objects. Assuming that prosody does use pitch variation for grouping and that the representation of groups by pitch variation is based on the same general principles as we found in temporal grouping, we assumed that groups denoted by pitch variation would be created following the very same principles as those created by timing, i.e. by inherent grouping and recursion.

Material and methods

In order to preserve the abstract nature of the experiments and ensure that no factors other than abstract prosodic ones be considered, the settings of our original experiments were modified in such a way that, as visual stimulus, the original sequence of the character "•" was substituted by a sequence of capital letters such as *A*, *B, C* and *D*. In speech, they resulted in highly abstract prosodic objects which enabled us to study the role of pitch variation in prosodic grouping without the additional effects of lexical-semantic content on pitch. This time there were 13 subjects aged 18-50 involved in the experiment, including 11 university students and 2 others. They were presented with a sequence of capital letters and instructed that pairs of parentheses would suggest a structural relation: letters enclosed within them would constitute a single group, letters outside the parentheses would be outside that group. No further instructions were given.

The following patterns were used as stimuli:

(16) A B C D
(17) (A B) (C D)
(18) A (B C) D[6]

The readings of the stimuli by the subjects were digitally recorded and further analysed using the Macintosh program SoundScope/16 2.30. Boundaries of segments were marked manually and enhanced by a peak detecting function of the program.

Results and discussion

Interestingly, some of the subjects requested to repeat the recording saying they did not find their utterance representing what they had expected. It showed that in their responses they undoubtedly followed some internal faculty. This was exactly what our experiments were all about, consequently, we could be aware that we were heading in the right direction. *Table 8* shows the actual means and standard deviation data of the respected patterns:

Segments	**(16) A B C D**	**(17) (A B) (C D)**	**(18) A (B C) D**
a	548,42	441,5	903
Std Dev	319,53	144,91	313,12
b	578,25	899,75	387,62
Std Dev	284,65	301,86	72,39
c	501,67	351,83	666,46
Std Dev	297,19	96,81	177,11

Table 8

[6] Here and in the following experiment under *1.2.2* segment *a* represented the difference between the onset times for *A* and *B*, segment *b* the one between the onset times for *B* and *C*, etc.

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

First, we wished to see if timing retained its grouping function found in abstract grouping if associated with pitch variation. Comparing (17) and (18) to (2) and (5), their respective counterparts with visual variation only, we found that segment length and pattern type play exactly the same role: here, too, both parameters had a significant effect ($F_{(2,22)}$=10.046137, p<0.0001 and $F_{(1,23)}$=0.4660437, p<0.0033, respectively) with a similar interaction between the two ($F_{(2,22)}$=1.9888716, p<0.0001).

Segment length, i.e. whether the segment is short or long was similarly found to be structure dependent, as shown in the comparison of the first two segments of the above patterns; cf. *Figure 12*:

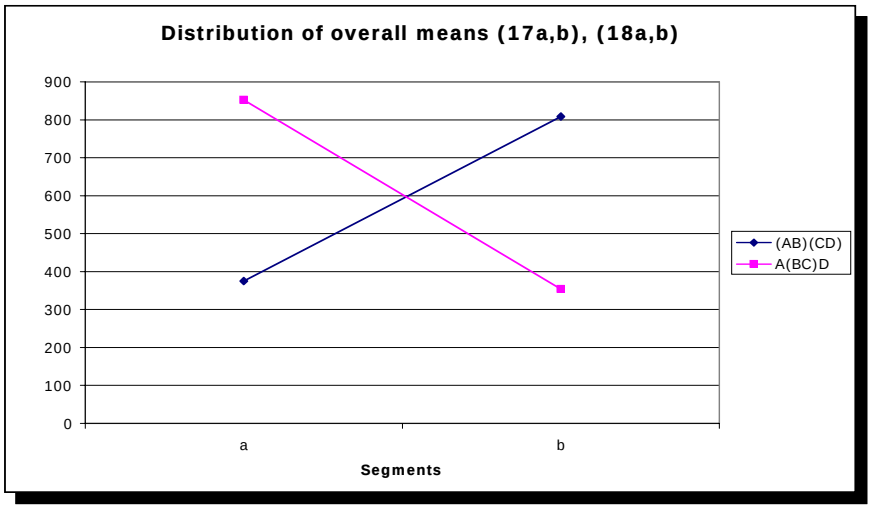


Figure 12

Segment length is found to be significant ($F_{(1,23)}$=8.1237594, p<0.0001) and segment type is not ($F_{(1,23)}$=0.0016991, p=0.8450). Again, no interaction was found between the two factors ($F_{(1,23)}$=0.0537401, p=0.2777). These results confirm that it is the internal composition of groups (the number of elements within a pair of parentheses) rather than the sequence of segments within the given pattern that determines the actual temporal representation of grouping, regardless of the nature of the stimulus.

We also wished to find out if inherent grouping (as in the abstract visual pattern (1)) could also be found in the grouping of pitch-related prosodic objects. We compared segment length in (16) *A B C D* and (17) *(A B) (C D)*; cf. *Figure 13*:

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

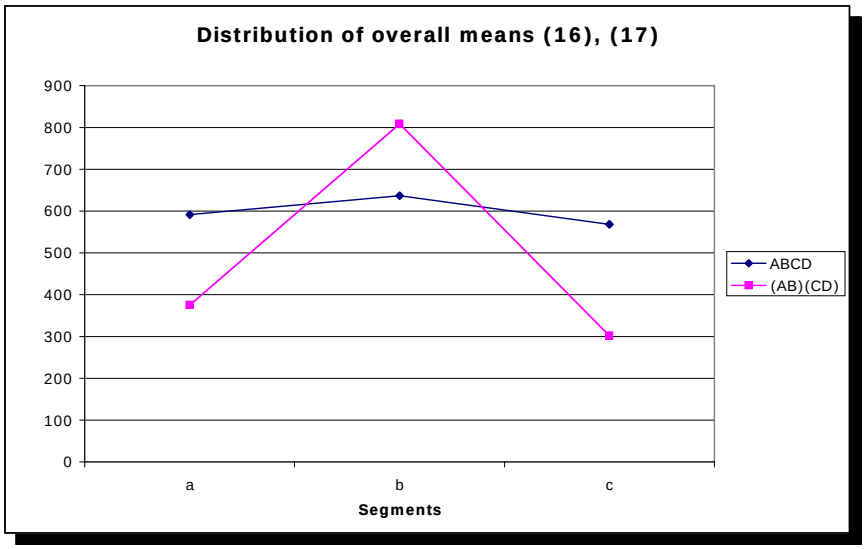


Figure 13

Similarly to abstract visual patterns, here again, segment length was a significant effect ($F_{(2,22)}$=9.466758, p<0.0001), but pattern type was also significant ($F_{(1,23)}$=0.2596755, p=0.6424). There was an interaction between the two effects ($F_{(2,22)}$=4.7717333, p<0.0001).

The within-subject comparison of (16) and (17) shows that segment length is a significant effect in all three segments:

segment a of (16) vs. segment a of (17): $F_{(1,23)}$=1.4511207, p<0.0001
segment b of (16) vs. segment b of (17): $F_{(1,23)}$=0.8293219, p=0.0002
segment c of (16) vs. segment c of (17): $F_{(1,23)}$=1.1897522, p<0.0001

It already suggests that (16) with no grouping denoted in the visual stimulus also consists of segments of different length. The inside-subject comparison gives further detail:

(16) segment a vs. segment b: $F_{(1,23)}$=0.3431009, p=0.0100
segment a vs. segment c: $F_{(1,23)}$=0.0714253, p=0.2127
segment b vs. segment c: $F_{(1,23)}$=0.6264271, p=0.0009

(17) segment a vs. segment b: $F_{(1,23)}$=6.3482201, p<0.0001
segment a vs. segment c: $F_{(1,23)}$=1.5087312, p<0.0001
segment b vs. segment c: $F_{(1,23)}$=8.1017008, p<0.0001

Accordingly, segments a and c of (16) are of similar length but the middle segment, b is significantly longer. It clearly demonstrates inherent grouping in (16).

Remember that inherent grouping for (1) was only established in its within-subject comparison with (2). How can we then account for the fact that in the corresponding pattern (16) it was available even inside its own pattern?

We suggest that, as it will be shown in section 1.2.3 in more detail, in addition to timing, pitch variation is an important means for the grouping of prosodic elements, and in the case of prosodic grouping it is pitch variation that is obligatory; timing plays a secondary, optional

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

role. This is what we see here, too: an overtly designated prosodic grouping should be primarily denoted by pitch variation and, optionally, temporal variation, as in (17). If grouping is not overtly designated (it is inherent), pitch variation should normally be avoided and temporal variation applied (as is the case of (16)). This is how we can account for the fact that in (1) inherent grouping is only relevant in the within-subject comparison. Namely, the only means of the grouping of visual elements (as in (1)) is temporal variation, accordingly, for inherent grouping it should only be applied to a lesser degree, with significance only in the within-subject comparison.

(One more remark: following pattern (2) we would expect the difference between the length of segment a and segment c in (17) not to be significant. The fact that it is significant, can be accounted for by a general tendency of acceleration in speech.)

Although, according to inherent grouping, one tends to group a set of elements even if such a structure is not overtly denoted in the stimulus, the difference between long and short segments will not be as big as when grouping is overtly marked in the stimulus; cf. Figures 14 and 15, the 100% stacked graphs for (16) and (17), respectively:

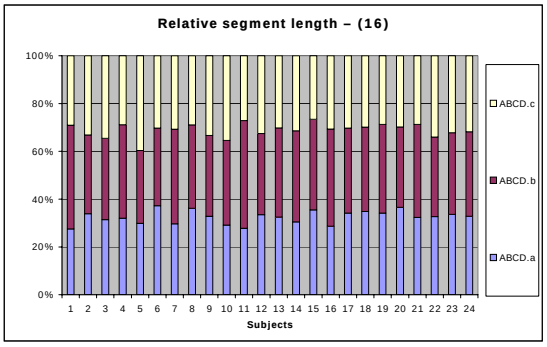


Figure 14

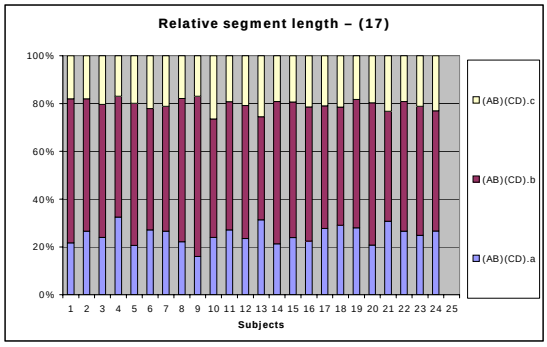


Figure 15

In sum: the experiments have shown that timing has essentially the same role of segmentation in both abstract visual and prosodic grouping. Temporal segmentation, the distinction between long and short segments plays a significant role in denoting grouping both when timing is the only means of grouping and when it is accompanied by other means (such as pitch-variation) as well. These experiments have also confirmed that grouping is inherent to the grouping of both abstract visual and prosodic elements.

1.2.2 Experiment 6: timing vs. grouping of abstract, pitch-related objects by embedding

Material and methods

The same subjects and methods were used as in the experiment described in *1.2.1*. The following structures with embedding were included in this experiment:

(19) A (B (C D)) E
(20) A (B (C D)) E F

Here again, we wanted to see if segment length was significant in these patterns. But in addition to the question whether a difference in the number of elements in a group is reflected by temporal segmentation we also wanted to find out if embedding itself has an effect on relative segment length, similarly to what we found in the abstract structures of (6) and (7).

Results and discussion

Table 9 shows the actual means and standard deviation of the respected patterns:

Segments	**(19) A (B (C D)) E**	**(20) A (B (C D)) E F**
a	1111.1	770.92
Std Dev	319.58	179.94
b	583.1	467
Std Dev	195.4	110.6
c	293.7	236.25
Std Dev	90.5	47.13
d	1249.3	1371.5
Std Dev	289.4	331.38
e		747.42
Std Dev		209.01

Table 9

The distribution of the overall means of (19) and (20) are shown in *Figure 16*:

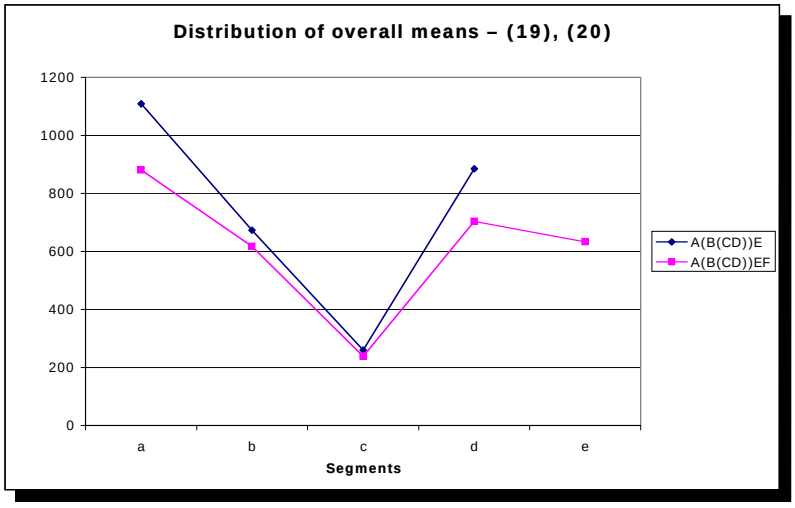


Figure 16

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

Comparing the corresponding segments in the above two patterns of embedding (from segment *a* through segment *d*), both segment length and segment type were significant ($F_{(3,20)}$=14.544487, p<0.0001 and ($F_{(1,22)}$=0.2936672, p=0.186, respectively). There was no interaction between the two effects ($F_{(3,20)}$=0.4457524, p=0.0563).

Comparing segments *a, b* (i.e. segments of embedding) and *d* (i.e. segment of de-embedding) in (19) and (20), we found that both segment length and segment type were significant ($F_{(2,21)}$=1.9410804, p<0.0001 and $F_{(1,22)}$=0.299637, p=0.3618, respectively) with no interaction between the two effects ($F_{(2,21)}$=0.2904888, p=0.0225). Similarly to the case of the corresponding abstract patterns of (6) **•(•(••))•** and (7) **•(•(••))••**, these data confirm that it is not just the number of elements in a group (enclosed in a pair of parentheses) that makes segment length significant but the level of embedding, too. Namely, the deeper the embedding, the shorter the segment becomes. This accounts for segment type as a significant effect. *Table 9* also confirms the difference we found in the length of embedding and de-embedding: similarly to (6) and (7), at the same structural level, the length of embedding is always shorter than the length of de-embedding. Accordingly, segment *b* is shorter than segment *d* in both (19) and (20). These data lead us to the conclusion that the temporal representation of embedding is recursive: the same rule ("make de-embedding longer than embedding") applies to any level of a complex structure.

In summary: the study of the relation between timing and grouping of abstract, pitch-related objects revealed that (a) the grouping of pitch-related objects has its temporal correlate of short and long segments, (b) timing is relevant to embedding and (c) the temporal segmentation of these objects is structure dependent and applied recursively. Finally, it was also confirmed that, although the grouping of abstract, pitch-related objects is equally represented by pitch and temporal variation, temporal variation continues to have the same two essential properties found in the grouping of abstract, non-prosodic elements: inherent grouping and recursive structure generation.

1.2.3 Grouping and tonal contour

Material

The experiment used the patterns and recordings already studied for their temporal representation above, i.e. (16)-(20).

Results and discussion

Since prosody groups pitch-related objects, variation by pitch was expected to play central role in grouping. The study of the tonal contour of these utterances showed that, in contrast to temporal variation whose share in the denotation of grouping varied according to pattern type, pitch variation proved to be mandatory and quite uniform. It was found that being part of a group or belonging to different groups was denoted by pitch unambiguously: belonging to one and the same group was denoted by a single, continuous tonal contour (i.e. a tonal contour with a virtually seamless succession of regular pitch intervals) and belonging to two different groups by two separate (non-continuous) tonal contours. Accordingly, a continuouscontour of fall–rise or rise–fall denoted two adjacent elements to belong to the same group (a fall–rise between *A* and *B* in (16) *ABCD*, *B* and *C* in (18) *A(BC)D* and a rise–fall between *C* and *D* in

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

(16)); in contrast, a non-continuous sequence of pitches denoted separate groups on the two sides of the boundary between adjacent elements (as, e.g., *B* and *C* in (17) (AB) (CD))[7].

Whereas in most cases belonging to a single group was denoted by a continuous fall–rise or rise–fall, in some instances belonging to the same group was denoted by both members of a single group produced with a continuous level tone. In this case group boundary was marked by a noticeable change of pitch. This was observed as an optional variation in (17) only. In this case the difference between two adjacent groups became even more straightforward: they became separated by two distinct pitch levels: the first group by a high (H) pitch, the second by a low (L) pitch, i.e. (H–H) (L–L), something like

A B
  C D

Accordingly, we observed pitch variation clearly used as the default and mandatory means for grouping. It was accompanied, as we saw earlier, by temporal grouping as well, much the same way as in cases of abstract visual (non-prosodic) grouping.

Importantly, pitch-variation was used recursively. It could be seen from the fact that a group consisting of two elements had the same tonal contour (fall–rise) regardless of the level of embedding: the same contour was found at the top of the hierarchy (as in segment *a* of (17)), in a linearly co-ordinated position (as in segment *b* of (18)) and embedded as segment *c* in both (19) and (20) – and this tonal contour was not found in a structurally different position, at group boundaries. The starting pitch of this contour also followed the embedding of the prosodic complex: the deeper the embedding, the lower the starting pitch of the contour of the given group was, whereas the character of the contour remained the same (fall–rise). It allows us to make the generalisation that recursion, i.e. the recursive use of pitch variation is a significant property of grouping of abstract, pitch related objects.

1.2.4 Experiment 7: prosodic grouping at the pre-school age

In order to filter out the chance that prosodic grouping is something that is the result of learning – this interpretation possibly enforced by the use of parentheses in the visual stimuli – we carried out a series of experiments with 12 pre-school children aged 4 to 6 as well. We wished to present these pre-school children with patterns suggesting groupings similar to those described in *1.2.1* and *1.2.2*, this time avoiding the use of parentheses. Similarly to our earlier experiments, we wished to see again how grouping is represented by prosody (the variation of the tonal contour).

Material and methods 1

In a pilot study preceding the actual experiments we wished to see how grouping by similarity was observed and denoted: similar and dissimilar objects were placed side by side, using equal space between all objects, such as:

(21) [pen] [pen] [pencil] [pencil]

(22) [pen] [pencil] [pencil] [pen]

[7] For the cognitive use of *rise* and *fall* in intonation see the analysis in Bolinger 1989.

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

Subjects were required to name the objects one after the other from left to right, using single words as 'pen' or 'pencil'. Recording was preceded by a familiarisation phase.

Results and discussion

Since these children had no prior experience with similar tests and since they differed noticeably regarding social background, they showed differences in their readiness for this kind of abstract experiment. As a tendency, however, we noticed that the production of some of the children was remarkably close to what we had found in our experiments with adults earlier: both pitch and temporal variation was used in grouping. However, in order to make the experiment more accessible to the majority of the children, we designed yet another test with less abstraction and carried it out with the same group at a later time:

Material and methods 2

We used pictures of dogs placed at distances suggesting grouping and the subjects were asked to say 'wow' as many times as they saw dogs. No instruction regarding grouping was issued. The patterns were as follows (spaces below represent the relative distance between the displayed objects):

(23) [dog] [dog] [dog] [dog]

(24) [dog] [dog] [dog] [dog]

(25) [dog] [dog] [dog] [dog]

This time we got more comprehensive results. We attribute the improved performance to the fact that the task was pragmatically more sensible. Whereas putting together words like 'pen' and 'pencil' without actually saying anything about their relation was pragmatically invalid for them, repeating 'wow' was not, an act they had performed several times before when imitating dogs. What was new (and essential) in the task, however, was that they were assumed (even if not told) to notice the given structure and repeat that word accordingly.

Results and discussion

As for the visually ungrouped pattern (23), all the 12 subjects represented it with a sequence of four words. One third of them, 4 denoted it as a set of two groups, all using temporal and three also pitch variation for grouping. As for (24), 11 of the 12 subjects (91.7%) gave a response and 8 of those (72.3%) produced a grouping as marked in the stimulus. Those three giving an incorrect response did not group into two halves. Three of the correct cases (37,5%) used both pitch and temporal variation, one (12.5%) only used pitch and four (50%) only time. The tonal contours used were as follows: rise–fall + rise–fall, rise–fall + rise–fall, rise–rise + fall–fall, fall–rise + fall–fall. As for (25), there were only 7 complete responses (58.3%) and only 3 (25%) were correct, all of which used the tonal sequence rise + fall–rise + fall. The incorrect responses either grouped incorrectly (as (*wow–wow)* + (*wow–wow*)) or did not notice grouping at all.

This latter experiment showed the following tendencies. The marking of the grouping in pattern (24) proved to be more successful than that in (25) suggesting that the interpretation of a sequence of two equal groups was more straightforward for them than that of a sequence

of three groups of alternating length. Since reaction time (time elapsed between seeing the stimulus and starting the response) in adults for the two corresponding patterns (2) (••)(••) and (5) •(••)• did not show such a relation with degree of difficulty (1231.65 ms and 1019.5 ms, respectively), we suggest that the degree of complexity of a structure (the number and internal composition of groups in a pattern) has a developmental effect on perception and interpretation.

But as a tendency we can see the ability of children at the age of 4 to 6 to use both temporal and pitch variation for grouping without any previous learning of how grouping can be denoted. The use of both time and pitch-variation follows essentially the same pattern as among adult subjects. (There have been earlier experiments showing that the perception of rhythm may be an innate capacity; cf. Petitto *et al.* 2003, Nazzi and Franck 2003. If so, our results with pre-school subjects may be further tested in experiments with younger subjects as well.)

Our data also show that temporal segmentation is more general at this age than pitch variation suggesting an important phenomenon of cognitive development: the grouping of pitch-related prosodic objects is based on the grouping of more abstract, non-prosodic objects. Timing, the principal means of grouping of abstract objects continues to have its share in more specific modalities, such as prosody as well, and may even play a central role until means specific to the given, less abstract modality are established in it.

1.2.5 Experiment 8: speech prosody and recursion

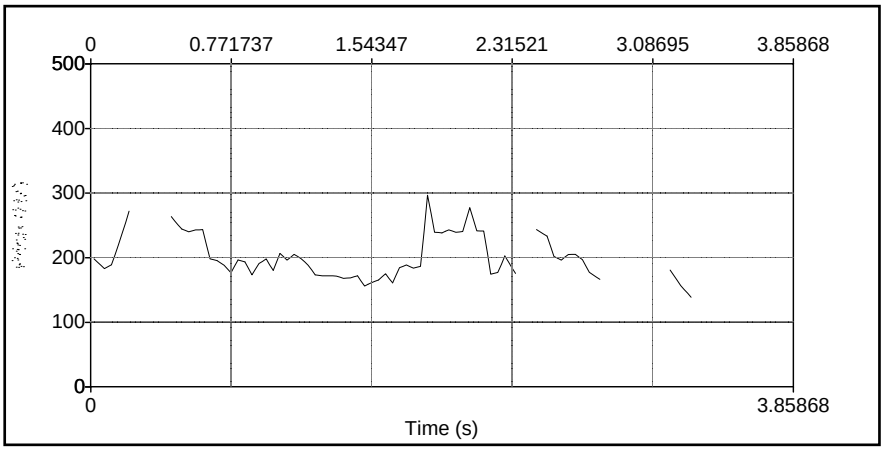
The analysis of sentences (14a), (14b) and (15) in section *1.1.4* showed that the temporal organisation of the speech representation of syntactic structures with recursive embedding is itself recursive. It suggests that at least one significant element of prosody, i.e. time/duration is used recursively to represent recursive structure. Next, we wanted to see if another important aspect of prosody, intonation through pitch variation also shows properties of recursion.

The following experiment was aimed at finding if linear discontinuity has its counterpart in intonation, a property which, if identified, would give support for an essential relation between abstract timing, speech prosody and syntax. As it is known from literature on prosody, there are important differences between structural segmentation in syntax and prosody. In particular, syntax is believed to display infinite recursion of embeddings whereas prosody appears to have a limited number of structural levels and, as a consequence, to have a flat structure (cf. Selkirk 1984). Those who allow recursion in prosody also limit it to some well-defined restricted cases (such as prosodic compounds, similar to lexical compounds; cf. Ladd 1996). In what follows we will examine how discontinuity of syntactic segments is represented by prosody, and, in particular, how the resulting tonal contours relate one to the other. We will re-examine the two kinds of sentences we studied for temporal organisation in section *1.1.4*: (14a) and (14b), cases of recursive embedding repeated here as (26a) and (26b), respectively, and (15), the case of insertion repeated as (27) and (28).

Material and methods

This time, an audio recording was made of (26). The change of fundamental frequency of an individual utterance was studied to see how embedding is represented by the tonal contour.

(26)
A macska, amit a kutya, ami megveszett, megharapott, elszaladt.
The cat that the dog that got mad bit ran away
'The cat that the dog that had got mad. bit, ran away.'

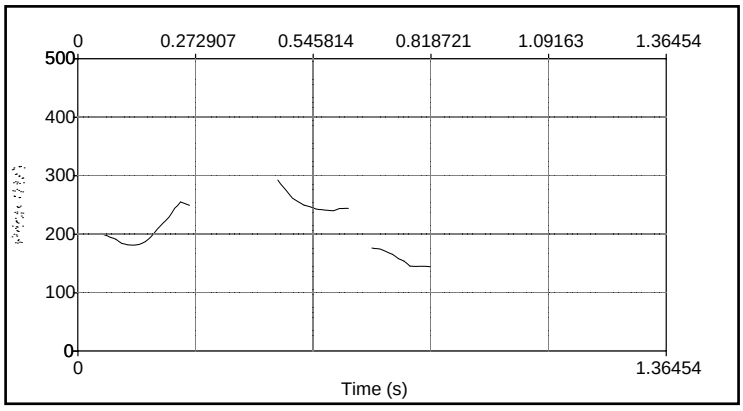


A macska, amit a kutya, ami megveszett, megharapott elszaladt

Figure 17

Figure 17 shows the fundamental frequency of the above utterance. Next, as shown in (26a), the embedded segments *amit a kutya, ami megveszett, megharapott* were removed and the remaining chunks *a macska* and *elszaladt* were concatenated to form a sentence *a macska elszaladt* 'the cat ran away'; cf. *Figure 18*:

(26a)
A macska [this segment removed] elszaladt.
The cat ran away
'The cat ran away.'



A macska elszaladt

Figure 18

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

As the tonal contour shows, the two, originally distant tonal segments join seamlessly to form a new contour which, as auditory perception confirmed, could be considered as if originally produced without any manipulation; i.e. there was no audible tonal indication of any removal. Next, in (26b), we took the embedded segments *amit a kutya, ami megveszett, megharapott* 'that the dog that was rabid bit', removed the embedded sequence *ami megveszett* 'that was rabid' and concatenated the remaining parts to get *amit a kutya megharapott* 'that the dog bit' again with the result that the remaining chunks fit seamlessly into a single contour (cf. *Figure 19*):

(26b)
amit a kutya [this segment removed] megharapott [this segment removed]
that the dog bit
'that the dog bit'

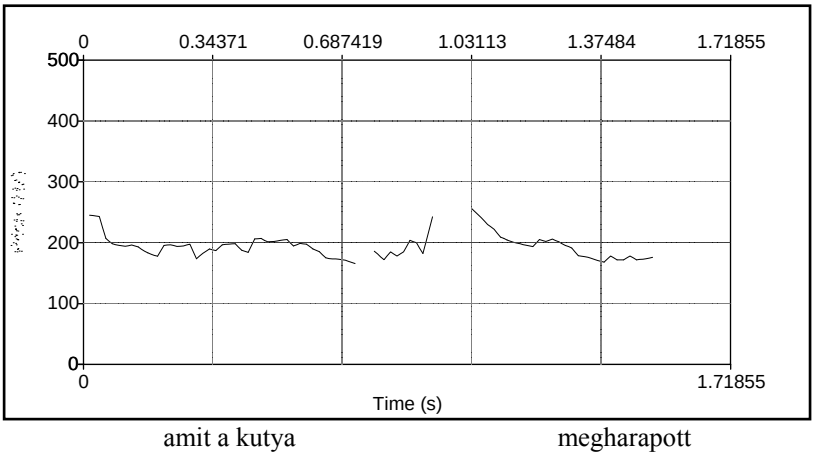


Figure 19

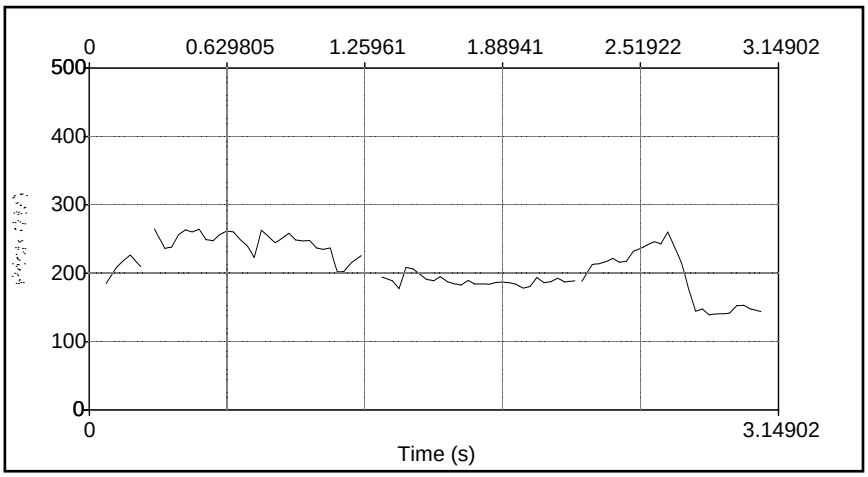
Results and discussion 1

This latter manipulation shows again that the syntactically matching discontinuous segments of the utterance are represented by matching tonal contours of the corresponding discontinuous prosodic phrases. Interestingly and importantly, this actual continuity of the tonal contour appears to defy tonal declination, the general observation that the tonal contour tends to gradually decline (with downsteps at certain phrasal boundaries) towards the end of the utterance. Namely, if declination were to apply, we should expect a pitch mismatch at the point of concatenation after the removal of embedded segments, however, this is not the case. This fact reveals that prosodic phrases are not just linearly ordered, but the prosodic relation between the discontinuous phrases is preserved by their virtually continuous tonal contour regardless of the distance of discontinuity. This allows us to suggest that such discontinuous chunks of tonal contours represent a single prosodic phrase (such as an IP; cf. Hunyadi *in preparation*).

In order to find out if discontinuous tonal matching can be considered an exclusive property of prosodic recursion we carried out the same kind of manipulation on an utterance which does not represent syntactic recursion; cf. (27) and (28):

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

(27)
Meg tudnád mondani, hogy – az én órám megállt – hány óra van?
Preverb you could tell that the I my watch stopped how many hour is
'Could you tell me – my watch has stopped – what time it is?'

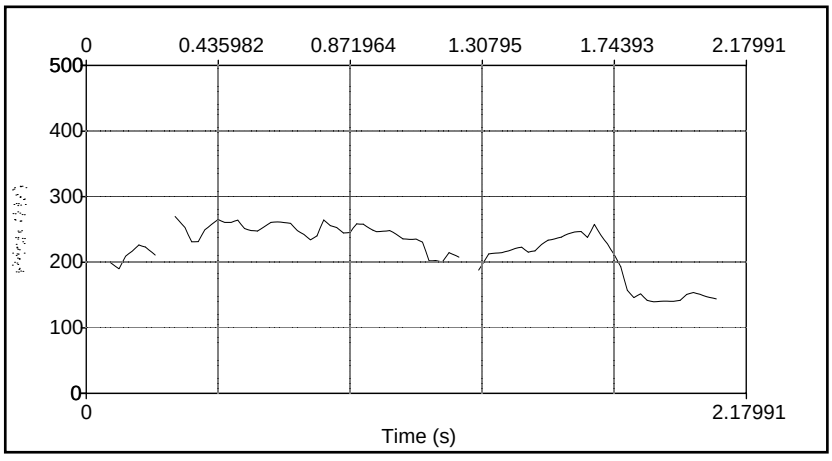


Meg tudnád mondani, hogy az én órám megállt hány óra van?

Figure 20

Figure 20 above shows the contour of the fundamental frequency for (27), a case of non-recursive discontinuity. In (27a) the inserted segment *az én órám megállt* 'my watch has stopped' is removed providing the contour of the fundamental frequency with a clear indication of tonal continuity again (cf. *Figure 21*):

(27a)
Meg tudnád mondani, hogy [this segment removed] hány óra van?
Preverb you could tell that how many hour is
'Could you tell me what time it is?'



Meg tudnád mondani, hogy hány óra van?

Figure 21

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

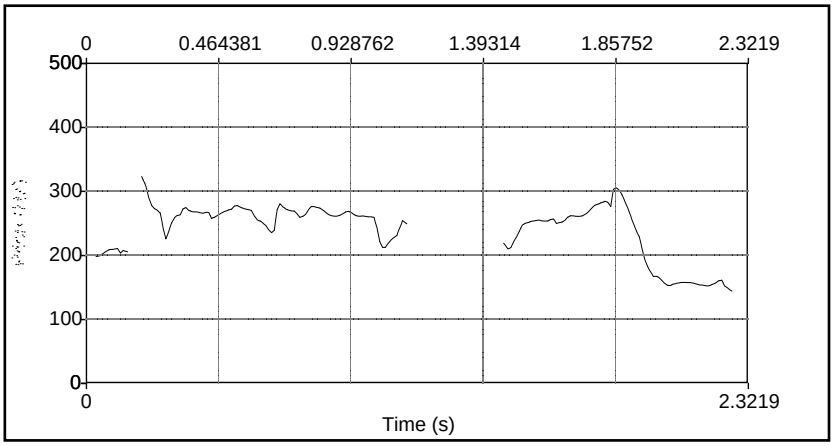
Results and discussion 2

Similarly to the previous experiment in (26), the discontinuous chunks thus concatenated form a tonal contour which was again confirmed by auditory perception as a natural single, continuous contour.

The analysis of these examples shows that the non-syntactic insertion of *az én órám megállt* 'My watch has stopped' has its clear prosodic correspondence: when this clause is removed from the sound wave, we can see that the tonal contour of the final clause *hány óra van?* 'what time it is' is a direct continuation of the first clause *Meg tudnád mondani, hogy* 'Could you tell me (that)'. It indicates that the IPs of the three respective clauses of (27) are not simply arranged in linear sequence (then they should have followed tonal declination), but the first and the third IPs are tonally closely related: they can be considered as two halves of a single IP. As a consequence, the IP of the second (inserted) clause, *az én órám megállt* 'my watch has stopped' needs to be considered prosodically inserted in the above discontinuous IP: [*Meg tudnád mondani, hogy* [...] *hány óra van?*].[8]

In order to confirm that the two chunks actually form a single prosodic phrase (such as an IP), we made an additional recording of an utterance of the form (27a) in real speech, without any manipulation; cf. (28):

(28)
Meg tudnád mondani, hogy hány óra van?



Meg tudnád mondani, hogy hány óra van?

Figure 22

Results and discussion

Comparing *Figure 21* and *Figure 22* we can see no essential difference between the two, both of them representing a single continuous tonal contour. The similarity of the two contours (one manipulated, i.e. concatenated and the other non-manipulated) indicates that whatever kind of interruption a prosodic phrase may undergo in the flow of speech (embedding or

[8] This issue of the prosodic representation of syntactic discontinuity has earlier been studied by Varga 2002 who arrived at a different conclusion. He studied the sentence "Találkoztam − jól ülsz? − az Angélával" 'I met – are you sitting all right? – Angela'. Since his example consisted of a question inserted in a declarative matrix, the overall intonation contour could not reveal an obvious discontinuous tonal matching, the rule that our experiment was able to show; accordingly, he accepted the three IP explanation.

insertion), we keep an eye on its contour and realise it in its entirety even in the lack of continuity.

This latter observation, i.e. the fact that discontinuous tonal matching is a property of prosodic phrase formation regardless of the type of the complexity of the prosodic structure (i.e. it equally applies to structures with and without embedding) indicates that discontinuous tonal matching is an important property of the overall organisation of prosodic phrases. While allowing for discontinuous chunks of the tonal contour to form a single prosodic phrase (such as an IP), it opens the way to further structural subdivisions by means of insertion as well as embedding.

Discontinuous Tonal Matching makes discontinuous structures available in prosody, a property which we can thus see is not unique to syntax, on the contrary, it is a property which extends the possibility of the mappings between syntax and prosody by broadening the conditions under which recursion can apply.

1.2.6 Temporal and tonal segmentation revisited: the bookmark effect

The comparison earlier between the temporal properties of embedding and co-ordination of abstract visual patterns showed a clear difference: we found that, whereas, in this most abstract case, a segment of embedding is, as a rule, shorter than the corresponding segment of de-embedding (cf. (6) •(•(••))• and (7) •(•(••))••), the co-ordination of the groups preceding and following the parenthesised group as found in the co-ordination of (5) •(••)• does not follow this rule (in fact, the last segment happens to be even shorter than the initial segment (the mean values of segment length for the corresponding segments of (5) are as follows: segment *a*: 837.41 msec, segment *c*: 711.20 msec, p=0.0008). We wanted to find out if the observed property of embedding denoted by the temporal variation of segment length is the distinctive rule for prosodic embedding as well.

The temporal segmentation of speech prosody of the sentence type (26) with embedding confirmed this property to be in effect (cf. our earlier analysis in the corresponding (14a), but, unexpectedly, the structurally distinct (27) with no embedding as analysed in (15) earlier proved to have a temporal structure corresponding to that of embedding again: the syllable durations for segments *a* and *c* were 185.68 ms and 416.39 ms, respectively. Since (26)/(14a) containing speech material has the same underlying grouping structure with embedding as the abstract patterns (6) and (7) with visual elements only, we would have all reasons to expect that temporal variation found in (6) and (7) should also apply to (26)/(14a). Consequently, temporal variation should be the underlying rule for the differentiation of embedding and de-embedding, regardless of the grouped elements (abstract visual patterns or elements of speech). Similarly, the structural identity of (27) with (5), the latter having no embedding and showing no temporal variation specific to embedding, would suggest that we should expect this kind of temporal variation not to apply to (27) either. Since, contrary to our expectation, it does, we need to seek for an account for the lengthening of segment *c* in (27) for reasons other than temporal variation.

We suggest that in speech prosody this segment lengthening of structures of co-ordination is the effect of the increased complexity of the objects of prosodic grouping: the addition of pitch variation to temporal variation as means of grouping results in the increase of the load on calculation.

The observed lengthening of the segment following the inserted segment is the blueprint of an increased calculation caused by the bookmark effect. In either embedding or co-ordination

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

in speech prosody, in addition to determining segment length, discontinuous tonal matching must also be satisfied. The bookmark effect is the prosodic means of denoting discontinuous pieces of structure in general: its function is to signal that two non-adjacent groups make one structural unit regardless of its kind, whether embedding or co-ordination. The bookmark effect involves an important cognitive process: in general, for a sequence of groups *x*, *y* and *z*, if groups *x* and *z* are structurally related and are separated by group *y*, this relation between *x* and *z* is prosodically marked by the discontinuous tonal matching of *x* and *z*. Accordingly, the parameters of the transition from segment *x* to segment *y* are stored and, when making the transition from segment *y* to segment *z* (the return to the original final pitch of segment *x*) those stored data are retrieved to be used as the starting pitch of segment *z*. Due to the increased complexity of tonal information (as compared to the temporal one), the cognitive process of its retrieval is computationally more expensive. Thus this bookmark effect results in the lengthening of segment *c* in the non-embedding structure of (27) as well. Since pattern (5) with the grouping of visual elements does not involve embedding and since no tonal variation is involved, the lengthening of segment *c* does not occur either. As for the lengthening of segment *c* of (26) with embedding, it is both the result of the rule of temporal variation (as also found in the case of abstract visual grouping in (6) and (7)) and that of the bookmark effect to satisfy discontinuous tonal matching. Since patterns of type (6) and (7) do not involve tonal grouping, this additional effect of lengthening does not apply and the lengthening of the segment of de-embedding (segment *c*) is fully due to temporal variation. Finally, our assumption that discontinuous tonal matching is structure-dependent, i.e. the given non-adjacent groups represented by segments *a* and *c* must form some discontinuous structure is shown by the comparison of speech segmentation, as in (26)/(14a) and (18) A (BC) D. They are similar in that (a) they have the same sequence of long + short + long segments and (b) in both patterns elements are also grouped by pitch variation. However, in contrast to the former pattern, segment *c* of (18) is shorter than segment *a* (666,46 ms and 903 ms, respectively), i.e. (18) does not show the bookmark effect. We can assume that these two segments are not in immediate structural relation (segment *BC* is not an insertion), thus no discontinuous tonal matching, supported by the bookmark effect, is required. The same accounts for the shortening of segment *c* of (5), and its observed difference from (15) is also based on (15) being an instance of insertion where the bookmark effect should apply.

1.2.7 Prosodic vs. syntactic phrasing

The comparison of the domains of prosodic phrases in (26) above with its corresponding syntactic domains suggest an obvious correspondence (with the only difference that the prosodic boundaries appear to be shifted right as compared to the syntactic ones). It is important to decide if prosodic phrasing is the result of its mapping onto syntax. Our earlier experiments on the relation between grouping and non-lexical tonal contour in section *1.2.3* demonstrated that tonal contour is determined recursively, depending on the given structure of embedding. Since it applies to structures with no syntactic categories but the recursive embedding of just abstract, pitch-related objects, we must see that the grouping of pitch related objects follows a more general, tonal principle of grouping. Accordingly, we can also suggest that tonal grouping in speech prosody also follows this more general, tonal principle rather than mapping onto syntactic structure. The virtual correspondence of syntactic and prosodic phrase formation in sentences like (26) can then be explained by the following:

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

(a) since the same basic propositional content should be expressed by both syntax and prosody, they should represent the same grouping relations between the given segmental elements, and
(b) the application of the principle of discontinuous tonal matching makes this representation in prosody possible.

Accordingly, discontinuous tonal matching structurally associates linearly discontinuous segments of a phrase, and it can be used recursively to denote hierarchical structure, including the appropriate levels of embedding: each level is embedded into a higher level denoted by the lowering of the pitch of the corresponding tonal contour.

To sum up: our earlier experiments showed that abstract temporal grouping is recursive and it can be extended to the grouping of prosodic objects (cf. section *1.1.4*); pitch related objects are grouped by both temporal and pitch variation, where, in addition to temporal variation, pitch variation is also recursive (cf. *1.2.3*). Temporal and pitch variation being the main properties of prosody we can conclude that prosody displays essential properties of being recursive: these two prosodic means, when applicable, go hand in hand in the recursive grouping of prosodic structure, based on the principle of discontinuous tonal matching which extends the range of structures prosodic recursion can apply to, resulting in a broader set of mappings between syntax and prosody.

2 Processing time as a function of complexity and length

Differences in segment length showed that variation by time is indeed a significant means of the representation of structurally different abstract patterns. The timing of boundaries was also shown to be dependent on the level of embedding. In order to show that the complexity of embedding appears in perception as well, we measured the time needed for the processing of the visual stimuli by the subjects before their response. In the following section we will show how processing time is related to the number of visual elements and the complexity of patterns.

2.1 Number of visual elements as stimuli vs. processing time

Average processing time is given in ms for each pattern:

1. 4 elements:
   •(• • •): 1136.74
   •(• (• •)): 1568.26
   •(• • (•)): 2181.98
2. 5 elements:
   •(• • (• •)): 1164.55
   •(• • • (•)): 1181.76
   •(• • • •): 1239.86
   •(• (• • •)): 1354.43
3. 6 elements:
   •(• • • (• •)): 1334.46
   •(• • (• • •)): 1837.47
   •(• • • • (•)): 2298.84

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

4. 7 elements:
   •(• • • • (• •)): 1435.04
   •(• • (• • • •)): 2049.96
   •(• • • (• • •)): 2057.63
5. 8 elements:
   •(• • • • (• • •)): 1387.06
   •(• • • (• • • •)): 1418.71
6. 9 elements:
   •(• • • • (• • • •)): 1771.56

Looking at these data we can notice that although there is some relation between the number of elements in a pattern and processing time (e.g., the greatest processing time values are found in rather long, 6- and 7-element patterns, or a 9-element pattern needs more processing time than an 8-element one), but this is not a straightforward relation (we can observe significant differences inside each group, i.e. according to the arrangement of elements as well). Accordingly, structure has an important effect on processing time. This is what we will look at in the next section.

2.2 Complexity of patterns of visual elements as stimuli vs. processing time

The following list demonstrates that processing time is not directly related to the number of visual elements in a pattern:
Average processing time is given in ms for each pattern:
In increasing order:

•(• • •): 1136.74
•(• • (• •)): 1164.55
•(• • • (•)): 1181.76
•(• • • •): 1239.86
•(• • • (• •)): 1334.46
•(• (• • •)): 1354.43
•(• • • • (• • •)): 1387.06
•(• • • (• • • •)): 1418.71
•(• • • • (• •)): 1435.04
•(• (• •)): 1568.26
•(• • • • (• • • •)): 1771.56
•(• • (• • •)): 1837.47
•(• • (• • • •)): 2049.96
•(• • • (• • •)): 2057.63
•(• • (•)): 2181.98
•(• • • • (•)): 2298.84

Processing time is quite telling with at least three observations. First, processing time was not directly related to the number of the elements "•", instead, to the internal structure of the pattern. Second, every pattern except for one consisted of three groups: a single character "•" (group *a*) followed by a complex of a pair of groups (group *b* and group *c*) consisting of these elements. In all cases of pairs with a symmetrical arrangement of the element "•" (group *b* (marked by segment *b*) of one pattern would be equal to group *c* (marked by segment *c*) of the other and vice versa, such as •(• • • (• •)) and •(• • (• • •)), processing time was always shorter for the pattern with group *c* consisting of fewer elements than group *b*. It suggests that the temporal representation of the patterns followed a real-time processing: the duration of group *c* served as a reference for the duration of the preceding group *b*. As we saw from the rhythmic analysis of such patterns in section *1.1.3*, there was a tendency towards symmetry, i.e. to represent the two groups with equal length and it was more straightforward to add the missing pauses (of the duration of a single "•") to the group on the right than to calculate such pauses for the duration of the left group in advance. (This observation supports an earlier finding according to which planning both in speech and music is facilitated by the events' metrical similarity and serial/temporal proximity and by developmental changes in short-term memory (cf. Palmer *et al*. 2003). Accordingly, we can suggest that it is computationally less expensive to adjust the temporal structure of the group on the right to the one on the left than vice versa: to look ahead, calculate the rhythmic structure of a group that is only realised later, and then calculate the rhythmic structure of the preceding group accordingly is computationally more expensive). Third, with the exception of the pattern •(• • (•)), the patterns with a single embedded character "•" on the right proved to be the most difficult of all patterns to process regardless of the number of these characters. Since such patterns are really difficult to interpret, the subjects needed more time to make a special decision how to represent them.

3 Grouping, sentence prosody and syntax

Our experiments have shown that we have reason to suggest that important properties of prosody are based on the more general principles of abstract grouping, including inherent grouping and recursion. When we talk about prosody, we may inevitably ask the question to what extent such principles can be applied to syntax as well. Since prosody is a way to acoustically represent the structural relations contained and expressed by syntax and as such it functions as an important interface between the speaker and the hearer, it is generally accepted that there must be some well-defined correspondence between the two structures. However, attempts to map prosody completely onto syntax have failed and it is widely agreed upon that one of the important reasons is that whereas syntax is of indefinite depth due to theoretically infinite recursion, prosody is flatter having a definite depth with no recursion (cf. Halliday 1966, Selkirk 1984 [cf. her Strict Layer Hypothesis], Nespor and Vogel 1986 etc.). From this significant difference it also follows that it is not possible to completely map one structure onto the other. However, when we are looking at prosody from the point of view of abstract grouping and find that, similarly to abstract grouping, prosody also displays features of recursion, we may suggest that this very aspect of prosody may nevertheless open our way to find its mapping onto syntactic recursion as well (Ladd 1996's proposal of the so-called compound prosodic phrase is an interesting idea which, however, limits recursion to just specific cases, cf. Ladd 1996).

Another approach may be to suggest that, in contrast to current generative syntactic theory syntax may not be the (only) module of grammar which prosody is based on or is a representation of, but that prosody has functions directly related to semantics and information structure. This is the approach prominently manifested in Steedman's proposal (for a recent discussion of the issue cf. Steedman 2000). His CCG approach leaves the question open if it actually models the way how, by what calculations a speaker actually maps prosodic structure onto syntax.

We suggest that prosody with its primary grouping function plays an important role in identifying and representing linguistic structure. In this sense it does not contradict the structure represented by syntax, but it also offers additional, especially semantic and pragmatic information.

The expression of grouping in general addressed in the experiments of this paper takes two important forms: linear (based on adjacent elements) and non-linear (the relation between non-adjacent, discontinuous elements). Whereas syntax is fundamentally based on the linear sequence of words and its possibilities are therefore limited to express additional, semantic and pragmatic relations, prosody appears to play a role in both kinds of grouping.

Namely, in a prosodic phrase like ((wx)(yz)) there is a linear relation between *w*, *x*, *y*, *z*, a similar relation between *w* and *x* and between *y* and *z (*expressed by temporal and/or pitch variation), but there is also a non-linear relation between *w* and *y* (this latter expressed by pitch as well as rhythm and accent).

Languages may use both ways to express such relations. Whereas linear grouping may be part of The Basic Variety (cf. Klein and Perdue 1997, Jackendoff 2002), non-linear grouping may be more sophisticated and related to the information content of the utterance (including focusing and prosodically expressed scope).

Steedman's examples (cf. Steedman 2000; using his numbering) (3a) *Marcel (proved completeness)* and (3b) *(Marcel proved) completeness* can both be possible groupings for the following reasons:

Both can be considered as prosodic variants of a yet simpler variant without any prosodic subdivision: *Marcel proved completeness*. This is the basic variety in which it corresponds to the structure suggested by syntax. As for its pragmatic content, it is a neutral utterance, without any further relation to be specified by further prosodic specification. The prosodic structure of both (3a) and (3b) is derived from this same prosodic utterance by recursive grouping: it is the additional (non-propositional) semantic/pragmatic content that determines derivation. As a result, (3a) and (3b) differ in their prosodic structure with respect to the groupings made to express focusing but go back to the same underlying prosodic structure which is not in contradiction with the given syntactic structure. These examples demonstrate that syntax determines linear order to express the fundamental propositional content. If no additional relation is to be expressed (as in a neutral utterance), the utterance receives a non-specific, (under-specified) prosodic structure. If some additional relation is to be expressed, this underlying non-specific prosodic structure is further specified by recursive grouping and, since it is a derivation, it does not contradict the underlying prosodic (and also syntactic) structure.

It depends on the typological characteristics of a given language what particular relations its syntax expresses by linear order. But the relation between syntax and prosody remains to be the same: the underlying prosodic structure will express whatever relations the given syntax expresses; any additional relation is expressed by recursive prosodic derivation.

László Hunyadi: Grouping, the cognitive basis of recursion in language
Argumentum, 2 (2006), 67-114
Kossuth Egyetemi Kiadó (Debrecen)

Hungarian is a relatively free word-order language with pre-verbal operator positions for focus and the universal quantifier (cf. E.Kiss 1987) so that scope can be expressed by syntax (linear order). However, there are instances when the expression of some contextual and pragmatic information (such as focusing) has priority in determining the linear order of words, and, as a result, scope cannot be denoted by the same linear order. Such is the case of inverse scope where it is prosodic derivation that expresses the given scope relation; cf. Hunyadi 2002. Consider now the following Hungarian examples (syllables in capitals denote heavy accent, parentheses indicate prosodic grouping):

(29)
(JÁNOS oldott meg minden feladatot.)
John solved Pref every exercise
'It was John who solved all the excercises.' (all the others did not solve every)

(30)
((JÁNOS oldott meg) (MINDEN feladatot.))
John solved Vmod every exercise
'It was John who solved all the excercises.' (all the others did not solve any)

According to standard syntactic analysis, the two sentences above have the same phrase structure. The prosodic analysis shows, however, a different segmentation. Whereas the same syntactic structure corresponds to the same overall, declarative intonation contour, in both utterances, this distinction in prosodic segmentation reveals an important semantic (logical) difference. (29) is built around a single IP where the focus *JÁNOS* has wide scope over the quantifier phrase (*minden feladatot*). (30) is, however, the case of inverse scope denoted by a distinct prosodic grouping as shown by the parentheses: as a rule, a stressed universal quantifier following the verb has wide scope over the pre-verbal focus.

What is the role of prosody in denoting inverse scope? Actually, this is the case of tonal reconstruction: both *MINDEN and JÁNOS* in (30) start at the very same pitch as in (31), a synonymous variant of (30):

(31) (MINDEN feladatot) (JÁNOS oldott meg)
every excercise John solved Pref
'It was John who solved all the excercises.' (all the others did not solve any)

In (31) the quantifier starts higher than the focus, a necessary prosodic condition for the quantifier to indicate wide scope over the focus. Tonal reconstruction as observed in (30) simply reconstructs this tonal pattern of (31): inverse linear order is accompanied by inverse tonal order (lower pitch followed by higher pitch), and interpretation is done by the reconstruction of direct tonal order.

To sum up, regarding the relation between intonation and its structural function, we have seen three kinds of tonal variation in Hungarian, each with its own role: (a) tonal continuity (higher pitch followed by lower pitch) represents a continuous prosodic phrase (IP) and it is the prosodic condition for direct scope, (b) discontinuous tonal matching (similar non-adjacent pitches) links discontinuous constituents together as a single prosodic phrase (IP), and (c) tonal reconstruction (lower pitch followed by higher pitch) serves the proper interpretation of scope in utterances with inverse tonal order.

Although forms of tonal segmentation can vary by language type, languages agree in that it serves as a means to extend the possibilities offered by linear ordering alone. The straightforward case is when syntax and prosody coincide both with respect to temporal and tonal segmentation. This is the case of the basic variety (the given prosodic structure coincides with the underlying prosodic structure). It can further be applied recursive derivation by temporal and tonal variation to produce further groupings to express additional, especially semantic or pragmatic relations.

An additional temporal boundary within the same tonal contour does not alter the expression of the basic propositional relations, it adds, however, further, information (such as focusing, hesitation, etc.) which linear order within the same tonal contour cannot convey. This is the function we find in the different groupings of Steedman's (3a) and (3b). A change in tonal variation, however, can result in the denotation of different logical relations, such as in the Hungarian examples of (29) and (30). In addition, as the comparison of (30) and (31) demonstrated, tonal variation in the form of tonal reconstruction can be used to virtually reconstruct linear order to derive information from the reconstruction of direct order.

In sum: the underlying prosodic structure of an utterance coincides with the corresponding syntactic structure as the basic variety. The scale of these syntactic phrases is essentially dependent on the language type. These prosodic properties can be further specified by recursive prosodic sub-grouping into smaller, linearly adjacent or non-adjacent groups to further specify the information content of the sentence. As such, linear and prosodic variation serve as structural dimensions with their own distinct functions without one contradicting the other.

As a result, syntax and prosody are strongly related. This relation is well manifested in the fact that linear and non-linear forms of prosodic grouping can represent both the syntactic phrase structure and the semantic, information content of the given sentence. Following these considerations, we can suggest the possibility of a direct lineage between abstract temporal grouping and syntax in which the recursive property of abstract grouping may have been established in grammar too. This assumption supports the claim according to which syntax and phonology, especially intonation are independent combinatorial systems so that certain aspects of one system are related to those of the other by interface constraints (cf. Jackendoff 2002), so that both have the same propositional content as their basis. Our experiments show that what makes this interface relation work is that, similarly to syntax, prosody also shares the important property of recursive grouping. Accordingly, each combinatorial system has its own main means for the expression of grouping. In a highly abstract combinatorial system this means is time (duration) with its principles of inherent grouping and recursion. In prosody it is pitch, and in syntax it is phrase structure. The main means of grouping are inherited in less abstract combinatorial systems and serve as means to express interface relations. Accordingly, although it is variation by pitch that is the main means of the grouping of abstract prosodic objects, temporal variation is also present to specify the interface between abstract prosody and other similarly abstract combinatorial systems, including music and dance. Its main principles of generating structure are inherent grouping and recursion. Syntax, in its turn, is based on the categories NP and VP; structuring of a sentence into groups consisting of these two can be seen as the instance of inherent grouping and recursion. The fundamental patterns of pitch variation serve as an interface between syntax and prosody (and, via prosody, other combinatorial systems as well). Accordingly, the underlying prosodic structure of an utterance corresponds to its syntactic structure due to the fact that both express the basic propositional relations; further prosodic (temporal and/or tonal) grouping is

determined by the interrelation with other combinatorial systems, including semantics and pragmatics.

4 Recursion in non-human cognition?

Our series of experiments have shown that recursion is not the unique property of syntax: temporal variation can be found at even more abstract cognitive levels as well. The question, however, whether recursion can be extended to non-humans is still open. The results of our experiments may contribute to this debate as follows.

Since temporal grouping is essentially what rhythm is about, we may want to find out how rhythm is represented in non-humans and what differences can be established. It has been shown (Ramus *et al.* 2000, Tincoff *et al.* 2005) that, similarly to new-borns, tamarin monkeys can also differentiate between languages of different rhythmic groups on the sentence level. Their comparison also showed that neither new-borns, nor tamarin monkeys can differentiate between languages belonging to two different rhythmic groups if a sentence is played backwards . Ramus *et al.* 2000 infer that at least some aspects of human speech perception may have been based on pre-existing sensitivities of the primate auditory system. The suggestion by Tincoff *et al.* 2005 is that, due to these similarities, the mechanism for rhythmic language differentiation is unlikely to have evolved specifically for language.

Remember that our experiments showed that temporal grouping at any level of abstraction was structurally (rather than linearly) dependent. Namely, the difference between long and short segments was not a function of linear order. Groups were not identified as long or short with regard to their linear order, instead, it was their structural position that determined length. We also saw that the length of groups in a hierarchical structure was calculated relative to their position in that hierarchy. Accordingly, we concluded that the temporal variation of segments (and rhythm, in general) was structure-dependent. Structure was evaluated top-down, but the highest node always coincided with the leftmost element. Consequently, linear order was significant in identifying hierarchy rather than actual segment length. If the above data on rhythm perception by new-borns and tamarin monkeys are suitable for making a generalisation, then our results suggest that the reversal of speech prevents the recognition of hierarchy and, consequently, rhythm in general, probably for the fact that the point of reference for the calculation of rhythm comes in this case last. In contrast, from the observation that new-borns and tamarin monkeys could in fact recognise speech if played in natural order we can suggest that both new-borns and tamarin monkeys had the ability to perceive the hierarchical organisation of elements. In that case, it is an indication that they may possess the general cognitive mechanism of recursion. If this assumption is correct, we can hope to have identified an important precursor of syntactic recursion beyond human cognition.

5 Summary

Our initial task was to account for the perception that utterances of one and the same sentence by different speakers appear to have virtually identical rhythm and relative timing. Our assumption was that this virtual identity is based on some general principle or principles of grouping originating outside language. It was further assumed that temporal variation is one of the fundamental forms of grouping that can be studied in structures consisting of elements of a highly abstract nature. Experiments were carried out involving the temporal representation of abstract visual elements and the findings were mapped onto results of experiments involving abstract pitch-related prosodic objects. Finally, these results were matched against temporal data from real natural language utterances. The experiments carried out showed that temporal variation is based on at least two basic principles: inherent grouping and recursion. Different levels of embedding were found to be denoted by differences in the duration of boundaries so that embedding at a given level was found shorter than de-embedding (the return to the original level of embedding). It was also found that the most straightforward function of temporal variation is the case of abstract visual objects with no means of representing grouping other than timing. In the case of abstract pitch-related objects, timing was observed to have similar functions as found in abstract visual grouping, however, priority was given to pitch variation. In the case of real natural language utterances both temporal and pitch variation took part in the representation of grouping, however, priority was given to syntax. The issue of the relation between syntactic structure and prosody was also addressed. Further support was found for the view that the two systems are related but are not identical. Syntax only determines the skeleton of the underlying prosodic structure of an utterance which essentially corresponds to the given syntactic structure,. Further small-scale prosodic phrasing is carried out by recursive temporal and pitch-related grouping to denote additional, especially semantic and pragmatic relations. This grouping of the prosodic material is made possible by the principle of recursion that makes possible to derive smaller embedded tonal contours (IPs) from the underlying larger tonal contour.

Recursion was found to be the fundamental principle that underlies abstract temporal as well as pitch-related grouping and connects them to grouping in speech prosody. Again, it is the principle of recursion that makes it possible for syntactic and prosodic structure to be directly related. Finally, on the basis of these data and considerations we can make the conclusion that recursion found in syntax is not some property that is only characteristic of language as part of human cognition, instead, it is a property which has its cognitive history in highly abstract temporal grouping in general.

6 References

Fitch, W.T., Hauser, M.D. & Chomsky, N.: The Evolution of the Language Faculty: Clarifications and Implications. *Cognition* (in press).

Fraisse, P. (1982): Rhythm and tempo. In: Deutch, D. (ed.): *The Psychology of Music.* New York: Academic Press, 149-180.

Fraisse, P. (1984): Perception and estimation of time. *Annual Review of Psychology* 35, 1-36.

Friederici, A.D., Gunter, T.C., Hahne, A. & Mauth, K. (2004): The relative timing of syntactic and semantic processes in sentence comprehension. *Neuroreport* 15(1), 165-9.

Halliday, M.A.K. (1966): The concept of rank: a reply. *Journal of Linguistics* 2, 110-118.

Hauser, M.D., Chomsky, N. & Fitch, W.T. (2002): The Faculty of Language: What Is It, Who Has It, And How Did It Evolve? *Science* 298, 1569-1579.

Hunyadi, L. (2002): *Hungarian Sentence Prosody and Universal Grammar*. Tübingen: Peter Lang.

Hunyadi, L.: Can prosodic structure be recursive? (*in preparation*)

Jackendoff, R. (2002): *Foundations of Language. Brain, Meaning, Grammar, Evolution*. Oxford University Press.

Pinker, S. & Jackendoff, R. (2005): *The Faculty of Language: What's Special about it? Cognition* 95, 201-236.

É.Kiss, K. (1987): *Configurationality in Hungarian*. Budapest: Akadémiai Kiadó.

Klein, W. & Perdue, C. (1997): 'The Basic Variety, or: Couldn't Language be Much Simpler?' *Second Language Research* 13, 301-347.

Kraut, M.A., Calhoun, V., Pitcock, J.A., Cusick, C. & Hart, J. Jr. (2003): Neural hybrid model of semantic object memory: implications from event-related timing using fMRI. *Int Neuropsychol Soc*. 2003 Nov. 9(7), 1031-40.

Ladd, R.D. (1996): *Intonational Phonology*. Cambridge University Press.

Lappin, Sh., Levine, R.D. & Johnson, D.E. (2000): The Structure of Unscientific Revolutions. *Natural Language and Linguistic Theory* 18, 665-671.

Lerdahl, F. & Jackendoff, R. (1983): *A Generative Theory of Tonal Music.* Cambridge: The MIT Press.

Mates, J., Müller, U., Radil, T. & Pöppel, E. (1994): Temporal integration in sensorimotor synchronization. *J. Cogn. Neuroscience* 6, 332-340.

Nazzi, Th. & Franck, R. (2003): Perception and acquisition of linguistic rhythm by infants. *Speech Communication* 41, 233-243.

Nespor, M. & Vogel, I. (1986): Prosodic domains of external sandhi rules. In: Hulst, H. van der & Smith, N. (eds.): *The structure of phonological representations*. Vol. I. Dordrecht: Foris, 225-255.

Palmer, C., Pfordresher, P.Q. (2003): Incremental planning in sequence production. *Psychol Rev.* 110(4), 683-712.

Petitto, L.A., Holowka, S., Sergio, L.E., Levy, B. & Ostry, D.J. (2003): Baby hands that move to the rhythm of language: hearing babies acquiring sign language babble silently on the hands. *J Cognition*, 2003.10.07.

Pierrehumbert, J. (1980): *The Phonology and phonetics of English intonation*. PhD thesis, MIT.

Pierrehumbert, J. & Liberman, M. (1982): Modelling the fundamental frequency of the voice. (Review of Cooper and Sorensen 1981) *Contemporary Psychology* 27, 690-692.

Pöppel, M. & Wittmann, M. (1998): Time in the Mind. *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge: The MIT Press, 841-842.

Pylkkanen, L., Stringfellow, A. &, Marantz, A. (2002): Neuromagnetic evidence for the timing of lexical activation: an MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain Lang* 81(1-3), 666-78.

Ramus, F., Hauser, M.D., Miller, C., Morris, D., & Mehler, J. (2000): Language discrimination by human newborns and cotton-top tamarin monkeys. *Science*, 288, 349–351.

Selkirk, E.O. (1984): *Phonology and Syntax: The Relation between Sound and Structure.* Cambridge: MIT Press.

Sanders, L.D., Newport, E.L. & Neville, H.J. (2002): Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. *Nat Neurosci* 5(7), 700-3.

Schiller, N.O., Bles, M. & Jansma, B.M. (2003): Tracking the time course of phonological encoding in speech production: an event-related brain potential study. *Brain Res Cogn Brain Res.* 17(3), 819-31.

Steedman, M. (2000): Information Structure and the Syntax-Phonology Interface. *Linguistic Inquiry* 31, 649-89.

Tincoff, R., Hauser, M.D., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005): The role of speech rhythm in language discrimination: further tests with a non-human primate. *Developmental Science* 8:1, pp. 26–35.

Varga, L. (2002): *Intonation and Stress, Evidence from Hungarian*. Basingstoke: Palgrave Macmillan.

Wertheimer, M.: Laws of Organisation in Perceptual Form (1923). Reprinted in Ellis, W.D. (ed.): *A Source Book of Gestalt Psychology*, 71-88. London: Routledge and Kegan Paul. (1938)