Effects of linguistic variance on sound-meaning connections in early stages of language acquisition

Eeva Koponen, Lisa Gustafsson and Francisco Lacerda
Stockholm University, Sweden
E-mail: eevak@ling.su.se, lisag@ling.su.se, frasse@ling.su.se

ABSTRACT

To explore the processes underlying early sound-meaning connections an artificial language was created and used in a series of infant speech perception studies. The subjects were tested using the Visual Preference Procedure. The subjects' responses were quantified in terms of looking times towards objects shown during the audio-visual exposure. Exposure to speech materials with large variance seemed to curtail the subjects' ability to establish stable sound-meaning connections. However, reducing the linguistic variance led to successful sound-meaning connections. These results indicate that linguistic variance is one of the primary determinants of sound-meaning connections for 1-year old subjects. The paper will discuss how structural differences in natural language settings may account for the infant's performance on word learning.

1. INTRODUCTION

Guttentag [1] suggests that the initial structuring of sensory input is an automatic process such that we simply connect what we hear with what we experience. More specifically, when coherent visual and auditory stimuli are stored they will eventually become associated [2]. Due to memory decay, less frequent sensory inputs are filtered out at the expense of more frequent ones [3], [4].

In the beginning of the ontogenetic development infants seem to have a limited working memory that presumably is able to process only simple linguistic structures [5]. But lack of variation in the linguistic input may cause the infant to make wrong generalizations. On the other hand, too much variation in the input may slow down the learning process until enough data are gathered to make correct generalizations [5]. With time, a more developed memory capacity will be able to handle larger variance in the input resulting in more economical processing. In evolutionary terms, emergence of grammatical structure can be seen as a result of processing an increasing amount of information. In this view syntax emerges under the pressure of expanding vocabulary [6].

Statistical learning mechanisms, or mechanisms which function like neural networks, have proved to be useful in segmentation tasks on synthetically produced syllables [7], [8]. In the present study we attempted to go beyond the acoustic signal per se to find out to what extent infants manage to extract words from continuous natural speech streams associating novel target-words with arbitrary objects. The hypothesis was that exposure to a nonsense language, characterized by the typical 'repetitive' structure of infant-directed speech (IDS: high target-word frequency rate, target-words in phrase final position), would be helpful in establishing correlations between the signal (speech sounds) and another co-occurring (visual) sensory input. Thus we assumed that statistical regularities conveyed by phonotactic and morphotactic constraints, part-of-speech (POS) and syntax may be helpful in making sound-meaning connections even for previously unheard, but well-formed utterances.

The reason for using an artificial language as a research tool in this study was thus to simulate a learning situation where the semantic content of words is arbitrary, and to control for subjects previous exposure. These artificial phrases, which were read aloud by a human speaker, were thus used to investigate infants' sensitivity to linguistic stimuli.

Two test conditions were created: one control condition in which the structure of the artificial phrases was based on the structure of IDS, and one experimental condition in which the structure of the artificial phrases was manipulated to contain less linguistic variance as compared to the structure of IDS.

2. METHOD

The Visual Preference Procedure – a version of Preferential Listening Procedure [9] – was used. The audio-visual exposure consisted of 3 minutes long films. The objects (puppets) in the films were presented visually and auditorily with corresponding nonsense target-words embedded in nonsense phrases. A video camera recorded the infant watching the film. The infant's looking behavior was later analyzed frame-by-frame.

2.1 SPEECH MATERIALS

The speech materials were based on a description of phonotactics, morphotactics, and syntax of Swedish IDS. A 20 minutes long recording of a mother interacting with her 3-month old infant [10] gave estimates of phoneme, morpheme and target-word frequencies, as well as typical proportion of declaratives, interrogatives and imperatives in Swedish IDS. The data collected was orthographically transcribed forming a mini corpus. As compared to Swedish adult directed speech, the word initial clusters in the mini corpus were rather simple, there were less morphological derivations and the phrases were short and characterized by lexical repetitions.

Context-free word rules were written to capture syllable structures of words corresponding to different POS in the infant-directed speech sample. Nonsense nouns (e g. **bummen*), pronouns (e g. **bu*), auxiliary verbs (e g. **fur*), verbs (e g. **skrett*), etc. were then randomly generated according to these rules. The nonsense words further corresponded to different structural constituents in declarative, interrogative and imperative phrases. Context-free phrase rules were written to capture these syntactical regularities. Finally nonsense phrases were randomly generated according to these rules (e g. **Fur bu skrett bummen?,* Swe. *Har du sett nallen?,* Eng. *Have you seen the teddy bear?*). To attain IDS typical modifications, such as frequent prosodic repetitions and expanded intonation contours, the phrases were read aloud by a human speaker.

2.2 DESIGN

Two conditions were created – one film with high linguistic variance (Svensiska 1) and another with low linguistic variance (Svensiska 2). To control for possible effects of the presentation order of visual materials another low variance film (Svensiska 3), consisting of exactly the same speech materials as in Svensiska 2, was created.

The high linguistic variance condition was characterized by many pronouns referring to objects, as well as target-words/pronouns occurring in other than phrase final positions. In contrast, the low linguistic variance condition was characterized by frequent target-word repetitions, the target-words always occurring in phrase final position. More specifically, in Svensiska 1 the probability for a word to be a target-word was low (0.11). About 70% of these target-words were in phrase final position and about 30% in other than phrase final position. In Svensiska 2 (and 3) the probability for a word to be a target-word was high (0.17). These target-words were always in phrase final position. Consequently, the probability for a phrase to contain target-word in phrase final position was low (0.71) in Svensiska 1, and high (1.0) in Svensiska 2 (and 3) (see Table 1).

	High variance condition	Low variance condition
Film	Svensiska 1	Svensiska 2 & Svensiska 3
Target-word probability (out of total words)	0.11	0.17
Probability for target-word to be in phrase final position	0.71	1.0

Table 1: The probability for a word to be a target-word out of total words, and the probability for target-word to be in phrase final position in the high variance condition (Svensiska 1) are shown in the left column. The corresponding probabilities in the low variance condition (Svensiska 2 & 3) are shown in the right column.

In sum, Svensiska 1 strictly mimicked the structure of the natural IDS dyad. Instead of systematic presentations of target-words, the materials also contained pronouns and context dependent references, whereas the linguistic variance in Svensiska 2 (and 3) was reduced. Exposure to systematic target-words in phrase final position in Svensiska 2 (and 3) was assumed to facilitate the subjects' establishing of sound-meaning connections during the experiment's limited amount of time.

2.3 VISUAL MATERIALS

Each film (Svensiska 1, 2, & 3) consisted of four phases:

- Phase 1) showed a split-screen of two puppets: one to the left, the other to the right. Since the infants' initial visual bias (spontaneous preference) towards future target was measured during this baseline phase, the audio of the phase was silent. Phase 1 lasted for 30 seconds.
- Phase 2) showed one of the puppets acting in the middle of the screen. The audio referred to the puppet with a target-word embedded in nonsense phrases. The infants' looking time was taken as a measure of attention to this first puppet. This first exposure phase lasted for 60 seconds.
- Phase 3) showed the other puppet acting in the middle of the screen. The audio referred to the puppet this time with a new target-word embedded in similar nonsense phrases as in Phase 2. The infants' total looking time was taken as a measure of attention to this second puppet. Also this second exposure phase lasted for 60 seconds.
- Phase 4) showed a split-screen of the two puppets again: one to the left, the other to the right. The audio referred to one of the puppets with a target-word embedded in nonsense questions. The infants' total looking time at the target-object, as compared to bias towards future target in Phase 1, was taken as a measure of changed preference (Test-bias gain) indicating a sound-meaning connection. Phase 4 lasted for 30 seconds.

Svensiska 1, 2 and 3 differed regarding the presentation order (Phase 2 and 3) of the puppets.

2.4 SUBJECTS

The subjects were randomly selected from the National Swedish address register (SPAR) on the basis of age. A total of 144 subjects participated in the study (age range 102-604 days, mean age 360 days) (see Figure 1). Three subjects were excluded due to interrupted recordings (infant crying or not wanting to look at the film).

The remaining 141 subjects were randomly assigned to watch one, two or all three of the films: 23 to Svensiska 1 (age range 235-604 days, mean age 399 days), 118 to Svensiska 2 and 3 (age range 102-518 days, mean age 352 days). The parents participated voluntarily and were not paid for their participation.

2.5 PROCEDURE

A video camera recorded a close-up image of the infant

watching the film. The infant was seated on the parent's lap. To reduce the risk of interacting with the infant, the parent listened to music through soundproof headphones during the whole procedure. Each infant's recording was analyzed frame-by-frame (with precision 0.04 sec) as eye movements to left-, right-, front-, or off-screen.

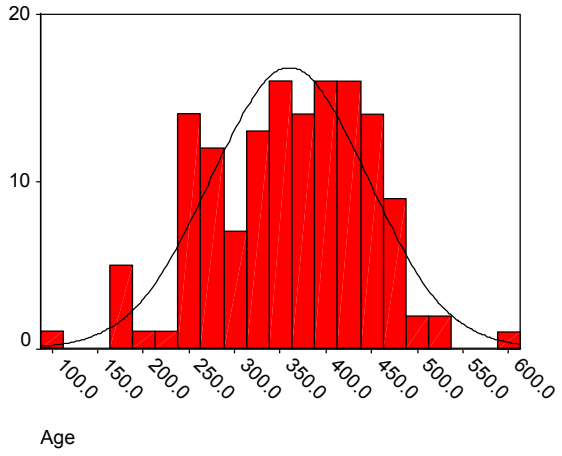

Figure 1: The subjects' age range varied in between 102-604 days (x-axis), mean age 360 days. A total of 144 subjects participated in the study (y-axis: subjects per age).

3. RESULTS

3.1 CHANGED PREFERENCE

The subjects in the low variance condition (Svensiska 2, and 3, N=118) showed increased mean looking preference at the correct object (longer Post-exposure target dominance than Bias towards future target), and the subjects in the high variance condition (Svensiska 1, N=23) showed decreased mean looking preference at the correct object (shorter Post-exposure target dominance than Bias towards future target). A repeated-measures ANOVA revealed significant interaction between the Test-bias gain (Post-exposure target dominance relative to Bias towards future target) and Condition (Svensiska 2, 3 vs. Svensiska 1) ($F(1,139) = 9.734$; $p<0.002$).

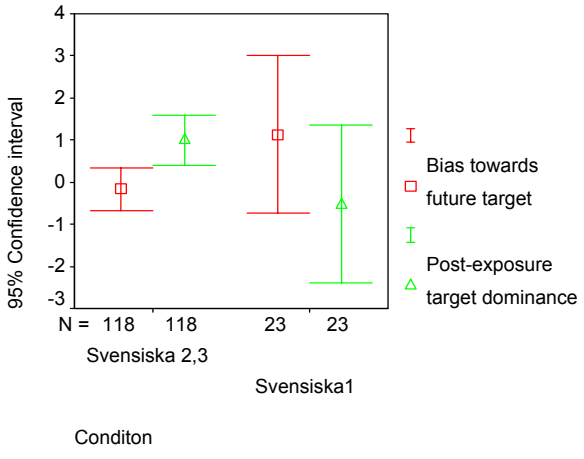

Figure 2: Mean looking times (Phase 1) at the future target, and mean looking times (Phase 4) at the target, in the low variance condition (Svensiska 2 & 3, N=118), and in the high variance condition (Svensiska 1, N=23).

3.2 EXPOSURE TIME

The results showed that an increased exposure (looking time) to the puppets (during Phase 2 and 3) was weakly correlated with the subjects' changed preference: the longer the exposure to the puppets, the greater Test-bias gain (see the left dashed line in the middle of Figure 3). The Test-bias gain was calculated as the net difference between looking time at the target in Phase 4 (after presentation of the puppets) and initial bias (Bias towards future target) in Phase 1.

3.3 AGE

The results showed that the subjects' age was not clearly correlated with the Test-bias gain (see the right dashed line in the middle of Figure 3).

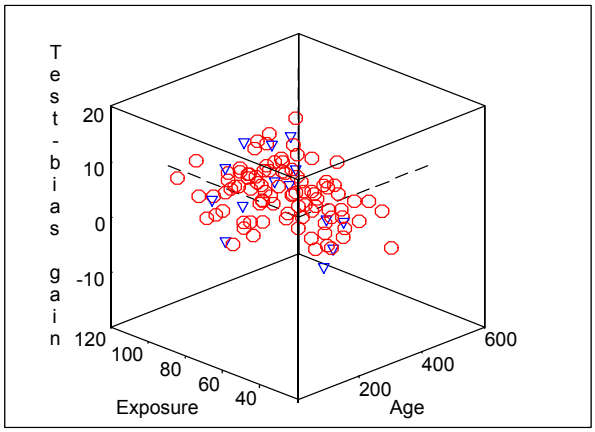

Figure 3: Exposure time to the puppets (during Phase 2 and 3) is shown on the X-axis (range 20-120 seconds), the age of the subjects is shown on the Z-axis (range 100-600 days), and the subjects' changed preference (Test-bias gain) is shown on the Y-axis (range -20 to +20 seconds). Longer exposure time to the puppets was weakly correlated with greater Test-bias gain. The age of the subjects' was not clearly correlated with the Test-bias gain. The subjects in Svensiska 1 are marked with triangles, and subjects in Svensiska 2 &3 are marked with circles.

4. SUMMARY AND DISCUSSION

Recent work on infant speech perception has shown that concrete sounds are learned along circumstantial information about time and space [1]. Phonological, morphological, and syntactical constraints in the speech signal form statistical regularities [7], [8]. Establishment of audio-visual contingencies is, along with frequent visual input, benefited by these regularities [2]. Initially the limited working memory reduces the search space so that infants only need to entertain a small number of hypotheses about the world [11]. However, the dynamic nature of the learning process makes continuous modification of the infant's hypothesis necessary as he/she is exposed to new information.

The aim of the present paper was to explore to what extent infants manage to extract and associate novel target-words to novel objects. Low target-word frequency rate, and varying target-word phrase position in the high variance

condition were strictly based on the structure of the natural IDS dyad that went on for 20 minutes. Thus the phrases in the high variance condition contained pronouns and context dependent references. The high target-word frequency, and constant target-word phrase final position in the low variance condition compressed the structure of the natural IDS dyad to better fit the experimental setting that went on for 3 minutes.

The results showed that the subjects in the low variance condition (Svensiska 2 & 3) had longer looking times at the target-object (Phase 4), as compared to Bias towards future target (Phase 1). This supports the notion that the nonsense phrases in the low variance condition could be associated with the objects just like semantically meaningful phrases do. Thus the infants' establishment of sound-meaning connections was presumably supported by this low variance structure.

The shorter looking times at the target-object (Phase 4), as compared to Bias towards future target (Phase 1), in the high variance condition (Svensiska 1) may indicate the infants' loss of interest as a consequence of too much variance during the exposure to the objects (Phase 2 and 3). As can be seen in Figure 2, the Bias towards future target in the high variance condition, probably due to more attractive puppets, was higher than the Bias towards future target in the low variance condition. This indicates that the infants in the high variance condition were indeed first interested in looking at the objects but their interest was apparently lost, perhaps because high variance exposure did not provide coherent enough information.

The subjects' age range in this study was quite large (102-604 days). However age did not seem to be correlated with the Test-bias gain (the net difference between looking time at the target in Phase 4 and initial bias towards future target in Phase 1). When using the Test-bias gain as a measure of infants establishment of sound-meaning connections, one also need to take the exposure (looking time) to the objects (Phase 2, and 3) into consideration. The results showed that an increased exposure to the puppets was weakly correlated with the Test-bias gain.

It seems that the infants in this experiment could take advantage of grammar. They were capable of recognizing word stems embedded in derivated, inflected, and compounded nonsense nouns. The statistics inherent in the structure of phrases were, due to naturally produced nonsense phrases, supplemented by pauses and intonational patterns. In natural settings infants presumably benefit from syntactic and prosodic statistical information when learning different aspects of language. New experiments will be conducted in the future to test infants' ability to establish sound-meaning connections between verbs and actions.

ACKNOWLEDGEMENTS

Swedish Research Council (Grant VR 421-2001-4876), Faculty of Humanities, and ERICSSON supported this research.

REFERENCES

[1] R. Guttentag, "Children's associative learning: Automatic and deliberate encoding of meaningful associations", *American Journal of Psychology*, **108**, pp. 99-114, 1995.

[2] F. Lacerda, "Phonology: An emergent consequence of memory constraints and sensory input", *Reading and writing: An Interdisciplinary Journal*, **16**, pp. 41-59, 2003.

[3] F. Lacerda and Lindblom, B., "Modelling the early stages of language acquisition", in *Cross-linguistic studies of dyslexia and early language development*, Å. Olofsson, S. Strömqvist, Ed., pp. 14-33. COST A8, European Commission, 1997.

[4] F. Lacerda and Lindblom, B., "On the perceptual consequences of the Tallal transform: Some observations from experiments on language-impaired and normal children", in *Basic neural mechanisms in cognition and language – with special reference to phonological problems in dyslexia*, C. von Euler et. al, Ed., Wenner-Gren Foundations and Rodin Remediation Academy, 1998.

[5] J. Elman, "The Emergence of Language: A Conspiracy Theory", in *The emergence of Language*, B. MacWhinney, Ed., pp. 1-27. Erlbaum: N.J., 1999.

[6] M. Nowak, J. Plotkin and V. Jensen, "The evolution of syntactic communication", *Nature,* **404**, pp. 495-498, 1999.

[7] G. Altman and Z. Dienes, "Rule learning by Seven-Month-Old Infants and Neural Networks", *Science*, **284**, 875, 1995.

[8] J. Saffran, R. Aslin And E. Newport, "Statistical learning by 8-month-old infants", *Science* **274**, pp. 1926-1928, 1996.

[9] A. Fernald, "Four-month-old infants prefer to listen to motherese", *Infant Behavior and Development* **8**, pp. 181-195, 1985.

[10] U. Sundberg, "Mother tongue – Phonetic aspects of infant-directed speech", Ph.D. thesis, *PERILUS XXI*, Stockholm University, 1998.

[11] J. Elman, "Learning and development in neural networks: the importance of starting small", *Cognition* **48**, pp. 71-99, 1993.