Major parts-of-speech in child language – division in open and close class words

E. Klintfors, F. Lacerda and U. Sundberg
Department of Linguistics, Stockholm University, Stockholm

Abstract

The purpose of this study was to assess relations between major parts-of-speech in 14-to 43-months-old infants. Therefore a division in open class and close class words was made. Open class words consist of nouns, verbs and adjectives, while the group of close class words is mainly constituted of grammatical words such as conjunctions, prepositions and adverbs. The data was collected using the Swedish Early Communicative Development Inventory, a version of the MacArthur Communicative Development Inventory. The number of open and close class words was estimated by summarizing items from diverse semantic categories. The study was performed as a mixture of longitudinal and cross-sectional data based on 28 completed forms. The results showed that while the total number of items in the children's vocabularies grew as the child got older; the proportional division in open vs. close class words – proximally 90-10% – was unchanged.

Introduction

This study is performed within the multidisciplinary research project: Modeling Interactive Language Learning[1] (MILLE, supported by the Bank of Sweden Tercentenary Foundation). The goal of the project is to study how general purpose mechanisms may lead to emergence of linguistic structure (*e.g.* words) under the pressure of exposure to the ambient language. The human subject part of the project use data from infant speech perception and production experiments and from adult-infant interaction. The non-human animal part of the project use data from gerbil discrimination and generalization experiments on natural speech stimuli. And finally, within the modeling part of the project mathematical models simulating infants' and animals' performances are implemented. In these models the balance between variance in the input and the formation of phonological-like categories under the pressure of different amounts of available memory representation space are of interest.

The aim of the current study is to explore the major parts-of-speech in child language. Therefore an analysis of questionnaire data based on parental reports of their infants' communicative skills regarding open and close class words was carried out.

Background

The partition in words that belong to the so called *open class* and those that belong to *close class* is a basic division in major parts-of-speech. The open class is "open" in the sense that there is no upper limit for how many units the class may contain, while the close class has relatively few members. The open and close class words also tend to have different functions in the language: the open class words often carry contents, while the close class words modify the relations of the semantically loaded content words.

Why would children pay attention to open class words? Children, as well as adults, look for meaning in what they see and hear. Therefore, the areas of interest and the cognitive development of the child are naturally factors that constrain what is learned first. Close class words seldom refer to something concrete that can be pointed out in the physical world in the way open class words do (Strömqvist, 2003). Also, close class words are not expected to be learned until the child has reached certain grammatical maturity (Håkansson, 1998). Perceptual prominence and frequency are other factors that influence what is learned first (Strömqvist, 1997). Prosodic features such as length and stress make some content words more salient than others. Also, if a word occurs

[1] A collaboration between Department of Linguistics, Stockholm University (SU, Sweden), Department of Psychology, Carnegie Mellon University (CMU, USA), and Department of Speech, Music and Hearing, Royal Institute of Technology (KTH, Sweden).

more often in the language input of the child, it is easier to recognize.

Estimations of children's use of open *vs.* close class words may be based on appreciations of *types* and *occurrences*. For example, in a longitudinal study on four Swedish children and their parents it was shown that the 20 most frequent *types* of words stand for approximately 35-45% of all the word *occurrences* in child language, as well as in adult's speech directed towards children (Strömqvist, 1997). And even more notably, there were almost none open class words among these 20 most frequent words in child language or in child-directed speech (CDS) in the Swedish material. On the contrary, close class words such as *de, du, va, e, ja, den, å, så* constituted the most common word forms. These word forms were most often unstressed and phonologically/phonetically reduced (*e.g.* the words were monosyllabic, and the vowels were centralized). Nevertheless, it should be mentioned that the transcriptions used were not disambiguated in the sense that one sound might stand for much more than the child is able to articulate. For example, the frequent *e* might be generalized to signify *är* (eng. *is*), *det* (eng. *it/that*), *ner* (eng. *down*) *etc.*

In the current study, the questionnaires based on parental reports prompted for words *types* produced by the child. The use of words was differentiated by whether the word in question was used "occasionally" or "often" by the child, but no estimations of number of word *occurrences* were made. Therefore the materials used in the current study allow only for comparison of *types* of words used.

Based on the earlier study by Strömqvist we should thus expect our data to show large and maybe growing proportion of close class words. For example, the proportion of open *vs.* close class words measured at three different time points, corresponding to growing vocabulary sizes, could progress as follows: 90-10%, 80-20%, 70-30% *etc.* But on the other hand, the typically limited amount of close class words in languages should be reflected in the sample and therefore our data should – irrespective of the child's vocabulary size – reveal large and stable proportion of open class words as compared to close class words, measured at different time points corresponding to growing vocabulary sizes (*e.g.* 90-10%, 90-10%, 90-10% *etc.*).

Eriksson and Berglund (1995) indicate that SECDI can to certain to extent to be used for screening purposes to detect and follow up children who show tendencies of delayed or atypical language development. The current study is a step in the direction for finding reference data for typical development of open *vs.* close class words. Atypical development of close class words might thus give information on potentially deviant grammatical development.

Method

The Swedish Early Communicative Development Inventory (SECDI) based on parental reports exists in two versions, one version on words & gestures for 8-to 16-months-old children and the other version on words & sentences for 16-to-28-months-old children. In this study the latter version, divided in checklists of 711 words belonging to 21 semantic categories, was used. The inventory may be used to estimate receptive and productive vocabulary, use of gestures and grammar, maximal length of utterance, as well as pragmatic abilities (Eriksson & Berglund, 1995).

Subjects

The subjects were 24 Swedish children (13 girls, and 11 boys, age range 6.1- to 20.6-months by the start point of the project) randomly selected from the National Swedish address register (SPAR). Swedish was the primary language spoken in all the families with the exception of two mothers who primarily spoke French and Russian respectively. The parents of the subjects were not paid to participate in the study. Children who only participated during the first part of the collection of longitudinal data (they had only filled in the version of SECDI for 8-to 16-months-old children) were excluded from the current study resulting in 28 completed forms filled by 17 children (10 girls, 7 boys, age range 14- to 43-months at the time point of data collection). The data collected was a mixture of longitudinal and cross-sectional data as follows: 1 child completed 4 forms, 1 child completed 3 forms, 6 children completed 2 forms, and 9 children completed 1 form.

Materials

To estimate the number of open class words the sections through A2 to A12, as well as A14 and A15 were included. The semantic categories of these sections are listed in Table 1. Section A1-Sound effects/animal sounds (*e.g. mjau*) and A13-Games/routines (*e.g. god natt*, eng. *good night*) were not considered as representative

open class words and were therefore excluded from the analysis. The sections A16-A21 constituted the group of close class words belonging to the semantic categories listed in Table 2.

Table 1. The semantic categories included for estimation of number of open class words.

Section	Semantic category	Examples of words
A2	Animals (real/toys)	*anka* (eng. *duck*)
A3	Vehicles (real/toys)	*bil* (eng. *car*)
A4	Toys	*boll* (eng. *ball*)
A5	Food and beverage	*apelsin* (eng. *orange*)
A6	Clothes	*jacka* (eng. *jacket*)
A7	Body parts	*mun* (eng. *mouth*)
A8	Small objects/things	*blomma* (eng. *flower*)
A9	Furniture and rooms	*badkar* (eng. *bathtub*)
A10	Objects outdoors	*gata* (eng. *street*)
A11	Places to go	*affär* (eng. *store*)
A12	People	*flicka* (eng. *girl*)
A14	Actions	*arbeta* (eng. *work*)
A15	Adjectives	*arg* (eng. *angry*)

Table 2. The semantic categories included for estimation of number of close class words.

Section	Semantic category	Examples of words
A16	Pronouns	*de* (eng. *they*)
A17	Time expressions	*dag* (eng. *day*)
A18	Prepositions/location	*bakom* (eng. *behind*)
A19	Amount and articles	*alla* (eng. *everybody*)
A20	Auxiliary verbs	*ha* (eng. *have*)
A21	Connectors/questions	*och* (eng. *and*)

Procedure

The materials were collected 2004-2007 by members of the Development group, Phonetic laboratory, Stockholm University. The subjects and their parents visited the lab approximately once/month. Each visit started off with an eye-tracking session to explore specific speech perception research questions, and then a video recording (app. 15-20 minutes) of adult-infant interaction was made. Towards the end of the visit, one of the experimenters entered the studio and filled the questionnaire based on parental information while the parent was playing with the child. Occasionally, if the parent had to leave the lab immediately after the recording session, she/he returned the questionnaire to the lab within about one week (Klintfors, Lacerda, Sundberg, 2007).

Results

The results based on 28 completed forms showed that the child with the smallest vocabulary (4 open class words) had yet not started to use words from close class. The child who produced the most of the open class words (564 open class words) had developed her/his use of close class words into 109 close class words.

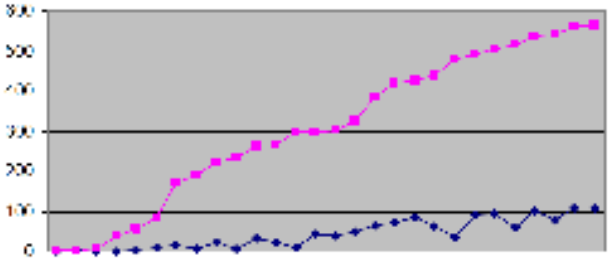

Figure 1. Total number of open class and close class words produced per each completed form. Number of open class words (the light line) and close class words (the dark line) – shown on the y-axis are plotted for each completed form – listed on the x-axis.

When a child knows approximately 100 open class words, she/he knows about 10 close class words – in other words the close class words constitute 10% of the total vocabulary (Figure 1). And further, when a child knows about 300 open class words, she/he knows about 35 close words – that is the close class words constitute 12% of the total vocabulary. And finally, when a child knows approximately 600 open class words, she/he knows about 100 close class words corresponding to 17% of the total vocabulary.

Discussion

The results showed that children's vocabularies initially contain proportionally more open class words as compared to close class words. Thereafter, the larger the vocabulary size, the bigger proportion of it is devoted for close class words. The proportion of open *vs.* close class words corresponding to total vocabulary size of 100, 300, and 600 words, was as follows: 90-10%, 88-12%, 83-17%.

Children might pay more attention to open class words since content words are typically stressed and more prominent (*e.g.* the vowel space of content words is expanded) in CDS (Kuhl et al., 1997; van de Weijer, 1998). Fur-

ther, the open class words often refer to concrete objects in the physical world and might therefore be learned earlier (Gentner & Boroditsky, 2001). Nor are children expected to use close class words until they have reached certain grammatical maturity (Håkansson, 1998).

The youngest subjects in the current study were 1.2-years old and some of the forms completed early on – with vocabularies < 50 words – did not contain any close class words. Shortly thereafter – for vocabularies > 50 words, all the children can be assumed to have reached grammatical maturity. The current study does thus not reveal the exact time point for starting to use close class words. Nevertheless, the age group of the current study ranged between 1.2-years to 3.6-years and likely captured the time point for onset of word spurt. The onset of word spurt has been documented to take place sometime between the end of the first and the end of the third year of life (Bates et al., 1994). Therefore, the proportional increase of close class words being almost twice as large (17%) for vocabulary size of 300 to 600 words, as compared to vocabulary size from 100 to 300 words (10%) is not surprising.

One reason for expecting close class words to later enter the children's vocabularies is that children might have more difficult to understand the abstract meaning of close class words. But closer inspection of the results shows that children start to use close class words although the size of their vocabularies is still relatively small. For example, one of the subjects showed at one occasion to have one close class word and five open class words. But a question to be asked next is how close class words are used in child language. That is, has the child understood the abstract functions of the words used? It is reasonable that children use close class words to express other functions than the original function of the word in question. For example, the word *upp* (eng. *up*) might not be the understood as the abstract content of the preposition up, but instead used to refer to the action *lyft mig upp* (eng. *lift med up*). Using the particle of a verb and omitting the verb referring to the action is typical in child language (Håkansson, 1998). Thus, the close class words are often phonotactically less complex (compare *upp* to *lyfta*) and therefore likely more available for the child. But the use of the word *per se* does not indicate that the child has understood the grammatical role of the close class words in the language. The close class words used by the 14- to 43-month-old children in the current study were Pronouns, Time expressions, prepositions/words for spatial locations, word for Amount and articles, Auxiliary verbs, Connectors and question words. It may thus be speculated that the children in the current study have started to perceive and explore the grammatical status of the close class words.

Acknowledgements

Research supported by The Bank of Sweden Tercentenary Foundation, and European Commission. We thank Ingrid Broomé, Andrea Dahlman, Liz Hultby, Ingrid Rådholm, and Amanda Thorell for data analysis within their B-level term paper in Logopedics.

References

Bates, e., Marchman, V., Thal, D., Fenson, L., Dale, P., Reilly, J., Hartung, J. (1994) Developmental and stylistic variation in the composition of early vocabulary. Journal of Child Language 21, 85-123.

Eriksson, M. and Berglund, E. (1995) Instruments, scoring manual and percentile levels of the Swedish Early Communicative Development Inventory, SECDI, FoU-nämnden. Högskolan i Gävle.

Gentner, D. and Boroditsky, L. (2001) Individuation, relativity, and early word learning. In Bowerman, M. and Levinson, S. (eds) Language acquisition and conceptual development, 215-256. Cambridge University Press, UK.

Håkansson, G. (1998). Språkinlärning hos barn. Studentlitteratur.

Klintfors, E., Lacerda, F., and Sundberg, U. (2007) Estimates of Infants' vocabulary composition and the role of adult-instructions for early word-learning. Proceedings of Fonetik 2007, TMH-QPSR (Stockholm, Sweden) 50, 69-72.

Kuhl, K. P., Andruski, J. E., Chistovich, I. A, Chistovich, L. A., Kozhevnikova, E. V., Ryskina V. L., Stolyarova, E. I., Sundberg, U. and Lacerda, F. (1997) Cross-language analysis of phonetic units in language addressed to infants. Science, 277, 684-686.

Strömqvist, S. (1997) Om tidig morfologisk utveckling. In Söderberg, R. (ed) Från joller till läsning och skrivning, 61-80. Gleerups.

Strömqvist, S. (2003) Barns tidiga språkutveckling. In Bjar, L. and Liberg, S. (eds) Barn utvecklar sitt språk, 57-77. Studentlitteratur.

Weijer, J. van de (1998) Language input for word discovery. Ph.D thesis. Max Planck Series in Psycholinguistics 9.