Mowgli in the jungle of words: Comprehension and later lexical development

Eef Ameel (eef.ameel@psy.kuleuven.be) and Gert Storms (gert.storms@psy.kuleuven.be)
Department of Psychology, University of Leuven, 3000 Leuven, Belgium

Barbara C. Malt (barbara.malt@lehigh.edu)
Department of Psychology, Lehigh University, Bethlehem, PA 18015 USA

Abstract

Production data suggest that meanings of common nouns continue to change well past the early years of language acquisition (Andersen, 1975; Ameel, Malt, & Storms, 2008). Here we used two comprehension tasks to further evaluate the nature of later lexical learning. In a name applicability task, seven- to 13-year old Dutch-speaking children judged whether each of three names applied to common household containers. In a typicality judgment task, participants judged how good an example the containers were of the three names. Both tasks revealed continued evolution of word use up to age 13. However, all names were overextended by the children, in contrast to the production data in which both over- and underextension were found. Moreover, the child lexical categories showed considerable overlap, indicating strong inter-category relatedness. With age these overextensions as well as the overlap gradually disappeared and the lexical categories became more distinct over time.

Keywords: word learning; later lexical development; lexical categories; typicality judgments

Introduction

The speed with which young children add words to their vocabulary seems to suggest that word learning is a relatively easy process. By age two, they are able to produce about 600 words. In the period from age two to six, they are estimated to acquire around 14.000 words, at a rate of ten words a day (Carey, 1978). The meanings assigned to these words, however, only partially correspond to the meanings adults attach to the words. The incomplete word meanings inevitably cause children to commit errors. Some words are used too broadly, a phenomenon called overextension, for example, when the word *cow* is applied to all four-legged animals. A category is underextended when its name is assigned to too few items, for example, when the word *cow* is only applied to the child's cuddly animal.

These errors disappear when the full conventional meaning is grasped, but this may take months, or even years. A number of studies have shown learning periods for verb meanings extending to age 8 or 9 (e.g., for *pour* and *fill*, Gropen, Pinker, Hollander, & Goldberg, 1991; see also Bowerman, 1974; Pye, Loeb, & Pao, 1996). But even for common nouns, the time frame for fully acquiring their meaning and use seems to extend beyond the early years of language acquisition. This reflects the broad and complex extension of common nouns referring to concrete, simple objects. *Ball* for English speakers, for example, often refers to smooth, bouncy, spherical, deflatable things made for play, but it can also refer to non-smooth balls of paper, non-bouncy beanbag balls, non-spherical footballs, non-deflatable baseballs and billiard balls, and things not for play such as balls of yarn, balls of string, and tea balls. A few studies clearly demonstrated the extended learning trajectory for nouns that are used to refer to common household objects.

Andersen (1975), for example, asked English-speaking children aged 3 through 12 to name drinking vessels, and she found that children's naming did not match adult naming until the age of 12. Ameel et al. (2008) had Dutch-speaking Belgian children of ages 5, 8, 10, 12 and 14 name common household containers, which were mostly called *fles*, *bus*, and *pot* by adults (Ameel, Storms, Malt, Sloman, 2005). They found that early use of *fles* and *pot* was overextended and gradually narrowed from 5-year-olds to adults, while the opposite pattern was found for *bus*. Gradually, the children converged toward the adult naming, but even the 14-year-olds still slightly differed in their word use from the adults.

The results of Ameel et al. (2008) were obtained using a production task in which participants named the object in each picture however they chose. A production task offers several advantages as opposed to a comprehension task in which participants are generally asked to make a forced choice decision between two or more options. A first advantage is that it does not constrain responses to be from a set pre-determined by the experimenter to be suitable. Production is also a purer measure of what participants know because they have to come up with a name without having any suggested to them. Given a forced choice between *fles* and *bus* (or any other word pair), participants will pick one regardless of how little knowledge they may have of the words. Finally, production has a certain ecological validity in that outside the laboratory, children and adults frequently select names for objects for purposes of communication, and they do so without having options presented to them.

However, there are also some disadvantages related to production tasks. First, production tasks might be cognitively more demanding than comprehension tasks, since they require not only a sense of familiarity with the presented material but also retrieval from memory. Children have had less language exposure than adults, so even if they have some knowledge of a word and of appropriate word-

object pairings, in a production task they may have greater trouble retrieving a particular word from memory. In such instances, they may resort to use of a more easily retrieved but less appropriate word. Also, they may sometimes successfully retrieve a word from memory in connection with an object to which it is strongly linked but fail to retrieve it in connection with an object to which it is more weakly linked. These retrieval failures could produce a pattern of overextension of some words and underextension of others, even if the child has a more adult-like pattern of knowledge encoded in memory. Second, it has long been noted that children's appropriate responses to words (such as picking up an object named by an adult) often precede their production in development (e.g., Benedict, 1979; Kuczaj, 1982; Rescorla, 1981). Children may have good implicit knowledge of the adult pattern but be unable to fully deploy that knowledge in production.

Children's implicit knowledge can be studied by means of a comprehension task in which participants are asked about the appropriateness of names for an object. For adults, artifacts can often be called by more than one name at the same level of abstraction (e.g., *bottle* or *jug* for a squat glass container with handles), indicating overlapping word meanings and extensions. However, production tasks are generally intended to elicit only a single name for an object which prevents overlap from being revealed. Comprehension tasks that do not require forced choice and allow children to accept as many names as they feel are appropriate for the object is more suited to tackle this job.

The present work uses a comprehension-based methodology to further investigate the artifact word learning process. We ask three questions about what comprehension data show relative to production data about later lexical development: (1) Is the evidence for an extended lexical development period still present in comprehension tasks? (2) If evidence for an extended learning period is not eliminated by comprehension measures, does the specific developmental trajectory for individual words resemble that seen in production? (3) What is the nature of later lexical development with regard to lexical category overlap?

Study 1: Name Acceptability Judgments

The first comprehension-type task was a name acceptability judgment task in which participants judged whether or not offered names were acceptable for a set of stimuli.

Method

Five age groups made name judgments: 20 7-year-olds, 21 9-year-olds, 20 11-year-olds, 20 13-year-olds, and 36 adults. For each object of the 73 storage containers, taken from Ameel et al. (2005), participants had to judge whether it did or did not belong to each of a series of category names (e.g., Is this object a *fles*?). There were two possible answers: yes or no. Participants made typicality judgments for three category names - *fles*, *bus*, and *pot* -which were the most frequently generated names by adults (Ameel et al., 2005).

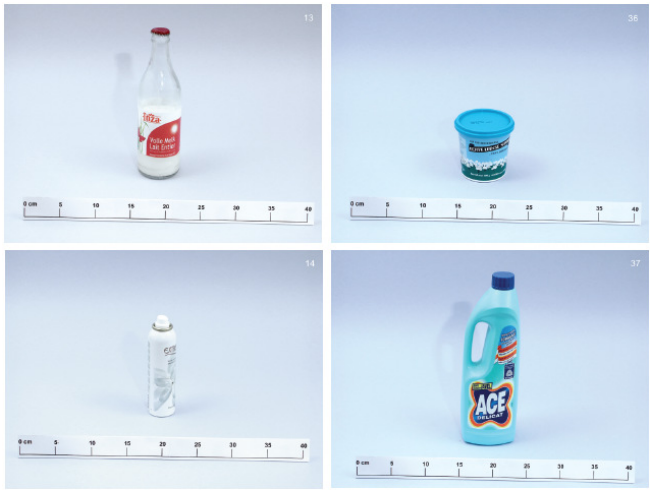

Figure 1: Typical examples of *fles* (upper left), *pot* (upper right), and *bus* (lower left and right).

Figure 1 shows typical examples of each of these names[1]. Participants were queried at two different points in time to maximize the ability to treat each judgment as independent of previous ones. A participant could agree that a particular object could be called *fles* and also agree that it could be called *bus* when asked at a different time. She also could accept just one name, or neither. Stimulus order and category name order were randomized across participants.

Results and Discussion

Reliability of the data for each name and age group was calculated by applying the Spearman-Brown formula to the split-half correlations between the frequencies of "yes" responses. The reliabilities were all very high, varying from .87 to .99, and they increased over age, though not significantly ($\rho = 0.82$, $p = .09$).

To evaluate evidence for an extended lexical development period for artifact names in comprehension tasks, we examined the size of individual categories at each age by calculating the percentage of objects that were accepted for each name by each participant in an age group. The percentages were averaged across participants of the same age group. An ANOVA was performed with two factors: age (5 levels: 7-, 9-, 11-, 13-year-olds and adults) and category name (3 levels: *fles*, *bus* and *pot*). There were significant differences among the age groups ($F(4,411) = 11.92$, $p < .0001$). Linear trend tests were used to test the change in percentages of objects granted a given name over time. All trends were significantly decreasing (*fles*: $F(1,285) = -6.07$, $p < .0001$; *bus*: $F(1,285) = -4.62$, $p < .0001$; *pot*: $F(1,285) = -5.60$, $p < .0001$). Thus, there is evidence for an extended learning trajectory when using comprehension measures rather than production measures.

The production data of Ameel et al. (2008) showed that among the names tested in the current study, *fles* and *pot*

[1] For *bus*, two typical examples are shown, since this word picks out a very heterogeneous category, encompassing a wide variety of objects (see also Ameel et al., 2008).

were used more broadly by the youngest children than by adults and their use narrowed over time, while *bus* was initially used more narrowly and broadened over time. In contrast, here we found that across ages, the names were gradually applied to smaller groups of objects. This trend toward narrowing over time is significant for all names, as indicated by the linear trend tests reported above. Thus, these data contrast with the production data in showing only narrowing of use over time.

Since the preceding analysis showed that the children's lexical categories were uniformly larger than adults', we can deduce that overall, there must be more overlap, or, less differentiation, among the categories for children than for adults. To quantify the overlap, we calculated the mean number of names accepted for each object across ages. A linear trend test of age showed that the mean number of names that each object is assigned to significantly decreased over age ($F(1,7295) = -24.29$, $p < .0001$).

To visualize the changes in overlap, we projected the lexical categories of the youngest children and adults onto a geometrical representation of the stimuli, which was also used in Ameel, Malt, Storms, and Van Assche (2009). The multidimensional scaling (MDS) representation was based on non-linguistic similarity judgments derived from adult sorting data collected by Ameel et al. (2005). Even though children's sorting data were available (Ameel et al., 2008), we decided to use the MDS solution based on adult sorting data to project categories onto because the adult data were more reliable than the child data. In such a geometrical representation, each exemplar of a category is represented by a vector of M coordinates, one for each of the M underlying dimensions. Distances between exemplars reflect similarity relations: The closer objects are located to each other in the multidimensional space, the more similar they are to each other.

Figures 2 and 3 show the two-dimensional MDS representations for the 7-year-olds and the adults. Shaded areas show the sets of exemplars that were accepted as *fles*, *bus*, and *pot*. An object was assigned to a category if at least 65 per cent of the participants judged the name to be appropriate for the object. The figures nicely show that with age, the categories shrink, their overlap decreases and they become more distinct.

In sum, in this study, name appropriateness judgments were used to assess lexical knowledge. For each name, the percentage of assignments decreased over time, from the youngest children to the adults, which is clear evidence for an extended lexical development period, as was found in the production task of Ameel et al. (2008). However, while they found a narrowing pattern for some categories (*fles*, and *pot*) over time and a broadening pattern for others (*bus*), the comprehension task used in here resulted in a narrowing pattern for each category name. Overlap between categories decreased with age, indicating that categories gradually become more distinct.

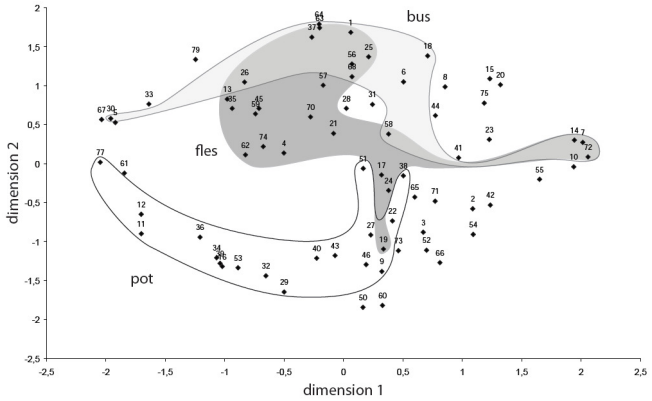

Figure 2: Two-dimensional MDS representation of the 73 stimuli (black diamonds) with the *fles* (dark grey cluster), *bus* (light grey) and *pot* (white) categories of 7-year-olds.

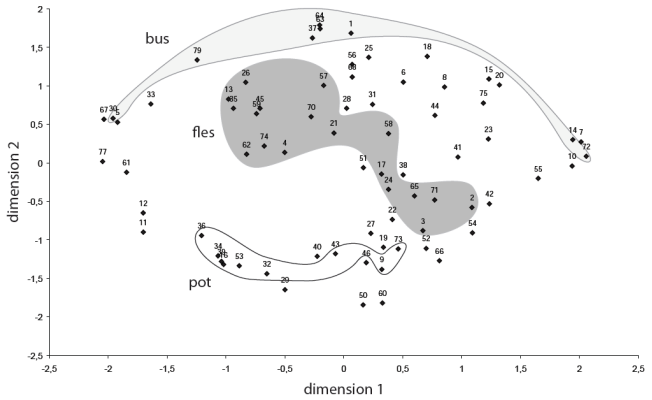

Figure 3: Two-dimensional MDS representation of the 73 stimuli (black diamonds) with the *fles* (dark grey cluster), *bus* (light grey) and *pot* (white) categories of adults.

Study 2: Typicality Judgments

Another way to assess lexical category knowledge using a comprehension measure is through typicality judgments. Some instances of a name are more representative or typical than others (Mervis & Rosch, 1981; Rosch, 1973). For example, cars are more typical examples of things called *vehicle* than submarines are. Typicality judgments allow us to investigate the extended learning trajectory of common nouns (question 1 and 2) and the nature of later lexical development with regard to lexical category overlap (question 3). Furthermore, the last question can be approached in a novel way: by computing the locations of the category centers in a multidimensional space and evaluating whether these typicality-determined centers approach each other or drift further apart from each other across development.

Method

Again, the 73 storage containers were used. These objects were randomly presented three times to 21 children aged 7, 21 children aged 9, and 23 children aged 11, as well as to 28 adults (taken from Ameel et al., 2005). None of the

participants had taken part in Study 1. Each time, they judged how good an example the objects were for one of the three names *fles*, *bus* and *pot*. The order of the names was randomized across participants.

To judge typicality, the participants used a 7-point rating scale, which was accompanied by three schematic faces, each having a different expression: the "frowning" face, located below number 1, corresponded to "very poor" examples of the target category, the "smiling" face, located below number 7, corresponded to "very good" examples of the target category, and the "straight" face, located below number 4, corresponded to "okay" examples of the category name. The schematic faces were similar to the ones used in a study of Bjorklund, Thompson, and Ornstein (1983) who used a 3-point rating scale with only schematic faces. The use of a 7-point rating scale with number was justified by the greater age of our participants. For the children, the test phase was preceded by a practice phase to ensure that children understood the instructions correctly.

To make sure that the children did not make preference judgments instead of typicality judgments (Maridaki-Kassotaki, 1997), the experimenter stressed that the child should not rate how much she liked/disliked the items.

Results and Discussion

Reliability of the typicality data for each category name and each age group was evaluated by applying the Spearman-Brown formula to the split-half correlations. The reliabilities were again very high, varying between .75 and .99, and they significantly increased over age (ρ = 1.00, p <.0001).

To evaluate evidence for an extended lexical development period for artifact names in comprehension tasks, we first calculated the mean typicality rating for each name for each age group. For all names children's ratings were higher than adults'. In an overall ANOVA with two factors: age (4 levels: 7-, 9-, 11-year-olds and adults) and category name (3 levels *fles*, *bus* and *pot*), significant differences were found among the age groups (F(3, 864) = 99.11, p <.0001). For all names, the mean ratings decreased linearly over age, as shown by significant linear trend tests (*fles*: F(1, 864) = -9.24, p <.0001; *bus*: F(1,864) = -8.64, p <.0001; *pot*: F(1,864) = -11.92, p <.0001). These analyses provide more evidence that an extended developmental trajectory for artifact word knowledge exists in comprehension.

As noted earlier, Ameel et al. (2008) found that in production the use of *fles* and *pot* narrowed over time, while *bus* broadened over time. To derive the evolution in naming from the typicality data, we computed for each participant the number of objects receiving a typicality score of 5 or more for each name. These frequencies were subjected to an ANOVA with two factors: age (4 levels: 7-, 9-, 11-year-olds and adults) and category name (3 levels: *fles*, *bus* and *pot*). Only the age effect was significant (F(3, 251) = 37.96, p <.0001), showing that for each name children consider more objects to be good examples of the lexical category than adults do. With age, this discrepancy is reduced. So, again, comprehension and production data yielded contradicting results as far as evolution of lexical category breadth.

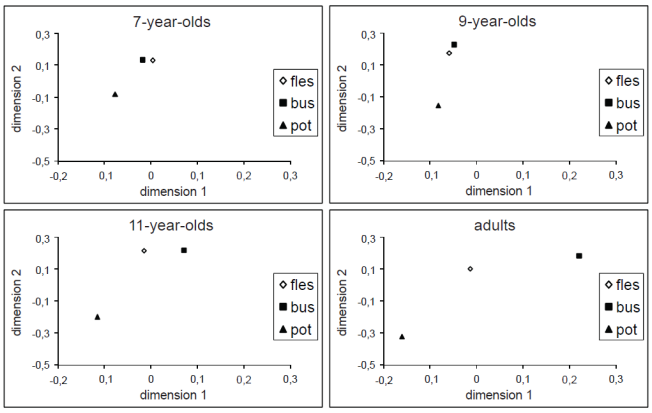

Figure 4: Two-dimensional MDS representations of the category centers of *fles*, *bus* and *pot* for each age group.

The evolution of overlap was investigated by looking at how close the centers of lexical categories are at each age. Given the result from the name applicability study that categories become more distinct over time, we expect that the distances between category centers will increase with age. Increasing distances with age, together with the finding of decreasing category size (see Study 1), implies that overlap between categories decreases. Analogous to the method applied in Ameel et al. (2009), the positions of the category centers for the different age groups were computed across all the objects of a stimulus set, weighed by their mean typicality rating for the name, and imposed on the MDS solutions[2]. This method allows objects rated as more typical for the name to affect the position of the category center to a larger degree than objects rated as less typical for the name. For example, to compute the coordinates in the MDS solution of the category center of *fles* for the 7-year-olds, the coordinates of each object given by the sorting data were multiplied by its typicality rating for *fles* averaged across all 7-year-olds. Next, the weighted coordinates were summed and the coordinates of the weighted category center were calculated as the weighted sum divided by the sum of all typicality ratings. Figure 4 shows, for each age group the weighted prototypes for the three target categories, projected in the 2-dimensional MDS representation. Note that only the weighted prototypes are displayed, not the positions of the individual stimuli, to enhance the clarity of the figures. From Figure 4 it is clear that the distances between each pair of lexical categories (*fles-bus*, *fles-pot*, *bus-pot*) gradually increased over age. The increasing pattern of distances between category centers indicates that the categories gradually become more differentiated over

[2] In Study 1, MDS is used as a visualization technique, in Study 2, MDS is a means to quantify the inter-category relations. We could have used them for both purposes in both studies, however, due to space restrictions, only one purpose was elaborated in each study.

age. In other words, overlap between categories gradually decreases over time. To quantify this change, distances between the weighted category centers, averaged across name pairs (rather than the locations of the individual centers, as in Figures 4) were computed in 2 to 5 dimensional representations, since the choice of the number of underlying psychologically relevant dimensions is not always obvious (see Verheyen, Ameel, & Storms, 2007). For each dimensionality, we found the clear-cut age effect of gradually increasing distances, but distances increased with increasing dimensionality. The age differences in distances between the different category pairs in 2 to 5 dimensions were tested by means of a linear trend test which was significant ($F(1,44) = 65.34$, $p < .0001$). This analysis confirms that on average, lexical categories become more differentiated over time.

In sum, Study 2 assessed lexical knowledge through typicality judgments. As in Study 1, evidence for an extended lexical development period was found, as indicated by the gradually decreasing mean typicality ratings from 7-year-olds to adults. The results suggest that, generally, children's lexical categories are broader than adults in comprehension, consistent with the name acceptability judgments and not entirely consistent with the production data of Ameel et al. (2008), where some categories were found to start narrow and broaden over time. To assess the changes in overlap, distances were calculated between weighted prototypes of the different categories in multidimensional scaling representations. We found that the distances between name pairs increased over age, suggesting that the lexical categories become more distinct over time, or, that categories gradually show less overlap over time.

General Discussion

While past research on the artifact word learning process has been based on production tasks (Andersen, 1975; Ameel et al., 2008), the present paper used two comprehension tasks to further understand later lexical development: a name applicability task and a typicality judgment task. In both comprehension tasks, evidence was still present for an extended lexical development period, as was found with a production task. The percentages of assignments to the names as well as the mean typicality ratings gradually decreased with age and converged upon, respectively, the adult percentages and the adult mean typicality ratings. The specific developmental trajectory for individual words, however, did not completely resemble that seen in production. While Ameel et al. (2009) found narrowing for *fles* and *pot*, and broadening for *bus*, the comprehension data showed only narrowing of use over time. Finally, in contrast to the production data, the comprehension data allowed us to investigate the evolution of overlap between the lexical categories. Overlap gradually decreased over age, as demonstrated by the decreasing number of names assigned to objects and as visualized in the MDS plots (Figure 2 and 3). Further, the distances between the typicality-based category centers increased with age, indicating that the categories gradually became more distinct over time.

The difference in the developmental trajectories between comprehension and production data can be explained in terms of the performance constraints imposed by production, in combination with the evolution of featural knowledge (see Ameel et al., 2008, in which word use was predicted by features at each age). In production, underextension only occurred with names that entered the vocabulary at a later age. The word *bus*, for example, did not appear into the productive vocabulary until the age of 8, suggesting that this word is less familiar, less frequently encountered, and less retrievable than the ones produced earlier (*fles* and *pot*). The 8-year-olds have this word represented with a few adult features plus some irrelevant features that do not help differentiate this word from contrasting words (e.g., "is round", see Appendix C, Ameel et al., 2008). Given this representation, overextension would be expected. However, other more easily retrieved words may out-compete *bus* for production at this age. In this case, especially the word *fles*, which shows considerable overlap with *bus* in comprehension (see Figure 2), takes over a large part of the extension of *bus* in production. In the comprehension task, the incomplete (by adult standards) representation of *bus* makes children excessively liberal in applying the name to objects, resulting in overextension. Features may be added to the representation of *bus* and others deleted or weighted less over time so that its representation becomes more distinct from the representation of *fles* (and of possibly other overlapping categories) over time.

Ameel et al.'s (2008) production study also revealed that children initially attended to different features from adults when naming objects, but gradually learned to attend to the adult set of features and assign them the appropriate weights. This finding is in line with Mervis' (1987) view of lexical development which implies that both over- and underextensions can occur. The findings of our comprehension study – only overextensions –, however, seem to imply that Mervis is wrong. Given the current findings, Mervis may be right about the *process* children go through to arrive at adult representations, which may involve adding relevant features, dropping irrelevant ones, and/or adapting feature weights to arrive at adult featural knowledge. Regarding the initial category *representation*, Mervis may be wrong in one respect. Consistent with her view, naming (both comprehension and production) may be initially controlled by a subset of adult features plus some irrelevant features. The number of features for adults and children will be similar due to the addition of these irrelevant features. However, according to Mervis, these irrelevant features, if too specific and assigned a high feature weight, can produce underextension. But no underextension was found in comprehension, suggesting that the irrelevant features generally are rather too broad and together with the smaller subset of the adult features cause

overextension. Over time, the irrelevant features are replaced by relevant ones, so that children gradually arrive at the correct set of adult features.

It must be acknowledged that comprehension tasks are - like everything else - an indirect measure of actual featural knowledge. Given the finding that children easily overestimate their own capacities (Stipek & MacIver, 1989), it could be that younger children also tend to accept offered words more readily in the name acceptability task, and rate things higher in typicality than adults. So, we have to take into account that more than just featural changes may contribute to developmental changes in patterns of production and patterns of comprehension judgments. However, the patterns of word acceptance make clear that a bias to accept names more generously than adults cannot be the whole story. Even the youngest children are not just saying *yes* to whatever is offered to them too often. The regions of semantic space covered by each adult word and the peculiarities of the irregular shapes of the adult regions are substantially mirrored in the children's choices (see Figures 2 and 3), indicating that their responses are guided by specific word knowledge.

Our data highlight the importance of understanding later lexical development in order to develop a complete view of word learning. Like Mowgli who lives in a jungle full of intertwining lianas and gradually has to disentangle them on his developmental path through the jungle, children's words meanings are initially strongly intertwined. Children gradually have to disentangle them in order to arrive at adult featural knowledge.

Acknowledgments

Eef Ameel is postdoctoral fellow at the Fund for Scientific Research, Flanders. We thank An Van Brusselen, Caroline Marsinelle and Sofie Tijskens for assisting in the data collection.

References

Ameel, E., Malt, B. C., & Storms, G. (2008). Object naming and later lexical development: From baby bottle to beer bottle. *Journal of Memory and Language, 58*, 262-285.

Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language , 60* , 270–290.

Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language, 53*, 60-80.

Andersen, E. S. (1975). Cups and glasses: learning that boundaries are vague. *Child Language, 2*, 79–103.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language, 6*, 183-200.

Bjorklund, D. F., Thompson, B. E., & Ornstein, P. A. (1983). Developmental trends in children's typicality judgments. *Behavior Research Methods and Instrumentation*,*15* , 350–356.

Bowerman, M. (1974). Learning the structure of causative verbs: a study in the relationship of cognitive, semantic and syntactic development. *Papers and Reports on Child Language Development, 8*, 142–178.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65–110). New York: Academic Press.

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991). Syntax and semantics in the acquisition of locative verbs. *Journal of Child Language, 18*(1), 115–151.

Kuczaj, S. A. (1982). Young children's overextensions of object words in comprehension and/or production: Support for a prototype theory of early object word meaning. *First Language, 3*, 93-105.

Maridaki-Kassotaki, K. (1997). Are rating-based procedures reliable for derivation of typicality judgments from children? *Behavior Research Methods and Instrumentation, 29*(3), 376-385.

Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (Ed.), *Concepts and conceptual development: ecological and intellectual factors in categorisation* (pp. 201–233). Cambridge: Cambridge University Press.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89-115.

Pye, C., Loeb, D. F., & Pao, Y. Y. (1996). The acquisition of breaking and cutting. In E. V. Clark (Ed.), *The proceedings of the twenty-seventh annual child language research forum* (pp. 227–236). Stanford: Center for the Study of Language and Information.

Rescorla, L. (1981). Category development in early language. *Journal of Child Language, 8*, 225-238.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). New York: Academic Press.

Stipek, D., & MacIver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development, 60*(3), 521-538.

Verheyen, S., Ameel, E., & Storms, G. (2007). Determining the dimensionality in spatial representations of semantic concepts. *Behavior Research Methods, Instruments, and Computers, 39*, 427-438.