What NN compounding in child language tells us about categorization

Maria Rosenberg
Dept. of Language Studies
Umeå university
SE-901 87 Umeå
SWEDEN
maria.rosenberg@umu.se

Ingmarie Mellenius
Dept. of Language Studies
Umeå university
SE-901 87 Umeå
SWEDEN
ingmarie.mellenius@umu.se

1 Introduction

The present study examines novel NN compounds, produced on line, in Swedish child language, with focus on categorization. Given that NN compounds denote objects, we concentrate on the categories those objects belong to. In that way, our study aims to provide evidence of object categorization in preschool children. Two questions are put forward:

(i) Does perception play a crucial role for the children's coinages?

(ii) In what way do structural and processing views on categorization apply to the data?

Swedish children produce compounds already at age two, reflecting the fact that compounding is a productive word formation device. In short, Swedish compounds are right-headed, written as one word, pronounced with a two-peak-intonation, and can exhibit liaison forms.

2 Theoretical background

Clark (2004) argues that language acquisition builds upon already established conceptual information, which enables the child to categorize objects, relations and events. Children rely mainly on shape as they embark on the mapping of words for objects onto their conceptual categories of objects, but also pay attention to texture, size, sound, motion and function. Even into adulthood, children continue the mapping of unknown linguistic items onto conceptual representations. Young children occasionally form emergent categories, based on non-conventional distinctions (e.g. *ball* for round things). Clark (2004) notes that the pairing of word to object enables the child to perceive similarities between cognitive categories, and allows for alternate perspectives on objects (cf. Waxman and Markow, 1995). "Do their [children's] categories reflect only what their language offers, or do they – must they– make use of other representations too?" (Clark, 2004:472).

Berman (2009) emphasizes that there is a substantial difference in adults' vs. children's lexicons of established compounds, and that children have to grasp inter alia the idea of subcategorization. Clark and Berman claim that "knowledge of the pertinent lexical items, and not the constructions they appear in, is more important for [children's] compounding" (1987:560).

In conceptual development, category structures change with age (Keil and Kelly, 1987). Object categorization allows generalization over properties of objects and of novel category members (Mandler, 2000).

Bornstein and Arterberry (2010) mention two complementing views of categorization: processing and structural. On the processing view, categories are flexible and category membership of objects can vary in different situations (cf. e.g. Jones and Smith, 1993). On the structure view, categories are hierarchically organized taxonomies (cf. e.g. Murphy, 2002). Instead of Rosch's (1978) superordinate-basic-subordinate, levels of category inclusiveness can be ordered in a neutral way, such as L_1 (animal), L_2 (cat, dogs), L_3 (collies, shepherds), L_4 (scotch collies, border collies) (Bornstein and Arterberry, 2010:3).

Whether categorization proceeds from concrete to abstract or the other way around is still under debate. Differentiation theory (e.g. Gibson, 1969) stipulates that the ability to make finer differentiations emerges after broad conceptions are acquired. Likewise, Bornstein and Arterberry (2010) indicate that more inclusive levels of categorization appear before less inclusive ones,

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

and that high perceptual contrasts have precedence over low. Fisher (2011) suggests that at age 3-5, perceptual information is anchored more strongly than conceptual information; cognitive flexibility develops with age.

Yet, according to Smith (1984), preschool children show the ability of both concrete categorization, due to perceptual characteristics, and abstract categorization, leaning on conceptual relationships. Nguyen and Murphy (2003) posit three categorization forms: taxonomic (see above), script and thematic. Script-based categories include objects (e.g. egg, cereal) with the same functional role in a routine event (e.g. eating breakfast). Thematic categories involve objects that usually appear together (e.g. bowl-cereal). Nguyen and Murphy (2003) show that children, aged 3 to7, use taxonomic and script categorization in a flexible way.

3 Data and method

The data consists of 383 spontaneously produced NN compounds from three monolingual Swedish children, aged 1 to 6, collected longitudinally and including contextual information. The children often give an explanation of the intended meaning, e.g. *hundstall* 'dog-stable', 'where dogs live, outside'. Hence, they seem to understand the semantics of their novel compound. We use a strict selection criterion: only non-established compounds in contemporary Swedish are considered.

As a first step to analyze our data, we sort the compounds in two ways: (i) based on N_1; (ii) based on N_2. This is a way of locating items belonging to a same morphological family (cf. Schreuder and Baayen, 1997). As a second step, the data is analyzed according to: (iii) level of inclusiveness; (iv) script; (v) thematicity; (vi) perception (real-world referent or not, high contrasts vs. low). As a third step, other characteristics appearing from the children's compounds are analyzed.

4 Analysis

In the analysis we provide evidence of categorization concerning larger groups of compounds. Below follows some preliminary findings. Note that the compounds can be analyzed according to different parameters and, thus, some of them go into several labels, depending on the parameter taken under account.

4.1 N_1 and N_2 sorting

The sorting of N1 and N2 shows that several nouns reoccur in the children's compounds. 126 N1 of the 383 compounds were either identical or belonging to the same morphological family, such as *morotsvatten* 'carrot-*s*-water and *morötter-macka* 'carrots-sandwich'. With respect to N_2, this number was as high as 143.

The largest morphological family found in our data contains *vatten* 'water'. 12 compounds are attested (7 compounds from one child, whereof 4 have *vatten* as N_1, and 3 as N_2). The two other children used *vatten* in 4 and 1 instances respectively, such as *vattenkaffe* 'water-coffee' or the aforementioned *morotsvatten* 'carrot-s-water'. Other nouns that reoccurred nearly ten times among the innovations of all three children were *bil* 'car' *kläder* 'clothes', *mamma* 'mommy' and *väg* 'road' (cf. 4.6).

It is worth noting that although the same nouns were used in several compounds, they did not always uphold the same relation to the other constituent: *pizzabil* 'pizza-car' was used for a car with a pizza print on it (viz. perceptually), whereas *dimbil* 'fog-car' referred to an imaginative car spraying fog (viz. abstractly).

Overall, the overlap between the same nouns being used in several compounds and as first or second constituent, can be taken as support for Clark's and Berman's (1987) claim (cf. 2) that children use lexical items that they are familiar with in their compounding.

4.2 Level of inclusiveness

As for the level of inclusiveness, the compounds in our data are situated on L_1 (*björkgrej* 'birch-thing'), L_2 (*brödrosta* 'toaster'), L_3 (*äppelsvans* 'apple-tail') or L_4 (*hjärtklackskorna* 'heart-heel-shoes), with L_3 as the predominant level. If we look only at N_1 or N_2 in isolation, they can also correspond to items located at L_1 (*djur* 'animal', L_2 (*björn* 'bear'), or L3 (*äppeljuice* 'apple-juice') in three-psart compounds.

Moreover, there are some compounds in our data containing a taxonomic relation between the constituents: two examples are *ugglafågel* 'owl-bird' and *skinndjur* 'skin-animals'.

4.3 Script-based categories

Entire sets of the compounds can be analyzed as having the same role with respect to a script, in which the compounds fulfill the same part. All three children categorize clothes according to season or weather, as indicated by N_1: *sommar-*

vantar 'summer-gloves', *snöstrumpor* 'snow-stockings' or *vinterficka* 'winter-pocket'.

There are also compounds in our data where N_1 and N_2 participate in the same scripts that concern different types of edibles: *grötmjölk* 'porridge-milk' (eating breakfast") or *pizzahamburgare* 'pizza-hamburger' (eating dinner") or *saftglass* 'syrup-ice cream' (eating dessert).

4.4 Thematic categories

Thematic categories, items with close semantic association based on, e.g., contiguity, are numerous within the compounds. An example is *häxafiskspö* 'witch-fish-wand', where the child aims at a wand used by a witch, but confuses *trollspö* 'magic-wand' with *fiskespö* 'fishing-pole', and than adds the user of the item in question (actually a case of "overcategorization", cf. 4.6).

Several themes are found. One is "sweets", giving rise to numerous compounds, semantically associated or not, such as *silvergodis* 'silver-candy' and *godisstrumpor* 'candy-stockings'.

Most of the thematic categorization found in the children's innovation is abstract and grounded in conceptual information. Furthermore, the thematic relations are mostly of an inherent nature, such as manifested by *djungelträd* 'jungle-tree', rather than temporal, such as *fotbollsplanet* 'football-planet'.

4.5 Perception

Compounds categorized according to Shape are attested, such as *R-paprika* 'a piece of paprika that looks like a R', or *mössaboll* 'hat crumpled into the shape of a ball'. Shape may concern either the head or the non-head of the compound. Texture is involved in many of the children's compounds, such as: *pälsmatta* 'fur-carpet'. Prints are also a frequent way to distinguish among clothes they want to wear, or vehicles that they see, such as the above-mentioned "pizza-car".

Yet, note that many of the children's coinages, which involve perception, can do so in an imaginary way, or in other words, as mental imagery. A compound, such as *champagnetröja* 'champagne-sweater', was uttered to denote a non-existent sweater that the child just dreamt up when playing.

4.6 Overcategorization

We will use the term "overcategorization" to label some striking features among the children's compounds. Underextension, often involving redundancy, is one way to arrive at overcategorization, as we see it. For instance, *kogräs* 'cow-grass' denotes 'ordinary grass, that cows eat' according to one child. Additionally, an ordinary car is referred to as *motorbil* 'motor-car', or *handfinger* 'hand-finger' is used instead of just finger for the body part. In these three examples, N_2 alone would have been the target like word to use, but the children limit its use further.

A quite odd categorization made by all three children, independently, is to add the goal of a direction to the direction: *kalasväg* 'party-road' or *mormorväg* 'granny-road'. Recall that *väg* 'road' was one of the nouns that reoccurred frequently among the novel compounds (cf. 4.1). Hence, the three children seem to find it important to name particular roads.

Furthermore, nearly 20 of the children's compounds contain one of the words *mamma* 'mommy', *pappa* 'daddy' or *bebis* 'baby' as N_1 or N_2, such as *mammfluga* 'mommy-fly', *fågelpappa* 'bird-daddy' or *bebismyra* 'baby-ant'. All three children coined such compounds, which we interpret as a kind of emergent categorization, as well as of overcategorization. There were two types of relations involved in these compounds: animals or insects subcategorized according to human kinship terms as in the preceding examples; mommy or daddy subcategorized according to some habit, such as *cigarettpappa* 'cigarette-daddy'.

4.7 Ad hoc categorization

Barsalou (1983) uses the label ad hoc categories for categories constructed on the spot to achieve certain goals, such as "things to sell at a garage sale". These categories are much less established in memory than common categories. We interpret ad hoc categories to encompass compounds such as Downing's (1977) "apple-juice seat", and also the examples from Clark, Gelman and Lane (1985) claimed to involve a temporal relation, in contrast to compounds with inherent relations. According to Clark, Gelman and Lane (1985), children would more often use novel compounds to express inherent relations among objects. The opposite stand is taken by Mellenius (1997), supported by Berman (2009), who claims children's novel compounds are "highly 'context-dependent' and hence more likely to express temporary rather than intrinsic relations" (2009:311).

Some innovations in our data can be analyzed as ad hoc instances that the children coin spontaneously without a real naming demand. They are

typically difficult to understand, or does not make sense, outside the context of the utterance. An example is one child in the data that invents a triplet of compounds with *glass* 'ice cream' with the goal "things that could possibly constitute ice cream": *träglass* 'wood-ice cream', *sockerglass* 'sugar-ice cream' and *glassögon* 'ice cream-eyes'; the latter denotes, according to the child, 'eye-glasses but made of ice cream (*glass*) instead of glass (*glas*)'. Another example is *kungtröja* 'king-sweater', coined on the spot when playing: 'if you wear that sweater you will be the king'.

However, our data points in the direction that the children's innovations more often express inherent relations than temporal relations, but this issue certainly merits further investigation.

5 Conclusion

The study provides evidence of on-line categorization based on spontaneous production of novel NN compounds from three Swedish children. Compared to experimental situations, limited by the material used and the children's will and energy to participate, our collection of data is unique. It shows that high contrast perceptual features give rise to much subcategorization, however not at the expense of conceptual subcategorization, equally important in our data.

Since we lack clear longitudinal facts of how object categorization emerges within the children, the structure view is hard to apply. We can state that L_3 and L_4 categories appear around age 2, but lack numbers about their overall frequency in relation to more inclusive categories. Given that the children show cognitive flexibility in their categorization of an object in a particular way by producing an NN compound, the processing view conforms better to our data. To conclude, the children often categorize objects in a much more detailed way than adults do.

References

Lawrence W. Barsalou. 1983. Ad Hoc Categories. *Memory & Cognition*, 11(3):211-227.

Ruth A. Berman. 2009. Children's Acquisition of Compound Constructions. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 298-322. Oxford: Oxford University Press.

Marc H. Bornstein and Martha E. Arterberry. 2010. The Development of Object Categorizaion in Young Children: Hierarchical Inclusiveness, Age, Perceptual Attribute, and Group versus Individual Analyses. *Developmental Psychology*, 46(2):350-365.

Eve V. Clark. 2004. How Language Acquisition Builds on Cognitive Development. *TRENDS in Cognitive Sciences*, 8(10):472-478.

Eve V. Clark and Ruth A. Berman. 1987. Types of Linguistic Knowledge: Interpreting and Producing Compound Nouns. *Journal of Child Language*, 14(3):547-567.

Eve V. Clark, Susan A. Gelman and Nancy M. Lane. 1985. Compound Nouns and Category Structure in Young Children. *Child Development* 56(1):84-94.

Pamela Downing. 1977. On the Creation and Use of English Compound Nouns. *Language* 53(4):810-842.

Anna V. Fisher. 2011. Processing of Perceptual Information Is More Robust than Processing of Conceptual Information in Preschool-Age Children: Evidence from Costs of Switching. *Cognition*, 119(2):253-264.

Eleanor J. Gibson. 1969. *Principles of Perceptual Learning and Development*. New York: Appleton-Century-Crofts.

Stevan R. Harnad (ed). 1987. *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.

Susan S. Jones and Linda B. Smith. 1993. Place and Perception in Children's Concepts. *Cognitive Development*, 8:113-139.

Frank C. Keil and Michael H. Kelly. 1987. Developmental changes in category structures. In S. Harnad (ed.), *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.

Jean M. Mandler. 2000. Perceptual and Conceptual Processes in Infancy. *Journal of Cognition and Development*, 1(1):3-36.

Ingmarie Mellenius. 1997. *The Acquisition of Nominal Compounding in Swedish*. [Travaux de l'Institut de Linguistique de Lund 31]. Lund: Lund University Press.

Gregory L. Murphy. 2002. *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Simone P. Nguyen and Gregory L. Murphy. 2003. An Apple is More Than Just a Fruit: Cross-Classification in Children's Concepts. *Child Development*, 24(6):1783-1806.

Eleanor Rosch. 1978. Principles of Categorization. In Eleanor Rosch and Barbara B. Lloyd (eds.), *Cognition and Categorization*, 27-48. Hillsdale, NJ: Erlbaum.

Robert Schreuder and Harald B. Baayen. 1997. How Complex Simplex Words Can Be. *Journal of Memory and Language*, 37:118-139.

Linda B. Smith. 1984. Young Children's Understanding of Attributes and Dimensions: A Comparison of Conceptual and Linguistics Measures. *Child Development*, 55(2):363-380.

Sandra R. Waxman and Dana B. Markow. 1995. Words as Invitations to Form Categories: Evidence from 12- to 13-Month-Old Infants. *Cognitive Psychology*, 29:257-302.