

ASA NORDQVIST
Göteborg University

THE USE OF DIRECT AND INDIRECT SPEECH BY 1 " - TO 4-YEAR-OLDS*

This study examines how direct and indirect speech are used by 30 Swedish children in the age range 18 months to 4 years. The material combines naturalistic case study data of two children (1;9-4;0 years), with data of 14 three-year-olds playing with a doll house jointly with their mothers, as well as longitudinal narrative data from 14 children at two datapoints (3 and 4 years). It is argued that the attribution of speech to others by means of direct and indirect speech develops in parallel with the emerging awareness of other people's minds and intentions. The results showed that the first forms of direct speech in the longitudinal case studies appear around 2;2 years and indirect speech around 3;0. In addition, it was found that three-year-olds adapted their use of the forms to what type of activity is engaged in, but that they managed less well than the four-year-olds in assessing shared knowledge with the interlocutor and in conveying shifts of perspective.

Introduction

The reporting of speech of another speaker involves selection of an appropriate linguistic form: direct or indirect speech. In the case of direct speech, the perspective is shifted from the reporter to that of the reported speaker, and the deictic elements need to be adjusted accordingly. In indirect speech, in contrast, the perspective remains with the reporter and deictic elements in the speech reported must conform to the here-and-now of the reporter. Thus, the reporting of speech requires an ability to take different perspectives and to encode them linguistically. In addition to this, the speech reporter needs to consider the interlocutor's state of mind and assess shared knowledge in order to provide sufficient information about the reported speech event. A study of children's acquisition of direct and indirect speech should thus take children's emerging pragmatic awareness and theorizing of mind into account along with the linguistic development.

Studies of children's skills in speech reporting are relatively rare (see Nordqvist, 2001, for a review). However, Ely & McCabe (1993) investigated three English-speaking longitudinal case studies, and found that the first forms of direct and indirect speech appeared

* This article is based on the paper presented at the Turku Symposium on First Language Acquisition, September 1-2, 2000. The author would like to thank Sven Strömqvist and Sally Boyd for valuable discussions and comments regarding the contents of this text. Address for correspondence: Asa Nordqvist, Department of Linguistics, Göteborg University, Box 200, SE-405 30 Göteborg, Sweden. Tel: +46 31 7734627; Fax: +46 31 7734853. E-mail: asanord@ling.gu.se

between 2 and 2;6 years in these children. This indicates that direct and indirect speech can be produced by children at an early age. Ely & McCabe report that the children primarily reported their own immediate speech, suggesting that speech quoting is used for a relatively limited purpose at this stage. Goodell & Sachs (1992) used an experimental design to elicit personal narratives from subjects in the age groups 4, 6, 8 and adults, and Hickmann (1993) had subjects at ages 4, 7, and 10 and adults retell a previously heard puppet dialogue. In both studies, it was found that the 4-year-old children to a greater extent than the older subjects quoted speech directly without making clear who the reported speaker was. These findings can be summarized as follows: children can produce direct and indirect speech before age 3, but children at these ages may fall short in making pragmatic adaptations to the speech situation, such as providing the interlocutor with sufficient information about the reported speech event.

The present study examines the use of speech reporting by Swedish-speaking children at the ages 1 1/2 to 4 years, engaged in various types of activities. More specifically, I address the questions as to when direct and indirect speech first appear, what the course of development looks like, and how direct and indirect speech are used in relation to the different types of activities. Moreover, I try to determine to what extent the children provide their interlocutors with sufficient information and successfully convey shifts of speaker perspective when they report speech. The findings will be discussed and related to the children's linguistic, cognitive, and pragmatic development.

Direct and indirect speech

Linguistically, direct speech typically consists of two (or more) main clauses, representing the quoted utterance(s) on the one hand, and a framing clause on the other. This type will be referred to as *framed direct speech* henceforth, and an example of the type is provided in (1) below. The framing clause contains a reference to the quoted speaker and a verb of saying, and this clause either precedes, interrupts or follows the quote. The direct quote can also be free-standing, and is then an example of *free direct speech* (cf. example (2) below). In such a case, other means may be necessary to provide information about who the quoted speech refers to. For example, the voice quality can be changed in order to depict the referent (Clark & Gerrig, 1990). Indirect speech, illustrated in (3) below, consists of a main clause (typically including verbal information about the reported speaker and a verb of saying) and a subordinate clause representing the quote.

- (1) Joanna said, "I like this game".
- (2) "I like this game!"
- (3) Joanna said that she liked that game.

Direct and indirect speech are often used to refer to past speech events and are embedded in personal narratives (this is what the common term *reported speech* typically refers to). However, these forms may be used even when there is no actual speech event to report and to take a stand on, for example in fictional contexts like in make-believe play. Elsewhere, Nordqvist (2001) refers to the former as *speech reporting* (i.e., when a past speech event is reported on), and to the latter as *speech projection* (i.e., when speech is projected onto characters in play or in a fantasy story). The data design of the study below makes

possible comparisons between the use of direct and indirect speech in different types of activities by young children.

Design of the study

Subjects and data

The study includes 30 monolingual Swedish-speaking children, and combines different types of data: Two children who are followed longitudinally (age range 1;7-4;0), fourteen 3-year-olds performing a make-believe play task, and fourteen children telling a picture elicited story at 3 years and at 4 years. An overview of the data is given in table 1.

Table 1. Overview of the data (M = mean)

Type of data	N of subjects	Age
Naturalistic longitudinal case studies	N = 2	1;7-4;0
Make-believe play task Doll house play	N = 14	M = 3;2
Narrative task The frog story	N = 14	M = 3;2 M = 4;2

The two longitudinal case studies refer to one boy, Harry, and his baby sister Tea, and these are drawn from the Strömquist and Richthoff corpus (Strömquist, Richthoff, & Andersson, 1993; Richthoff, 2000). Harry and Tea have been audio and/or video recorded individually during typical everyday activities (e.g. free play, book reading, bedtime and eating routines) with their close family members. The recordings were made at approximately three-week intervals and the material has been transcribed and coded according to CHAT-format (MacWhinney, 1995). For each child, 28 age-matched recordings were selected for analysis, yielding a corpus size of around 175,500 words.

The next type of data consists of video-recordings of fourteen 3-year-olds playing with a doll house with their mothers in a laboratory setting (Winberg & de Chateau, 1982; Söderbergh, 1982; Wiberg, Humble, & de Chateau, 1989). The mothers and their children were invited to play for about half an hour with a doll house consisting of four rooms with different kinds of furniture. They also had a number of dolls to choose from in their play. The recordings were transcribed, and the total size of the corpus approximates 23,000 words.

The final type of data, narrative data elicited by means of a wordless picture story book, was collected in a project by Berglund and Eriksson (2000). The picture series used was *Frog, where are you?* (Mercer Mayer 1969), which consists of 24 pictures and depicts the adventures of a little boy and his dog as they go out into the woods to search for a frog that

has disappeared from the boy's room. The parents of the 3-year-olds were provided with a tape recorder, and the children were recorded as they related the frog story. The procedure was repeated one year later, when the children were 4 years old. The recordings were transcribed in CHAT-format, and the corpus comprises around 17,000 words in total.

Types of analysis

Speech projections as well as speech reporting were considered in the analyses. Thus, instances of direct and indirect speech produced in make-believe play, in fantasy story telling, and reading of books were included in addition to discussions and reports of past events. The identification and coding of an instance (or candidate) of direct and indirect speech in the data, was based on syntactic and deictic criteria. Indirect speech was operationalized as utterances with the form "X said that ___", where the embedded clause had to be a proposition (cf. example (3) above). Framed direct speech typically had the form "X said, ___" or "___, X said" (cf. example (1)), and free direct speech had the form of "___!" (example (2)). The reported or projected utterance could be a proposition or an animal cry or the like. Besides typical linguistic markers, such as the use of a pronoun or an infinitive clause, transpositions in syntactic person or verb tense, and speech act verbs, changes in tone of voice, gestures and other contextual cues helped determine what was a candidate for direct and indirect speech.

The analyses were qualitative rather than quantitative, although the analyses of the doll house and the frog story data included some quantitative measures. The reason for not quantifying the findings of the longitudinal case studies is that, firstly, the forms of direct and indirect speech develop gradually and some instances are only candidates for the forms investigated; secondly, the recordings cover a wide range of activities, and are of different sizes, which makes quantitative comparisons among data points and children difficult. The extent to which the forms of direct and indirect speech were spontaneously produced by the children, or rather were imitations of, or prompted by adult utterances, was also considered in the analyses.

The results of the analyses of the longitudinal case studies, the doll house data, and the frog story data, are discussed separately below.

Results and discussion

The longitudinal case studies

In an early phase of Harry's and Tea's language development (before age 2), mimicking of animal cries in relation to book reading is frequent. Although this is not about speech projection proper, the animals are ascribed sounds by the children (e.g., *bää*, 'baa') and the adults are likely to put the mimicking into a frame of direct speech (*ja sí säger lammen, bää säger lammen*, 'yes that's what the lambs say, baa the lambs say'). Around the children's second birthday, Harry and Tea start to produce their first forms of direct speech. These are typically prompted by the adults and the quotes are free rather than framed. Below in (4) is an example from a recording of Harry at age 2 years.

- (4)
 Harry 2;0
 *MOT: va gjorde dom dí trolle?

%eng: what did the trolls do then?
 *HAR: haj
 %eng: hi
 *MOT: sa du hej te dom?
 %eng: did you say hi to them?

The use of free direct speech is explained by two facts: first, in many cases the preceding adult utterance itself consists of a framing clause, thus, there is no need for the child to frame the subsequent quote; second, the children's ability to coordinate two clauses at this stage of development is limited (the Mean Length of Utterance value is between 1-2). Hence, it seems more functional for the children at this stage to express mere quotes rather than to express clauses that frame quotes. Another way to put it is to say that the children at this age quote and enact rather than frame and narrate.

The first forms produced by Harry and Tea that are not prompted by an adult, are found around 26 months. Harry makes use of a free direct speech utterance in a conversation about personal experiences, and Tea in the context of make-believe play. Nevertheless, it is not until the children approach their third birthday that the forms are used more often spontaneously (i.e., not prompted or contingent on previous adult utterances), and also indirect speech appears (although the deictic elements are not adjusted appropriately in all cases) in addition to full-blown forms of framed direct speech. We can thus summarize the development as: free direct speech to be the first form to appear (around 26 months), followed by framed direct speech (28-30 months of age), and a little later (and not as common in use) indirect speech (35-36 months).

When considering the linguistic structures of these forms, this type of development is the one to be expected. Thus, the simplest form from a structural point of view, free direct speech, may consist of only one word (or similar), whereas framed direct speech requires a combination of at least two clauses. This is true for indirect speech, too, but in addition, this form involves subordination of clauses. The ability to subordinate clauses is indeed a later emerging skill in children's language development (e.g. Plunkett & Strömqvist, 1992). Also, from a pragmatic and deictic point of view, this course of development is to be expected. The use of free direct speech requires no shift of deixis and in the case of speech reporting the utterance is more or less "repeated". When direct speech is framed, at least two clauses are combined. The perspective is shifted from the reporting speaker to the reported speaker, and the deictic elements need to be shifted accordingly. Indirect speech is additionally complex in that the speech of another speaker is reported, but the perspective remains with, and the deictic elements are adjusted to, the reporting speaker.

The child's emerging awareness that the reported or projected speakers have independent minds has consequences for the encoding of perspectives and the use of direct and indirect speech. Previous research has shown that children start to attribute intentions to others between 2;6 and 3 years of age (Poulin-Dubois & Shultz, 1988). Kavanaugh, Eizenman, & Harris (1997) point out that children at this age have toy figures carrying out make-believe actions in play, and Wolf, Rygh, & Altshuler (1984) found that children begin to ascribe intentions to the figures in play with small replicas around the age of 2;7. The fact that both Harry and Tea increase their use and control over speech reporting and speech projection exactly during the months before they turn three indicates that the ascription of speech to others is part of the attribution of intentions to others. Tea and her

caretakers are often and from an early point in time engaged in make-believe play, but it is not until Tea is 2;7 that she herself occasionally projects speech in terms of free direct speech onto the dolls (still, in comparison to the adults, she is not doing it frequently at this point). From 2;11 and onwards, we find longer episodes of projected dialogue exchanges between the dolls with which Tea is playing.

In order to report and project speech successfully, the narrator also has to take his interlocutors into consideration. For speech reporting/projecting this means for the child to take into account the perspective of the interlocutor and make use of shared knowledge. Thus, in order for Tea's interlocutors to be able to keep track of the changes of speaker perspective in Tea's make-believe play, and for Harry's parents to understand who the reported speakers are in his vivid speech reports, the children need to provide their interlocutors with sufficient information. By age 3 both children do this to some extent; for instance, Tea changes her tone of voice in order to depict the speech of a certain doll, Harry frames the quotes with framing clauses, and both children weave in direct and indirect speech in narrative frames.

Playing make-believe with a doll house

In the data of the 3-year-olds playing with the doll house, 166 speech projections were found in total and these were distributed over all fourteen children. The overwhelming majority (n=159) of the speech projections took the form of free direct speech. This finding is in accordance with what was found in the play sequences of the longitudinal study of Harry and Tea. The free direct speech utterances are parts of dialogue exchanges, where the child typically enacts the role of one of the dolls and the mother takes the role of the other. Yet some 3-year-olds create entire dialogues on their own. Below in (5) is an example of a girl enacting a dialogue between two dolls within the play, and then she turns to her (real) mother and makes a comment about the dialogue exchange.

(5)

Situation: The child (CHI) holds onto two dolls, one in each hand, places them outside the doll's house and makes them speak with each other:

***CHI: gomorron gomorron har du sovi gott inatt?**

%eng: Good morning good morning have you slept well tonight?

%act: the child glances at the doll she is holding in her right hand; uses a squeaky voice

***CHI: ja de har ja verkligen [=!] gjort**

%eng: Yes I have indeed [=!]

%act: gazes at the doll she is holding in her left hand; uses a squeaky voice

***CHI: jaha**

%eng: uhum

%act: glances at the doll in her right hand; uses a squeaky voice

***CHI: visst var dom tokiga?**

%eng: They were pretty silly, weren't they?

%act: the child turns and looks at her mother and makes a comment on the dolls' conversation; normal narrator's voice pitch

There are functional explanations for the fact that free direct speech is used to a greater extent than framed direct and indirect speech in this type of activity. Since the children

have concrete objects (dolls) to hold, it is clear who is the projected speaker (e.g. by holding the dolls, wiggling them, directing the gaze towards them). Thus, there is no need to add a framing clause. Moreover, the projection of speech onto the dolls has consequences for advancing the plot in the play and may be employed as a strategy for this purpose (cf. a child accompanying the doll's walking in the house by projecting "now I'm going to see if I can find my friend").

In the doll house interactions, we also find evidence of packaging of information that serves to depict the speech of a certain referent and convey shifts of perspective. In (5), the 3-year-old uses a disguised and affected voice for both doll-speakers, and notably, she lets one of the doll-speakers use the word *verkligen*, 'indeed'. This word is not likely to belong to the child's own ordinary speech register, but is selected in order to depict the speech of an older distinguished woman. Figure 1 below is another example of how changes in voice quality can be used to signal changes of perspective. The figure shows the F_0 (fundamental frequency) correlates of two succeeding utterances delivered by a 3-year-old playing with two different dolls. In the first utterance, the child disguises his voice to depict a baby's voice (very squeaky), and in addition, in the adjacent utterance, the child uses an exaggerated deep voice to depict a father's voice.

In the first utterance, the utterance projected onto the baby doll, the mean pitch value is 370 Hz. The second utterance, projected onto the father doll, has a mean value of 203 Hz. In addition to these changes in pitch, the boy uses a childish pronunciation of *korv*, 'hot dog', namely *kojv*, in the first utterance, while in the second utterance *korv* is pronounced according to the adult standard form. The child also manipulates the temporal characteristics of his speech to indicate projected speaker shifts. The baby impersonation is slower than the impersonation of the father's speech. Playing house seems to be a good forum to show and practice not only linguistic skills, but, in addition, pragmatic and sociolinguistic skills.

Despite these fairly sophisticated ways of depicting different speakers and conveying the shifts of perspective, we find examples where the shifts are not made clear. One problematic aspect concerns what Sawyer (1996) refers to as *direct* versus *indirect voicing* (this should not be confused with the terms direct and indirect speech). Direct voicing deals with the play strategy where a child "becomes" a character, whereas indirect voicing

refers to when a child enacts a play role through the medium of a toy figure and projects speech onto the figure. The mothers in this study typically stick to indirect voicing whereas some of the children alternate between the two play strategies. To an outside observer (and perhaps also to the mothers) it is difficult to follow these changes of role identity. The fact that these types of ambiguities are found within the children's play and only rarely in the play of the mothers is consistent with Fein, Moorin, & Enslein's (1982) claim that indirect voicing (or *external agent play* in their terminology) is the most advanced form of make-believe play.

Relating the frog story

In the pictures of the frog story we can see how the boy and his dog search for their little frog in a forest. During the search they meet and interact with a variety of animals in the forest. The analyses of the 3-year-olds relating the story show that ten out of fourteen children ascribe speech or animal cries to any of the characters (the boy, the dog, or some other animal). In contrast to what was the case in the doll house play, most of the direct quotes were framed (e.g., *pojken skriker: "kom hit kom hit kom hit!"*, 'the boy yells, «come here come here come here»'). This was found in the data of the two longitudinal case studies as well (i.e., when Harry or Tea and their interlocutors were engaged in situations of book reading they used more framed direct speech than free direct speech). The preference for framed direct speech in the relating of the frog story can be explained by at least two facts. First, framed direct speech has been shown to occur frequently in children's books and fairy tales (Baker & Freebody, 1989; Perera, 1996). The children who are exposed to these literary forms thus have experience of this type of language use (for instance, if they have had stories read to them); then use of framed direct speech indicates an emerging sensitivity to genre by the 3-year-olds. Second, the frog story is a picture story, and in contrast to the doll house there are no concrete objects to hold. Consequently, in order to adapt to these conditions it is functional to refer to the characters by verbal means (like adding a framing clause).

The fact that the 3-year-old narrators often frame their quotes thus indicates an adaptation to the communicative conditions and an awareness that this type of activity requires verbal explicitness. However, in the cases where the 3-year-old children use free direct speech in this activity, they fail to make clear whether an utterance is intended to be a speech projection or not, and/or who is the projected speaker. For example, one 3-year-old frog story narrator exclaims *jag ska leta!*, 'I will search [for the frog]!', and it is difficult to know if *jag*, 'I', refers to the narrator/child himself, or if the narrator/child is intending to project the speech act to, for instance, the boy character in the story. This referential ambiguity is in accordance with Hickmann's (1993) claim that younger children tend to have problems marking explicit boundaries between narrative and narrated speech. In addition, this example shows that the boundaries between reality and fiction are also problematic.

Previous research on children's developing conceptions of other people's minds shows that 4-year-olds (but not 3-year-olds) understand false belief, the distinction between appearance and reality, and that something seen may present different visual experiences if the observer views it from different positions in space (e.g. Ferguson & Gopnik, 1988, Flavell, 1988). Hence, the period between three and four years of age is important for the development of perspective taking. The analyses of the frog stories produced by the chil-

dren one year later, at 4 years, indeed show that they at this point in time do better when it comes to making clear who the projected speakers are than they did at age 3. Half (49%) of all speech projections made by the 3-year-olds turned out to be problematic in that it was difficult or impossible to identify and code the quoted speaker. One year later, this percentage had decreased to 14%. Other indicators of development between these ages include the fact that the 3-year-olds needed more adult scaffolding to perform the task than the 4-year-olds did, and also the fact that the younger children used present tense and simply described what they saw in the pictures, whereas some of the 4-year-olds used a narrator's voice and anchored their stories in past tense (for a fuller account of the findings, see Nordqvist, 2001).

Conclusion

These analyses have shown that the attribution of speech to others (inanimate as well as animate objects) is related to the development of attributing intentions to others. Thus, direct and indirect speech appear between 2 and 3 years, and are used increasingly as the children approach their third birthday. Three-year-olds are aware of communicative conditions connected to different types of activities (cf. framed direct speech is preferred in story-telling, and free direct speech in make-believe play), yet they do not always adapt to the interlocutor and provide sufficient information about the speech reported/projected. Four-year-olds have a more developed theory of mind and awareness about different perspectives, and also manage better to convey shifts of perspective to the interlocutor.

References

- Baker, C. D. & Freebody, P. (1989). *Children's first school books*. Oxford: Blackwell.
- Berglund, E. & Eriksson, M. (2000). Communicative development in Swedish children 16-28 months old. The Swedish early communicative development inventory – words and sentences. *Scandinavian Journal of Psychology*, 41, 2.
- Clark, H. H. & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, 66, 764-805.
- Ely, R. & McCabe, A. (1993). Remembered voices. *Journal of Child Language*, 20: 671-696.
- Fein, G. G., Moorin, E. R., & Enslein, J. (1982). Pretense and peer behavior: An intersectoral analysis. *Human Development*, 25, 392-406.
- Flavell, J. H. (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. W. Astington, P. L. Harris & D. R. Olson (Eds.), *Developing theories of mind*. Cambridge: Cambridge University Press.
- Forguson, L. & Gopnik, A. (1988). The ontogeny of common sense. In J. W. Astington, P. L. Harris & D. R. Olson (Eds.), *Developing theories of mind*. Cambridge: Cambridge University Press.
- Goodell, E. W. & Sachs, J. (1992). Direct and indirect speech in children's retold narratives. *Discourse Processes*, 15, 395-422.
- Hickmann, M. (1993). The boundaries of reported speech in narrative discourse: some developmental aspects. In J. A. Lucy (Ed.), *Reflexive language: reported speech and metapragmatics*. Cambridge: Cambridge University Press.

- Kavanaugh, R. D., Eizenman, D. R. & Harris, P. L. (1997). Young children's understanding of pretense expressions of independent agency. *Developmental Psychology*, Vol. 33; No. 5, 764-770.
- MacWhinney, B. (1995). *The CHILDES project: Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mayer, M. (1969). *Frog, where are you?* New York: Dial Press
- Nordqvist, L. (2001). *Speech about speech: A developmental study on form and function of direct and indirect speech*. Doctoral Dissertation. *Gothenburg Monographs in Linguistics*, 19. Dept. of Linguistics, Göteborg University.
- Perera, K. (1996). Who says what? Learning to 'read' the punctuation of direct speech. In N. Hall & A. Robinson (Eds.), *Learning about punctuation*. Portsmouth, NH: Heinemann.
- Plunkett, K. & Strömquist, S. (1992). The acquisition of Scandinavian languages. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition, Vol 3*. Hillsdale, NJ: Lawrence Erlbaum.
- Poulin-Dubois, D. & Shultz, T. R. (1988). The development of the understanding of human behavior: From agency to intentionality. In J. W. Astington, P. L. Harris & D. R. Olson (Eds.), *Developing theories of mind*. Cambridge: Cambridge University Press.
- Richthoff, U. (2000). *En svensk barnspråkskorpus. Uppbyggnad och analyser*. Licentiate Thesis. Department of Linguistics, Göteborg University.
- Strömquist, S., Richthoff, U. & Andersson, A.-B. (1993). Strömquist's and Richthoff's Corpora – a guide to longitudinal data from four Swedish children. *Gothenburg Papers in Theoretical Linguistics* 66. Department of Linguistics, Göteborg University.
- Söderbergh, R. (1982). Linguistic effects by three years of age of extra contact during the first hour post partum. In C. E. Johnson & C. Larzon (Eds.), *Proceedings of the Second International Congress for the Study of Child Language*. Washington D.C.: University Press of America.
- Wiberg, B., Humble, K. & de Château, P. (1989). Long-term effect on mother-infant behavior of extra contact during the first hour post partum: Follow-up at three years. *Scandinavian Journal of Social Medicine*, 2, 181-191.
- Winberg, J. & de Château, P. (1982). Early social development: Studies of infant-mother interaction and relationships. *Review of Child Development Research*, 6, 1-44.
- Wolf, D., Rygh, J., & Altshuler, J. (1984). Agency and experience: Actions and states in play narratives. In I. Bretherton (Ed.), *Symbolic play*. New York: Academic Press.