Pronoun Interpretation in Direct and Indirect Speech

Franziska Köder (f.koder@rug.nl)
Faculty of Philosophy, University of Groningen,
Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands

Abstract

This study examines how the context shift in direct speech reports influences the interpretation of personal pronouns. 108 native speakers of Dutch played an online game where they had to identify the referent of the singular pronouns *I, you* and *he* that are either presented in direct speech (e.g. *Elephant said "I get the football"*) or indirect speech (e.g. *Elephant said that I get the football*). The error rate and reaction time was higher for pronoun interpretation in direct speech than in indirect speech indicating that the context shift in direct speech increases the processing load. I discuss how this can be reconciled with apparently contradicting evidence from crosslinguistic and developmental studies that suggest that direct speech is the easier speech reporting type.

Keywords: Personal pronouns; direct speech; indirect speech; context shift

Introduction

Many languages have two different types of speech reports: direct speech (1) and indirect speech (2).

(1) Direct speech: Anna zei "Ik ben gelukkig".
    'Anna said "I am happy".'

(2) Indirect speech: Anna zei dat ik gelukkig ben.
    'Anna said that I am happy.'

In contrast to English, there are clear syntactic and lexical cues in Dutch that distinguish direct from indirect speech. In direct speech, a reported declarative sentence has verb-second word order. In indirect speech, the report is presented in a subordinate clause with verb-final word order and with an obligatory complementizer *dat* ('that'). An important semantic difference between direct and indirect speech is that the context of evaluation shifts in direct speech from the actual utterance context to the reported utterance context inducing a shift in the interpretation of indexicals (like *I, tomorrow, here*) (Kaplan, 1989). A result of this context shift is that the quoted first person pronoun in example (1) refers not to the actual speaker (like in example (2)), but to the reported speaker Anna.

Direct vs. indirect speech

The present study compares the process of pronoun interpretation in direct and indirect speech. Based on different kinds of evidence, two conflicting hypotheses about the cognitive differences can be put forward.

The first hypothesis is that pronouns in direct speech are easier to interpret than in indirect speech. This is based on the following observations. First, direct speech is crosslinguistically more widespread (Li, 1986). Second, in languages where both constructions exist, direct speech is acquired before indirect speech (Ely & McCabe, 1993; Nordqvist, 2001) and in natural interactions it is used much more frequently by children and their interlocutors (Köder, 2013). Especially the fact that adults use predominantly direct speech when interacting with their children, suggests that direct speech is not only easier from a speaker's perspective, but also from a hearer's perspective.

Considerations from a formal semantic perspective point in the opposite direction: Since the context of evaluation needs to be shifted in direct speech, a higher processing effort is to be expected for the interpretations of pronouns in direct speech in comparison to indirect speech.

In this study, processing effort is operationalized as more mistakes and higher reaction times when identifying the referent of a personal pronoun.

First and second vs. third person pronouns

In addition to the comparison between direct and indirect speech, I also analyze whether there are differences between interpreting first, second and third person singular pronouns in a speech reporting environment.

Lyons (1968, 1977) sees a fundamental distinction between first and second person pronouns on the one hand and third person pronouns on the other hand. While *I* and *you* refer to the primary participants of an utterance situation, i.e. speaker and hearer, third person pronouns like *he, she* or *it* are negatively defined as referring to a person or thing other than speaker and hearer. This split between first and second vs. third person pronouns is supported by studies on children's comprehension of deictic pronouns that found that the understanding of third person pronouns clearly lags behind first and second person pronouns (e.g. Brener, 1983; Charney, 1980; Legendre & Smolensky, 2012).

A new contribution of this study to the field of pronoun interpretation is to investigate how a first and second person vs. third person split is affected by the context shift in direct speech in comparison to indirect speech.

Method

Participants

108 native monolingual speakers of Dutch (56 female, 52 male, average age: 24.4 years) participated in the study. They were recruited by university related mailing lists and the digital learning environment of the University of Groningen.

Materials

The experiment is designed as an online game that the participants can access via a link. It consists of short animated scenes with three animals, a dog, a monkey and an elephant. Animals were chosen as protagonists because the experiment was originally designed for children. All three animal protagonists are male, so that gender information of the third person pronoun cannot help to identify the correct referent. Following Brener (1983), I ensured that the experimental subjects are not possible referents of pronouns themselves, but observe the scenes from the outside. This guarantees comparable conditions for the interpretation of first, second and third person pronouns.

In each scene, one animal utters a sentence that indicates which of the three protagonists gets a particular object. The sentences are of a certain reporting type (no report (see example (4)), direct speech (5), indirect speech (6)) and include a pronoun of a certain type (*ik* ('I'), *jij* ('you'), *hij* ('he')). This leads to a total of nine possible combinations of reporting type and pronoun type. The no report condition was included as a baseline to assess participants' pronoun interpretation in a non-embedded environment. All pronouns are deictic pronouns with the exception of the third person pronoun in indirect speech which is an anaphoric pronoun.

Procedure

The experiment consists of an introduction phase and a test phase. In the introduction phase, the three animals introduce themselves and it is checked whether the participants remember their names (*Hond* 'Dog', *Aap* 'Monkey', *Olifant* 'Elephant'). All animals have different male voices. Then, the 18 objects used in the experiment are presented and named. 3 practice items follow to make sure that participants are familiar with the procedure of clicking on the person who receives the object. The practice items are comparable to the no report items, but contain proper names instead of pronouns (see example (3)).

(3) Practice item: Olifant krijgt het boek.
               'Elephant gets the book.'

The test phase is split in two parts. First the 15 no report items (5 per pronoun type) are presented in random order. At the beginning of each scene, the participants see the three protagonists and one object on the screen. Then one animal walks towards a second animal and tells him who gets the object (see figure 1b).

(4) No report: Ik/ Jij/ Hij krijg(t) de voetbal.
               'I/ You/ He get(s) the football.'

After the end of the sentence, the protagonists are presented in their original frontal position. A basket appears in front of each animal and they are highlighted with yellow bars (see figure 1c). These cues indicate that the participants can now click on the correct recipient of the object, that is the correct referent of the pronoun. Irrespective of participants' choice, the object jumps into the basket of the selected person. The program records for each participant and item the correctness of pronoun interpretation and the reaction time.

In the second part of the test phase, 15 direct and 15 indirect speech items (always 5 per pronoun type) are presented in random order. The testing procedure is similar to the no report items, but includes an additional step. Now one animal whispers – inaudibly for the participants – into another animal's ear which of the three animals gets the object (see figure 1a). The original addressee passes the information on to the third animal using either direct (5) or indirect (6) speech.
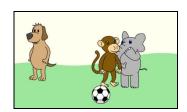
(5) Direct speech:
        Olifant zei "Ik/ Jij/ Hij krijg(t) de voetbal".
        'Elephant said, "I/ You/ He get(s) the football".'
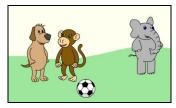
(6) Indirect speech:
        Olifant zei dat ik/ jij/ hij de voetbal krijg(t).
        'Elephant said that I/ you/ he get(s) the football.'

A characteristic of the direct and indirect speech items is that two utterance contexts overlap: the original utterance context in which the whispering takes place and the actual utterance context where the sentence is reported. In each utterance context, the protagonists assume different participant-roles (speaker, addressee, other person). Cues for identifying the speaker are mouth movement and the specific vocal qualities of each protagonist. The addressee can be identified by his proximity to the speaker and his body orientation towards the speaker. The protagonist with the participant-role 'other person' is clearly spatially separated from speaker and addressee and faces a different direction.

a. Whispering (only for direct and indirect speech)

b. Uttering critical sentence

c. Selection phase

Figure 1: Test materials

For all test items, the participant-roles of the animal protagonists, their spatial position, the sentences they utter and the objects they receive are counterbalanced.

Results

I used linear mixed-effects regression modeling with the software R (version 2.15.1) to explain the dependent variables error rate and log-transformed reaction time. The predictor variables are the categorical variables reporting type (no report, indirect speech, direct speech) and pronoun type (ik, jij, hij). Table 1 and 2 show how a stepwise inclusion of the predictor variables to the baseline model (including the random-effect factors) improves the goodness of fit of the models. This is indicated by the decrease of the Akaike Information Criterion (AIC) and the results of a chi-square test.

Table 1: Model for error rate

	AIC decrease	χ^2	Pr> (χ^2)
Random-effect factors (random intercepts per subject)			
+ Reporting type	65.4	69.4	p<0.0001
+ Pronoun type	293.6	297.5	p<0.0001
+ Reporting type * Pronoun	232.6	240.6	p<0.0001

Table 2: Model for (log-transformed) reaction time

	AIC decrease	χ^2	Pr> (χ^2)
Random-effect factors (random intercepts and slopes per subject)			
+ Reporting type	46.5	50.5	p<0.0001
+ Pronoun type	18.9	22.9	p<0.0001
+ Reporting type * Pronoun	53.9	61.9	p<0.0001

For both models, including the fixed-effect factors reporting type and pronoun type as well as the interaction between the two explains significantly more variance in the data.

Error rate

Figure 1 shows the mean error rate for each combination of reporting type and pronoun type. Participants make almost no mistakes in interpreting the pronouns *ik* (mean: 0.004) and *jij* (0.002) in the no report condition, but they have clearly more difficulties with the third person pronoun *hij* (0.319). The graph also already indicates that participants tend to make more mistakes in direct speech (ik: 0.096, jij: 0.148, hij: 0.248) than in indirect speech (ik: 0.063, jij:

0.083, hij: 0.091) and more in the interpretation of third person pronouns than of first and second person pronouns.
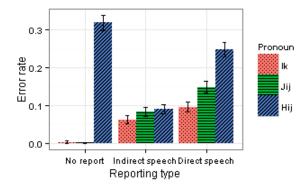

Figure 2: Mean error rate (±1SD)

A post-hoc analysis with the mcposthoc function from the 'LMERconveniencefunctions' package was performed to see which combinations of reporting type and pronoun type differ significantly. Table 3 presents all results of the pairwise, two-tailed test that are significant. Note that always pairs of combinations that include either the same pronoun or the same reporting type were compared. A positive estimate means that the element at the right-hand side of the dash in the first column has a higher error rate, a negative one that it has a lower error rate.

Table 3: Multiple comparisons, error rate

	Estimate	SD	z value	p value
Ik				
No – Ind	3.05	0.76	4.03	0.0001
No – Dir	3.60	0.75	4.79	p<0.0001
Ind – Dir	0.55	0.25	2.19	0.0289
Jij				
No – Ind	4.10	1.05	3.92	0.0001
No – Dir	4.92	1.04	4.72	p<0.0001
Ind – Dir	0.82	0.22	3.71	0.0002
Hij				
No – Ind	-2.16	0.21	-10.45	p<0.0001
No – Dir	-0.54	0.17	-3.23	0.0013
Ind – Dir	1.62	0.21	7.79	p<0.0001
No report				
Ik – Hij	5.67	0.74	7.63	p<0.0001
Jij – Hij	6.37	1.04	6.12	p<0.0001
Direct speech				
Ik – Jij	0.63	0.21	2.93	0.0034
Ik – Hij	1.54	0.20	7.52	p<0.0001
Jij – Hij	0.91	0.19	4.90	p<0.0001

For the first and second person pronouns, a clear difference between the three reporting types was found.

Participants make significantly more mistakes in both speech reporting conditions than in the no report baseline. Moreover, they are less accurate in identifying the referents of *ik*, *jij* and *hij* in direct speech as compared to indirect speech. A detailed analysis of the mistakes in direct speech reveals that in 93 percent of the cases, participants incorrectly evaluate pronouns with respect to the actual instead of the original utterance context.[1] These systematic mistakes suggest that participants ignore the lexical and syntactic cues that indicate direct speech and apply an indirect speech interpretation, thereby avoiding a cognitively demanding context shift.

Surprisingly, participants had the highest error rate for the interpretation of third person pronouns in the no report condition. A possible explanation for this is that it might be pragmatically dispreferred to use a third person pronoun in this context without further paralinguistic (stress) or nonlinguistic cues (eye gaze, pointing) that help to determine the correct referent.

The pairwise comparison also reveals significant differences between the three pronoun types. In the no report condition, the error rate was higher for *hij* than for both *ik* and *jij*. In indirect speech, however, no difference between first and second as compared to third person pronouns was found. The reason for this could be that third person pronouns in indirect speech are in contrast to all other pronouns in the experiment anaphoric pronouns. This is related to a different process of referent identification. While the deictic pronouns refer to the person in the extra-linguistic context that occupies a particular participant-role, the anaphoric pronoun *hij* refers to a linguistic antecedent, namely the proper names *Aap*, *Hond* and *Olifant* in the reporting clause, which in turn refer to the three animals. The results indicate that anaphoric pronoun resolution seems to be a cognitive process less prone to failure. This is consistent with evidence from child language acquisition. Charney (1980) found that children use anaphoric third person pronouns correctly before deictic ones.

In direct speech, not only the error rate for first and second person pronouns in contrast to third person pronouns differed significantly, also the difference between first and second person pronouns was significant.

Reaction time

The mean reaction time for the no report items was 1.2s, for indirect speech 1.7s and for direct speech 2.0s. Figure 3 shows the log-transformed reaction times for every combination of reporting type and pronoun type.

Figure 3: Mean log-transformed reaction time (±1SD)

In table 4, all significant results of a post-hoc analysis are summarized. The test only compared combinations with either the same reporting type or the same pronoun type. A positive estimate means that the element on the right-hand side of the dash has a higher reaction time.

There was no significant difference in reaction time between the no report items and the indirect speech items regardless of the type of pronoun used. Participants also did not need significantly longer to process the first person pronoun *ik* in direct speech than in the other two conditions. However, the high reaction times for *jij* and *hij* in direct speech clearly stand out. It takes participants significantly longer to identify a referent of these pronouns in direct speech than in indirect speech and the no report condition. Only in direct speech, a significant increase in reaction time from first person to second person to third person was found.

The results of the reaction time analysis support the hypothesis that direct speech is associated with a higher processing load. The context shift does however not affect all singular pronouns equally. The first person pronoun *ik* is interpreted equally fast as in indirect speech.

Table 4: Multiple comparisons, reaction time

	Estimate	SD	t value	p value
Jij				
No – Dir	0.26	0.04	6.19	p<0.0001
Ind – Dir	0.26	0.04	6.23	p<0.0001
Hij				
No – Dir	0.36	0.04	8.70	p<0.0001
Ind – Dir	0.34	0.04	8.19	p<0.0001
Direct speech				
Ik – Jij	0.25	0.04	6.62	p<0.0001
Ik – Hij	0.34	0.04	8.99	p<0.0001
Jij – Hij	0.09	0.04	2.37	0.0180

---

[1] A similar pattern was not found for the mistakes in indirect speech. In only 62 percent of the mistakes (50 percent chance level), participants wrongly selected the direct speech referent.

Discussion

This study found that participants make significantly more errors and have significantly longer reaction times when interpreting pronouns embedded in direct speech than in indirect speech. This supports the hypothesis that the context shift in direct speech is associated with higher processing demands. To identify the correct referent of a pronoun in direct speech, participants must shift from the actual to the original utterance context and select the protagonist with the respective participant-role there. The high number of systematic mistakes shows that there is a strong tendency to avoid this costly context shift and to evaluate pronouns with respect to the actual utterance context, analogous to indirect speech.

The results of this comprehension experiment seem to conflict with findings from cross-linguistic and developmental studies on the production of speech reports that suggest that direct speech is the cognitively less demanding speech reporting type. Direct speech occurs not only in more languages than indirect speech (Li, 1986), children also acquire it earlier and use it more frequently (Ely & McCabe, 1993; Köder, 2013; Nordqvist, 2001).

One possibility to reconcile this contradicting evidence is to claim an asymmetry between the production and the comprehension of speech reports. This is to say that direct speech is easier to produce for a speaker, but more difficult to interpret for a hearer. I will however argue for an alternative explanation, namely that the cognitive complexity of a speech report is dependent on the particular discourse context in which it is used.

The setup of the current experiment is in my view conducive to indirect speech. Other discourse contexts might however favor direct speech. In narratives for instance, the focus typically lies on the original utterance context while the actual context of the report (narrator telling story to audience) is back-grounded. To interpret pronouns in direct speech in this discourse context requires only a representation of the original utterance context and not a shift between actual and original utterance context.

To decide between the production-comprehension-asymmetry explanation and the discourse-context-dependency explanation, one could conduct a production variant of the current study. The experimental subjects assume the role of the protagonist who reports the original speaker's utterance to another person. I predict that in this discourse context, adults will use mainly indirect speech reports because that allows them to use pronouns that refer to the participants of the currently activated utterance context. With this strategy, they can avoid to refer to themselves with quoted *you*, to their addressee with quoted *he* and to the other person with quoted *I*. A production-comprehension-asymmetry explanation would however predict a preference for direct speech reports.

To avoid drawing too general and therefore false conclusions, the results of this study must be restricted to discourse contexts with a highly salient actual utterance context.

References

Brener, R. (1983). Learning the deictic meaning of third person pronouns. *Journal of Psycholinguistic Research, 12*(3), 235–262.

Charney, R. (1980). Speech roles and the development of personal pronouns. *Journal of Child Language, 7*(3), 509–28.

Ely, R., & McCabe, A. (1993). Remembered voices. *Journal of Child Language, 20*, 671–696.

Kaplan, D. (1989). Demonstratives. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–614). New York: Oxford University Press.

Köder, F. (2013). How children acquire reported speech in German and Dutch: a corpus study. In B. Sonnenhauser, C. Trautmann, & P. Noel Aziz Hanna (Eds.), *„Perspektiven" Diskussionsforum Linguistik in Bayern / Bavarian Working Papers in Linguistics 2* (pp. 15–28). München, Bamberg.

Legendre, G., & Smolensky, P. (2012). On the Asymmetrical Difficulty of Acquiring Person Reference in French: Production Versus Comprehension. *Journal of Logic, Language and Information, 21*(1), 7–30.

Li, C. (1986). Direct and indirect speech: A functional study. In: F. Coulmas (Ed.), *Direct and Indirect Speech* (pp. 29-45). Berlin, New York, Amsterdam: Mouton de Gruyter.

Lyons, J. (1968). *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge, London, New York, Melbourne: Cambridge University Press.

Nordqvist, Å. (2001). *Speech about speech. A developmental study on form and functions of direct and indirect speech*. Göteborg: Kompendiet.