The Influence of utterance-related factors on the use of direct and indirect speech

Li, Jianan; Jongerling, Joran; Dijkstra, Katinka; Zwaan, Rolf

Published in:
Collabra: Psychology

DOI:
10.1525/collabra.33631

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Li, J., Jongerling, J., Dijkstra, K., & Zwaan, R. (2022). The Influence of Utterance-Related Factors on the Use of Direct and Indirect Speech. *Collabra: Psychology*, *8*(1). https://doi.org/10.1525/collabra.33631

Cognitive Psychology

The Influence of Utterance-Related Factors on the Use of Direct and Indirect Speech

Jianan Li[1] [a], Joran Jongerling[1], Katinka Dijkstra[1], Rolf Zwaan[1]

[1] Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, the Netherlands

Keywords: Direct Speech, Indirect Speech, Non-verbal Information, Utterance Type

https://doi.org/10.1525/collabra.33631

Collabra: Psychology

Vol. 8, Issue 1, 2022

People routinely shift between direct and indirect speech in everyday communication. The factors that impact the selection between these two modes of reporting during language production are under-investigated. The present study examined how utterance-related factors (the vividness of non-verbal information and the utterance type) influence the use of direct and indirect reported speech in narratives. Participants were asked to watch and retell four movie clips. All narratives were videotaped and then transcribed verbatim for analyses. The data were analyzed using a mixed effects logistic regression model. The results showed that the utterances accompanied by vivid voice were more likely to be reported in direct speech. The vividness of facial expressions did not influence the form in which utterances were reported. In addition, we found that utterances that belonged to so-called *Main Clause Phenomena* were more likely to be reported in direct speech than in indirect speech. The current study helps us further understand the factors that influence structure choices during language production.

Introduction

People often quote their own or others' speech in daily communication, a phenomenon known as reported speech. Reported speech normally consists of two forms of constructions: direct speech and indirect speech, distinguished by the reporter's perspective (Coulmas, 1986). In direct speech (*Paul said: "I am hungry."*), the reporter talks in the original speaker's point of view. In indirect speech (*Paul said that he was hungry.*), on the other hand, the reporter presents utterances from his/her own point of view. Another marked difference between these two forms of reported speech is that direct speech conveys both the content and co-speech non-verbal information of previous utterances (e.g., voice, facial expressions, and gestures) while indirect speech only communicates the content (Li, 1986). Much of the literature has been devoted to describing the grammatical properties (Banfield, 1973) and discourse functions (Holt, 1996; Macaulay, 1987) of direct and indirect speech. However, little is known about the factors that account for their use, especially on the utterance level. The current study takes the first step to empirically address this gap in the context of a narrative.

Direct and Indirect Speech in Narratives

Because direct speech depicts the original speaker's voice, facial expressions, and gestures, it is often used in narratives to make stories more vivid and dramatic (Wierzbicka, 1974). It has been observed that people use direct speech to highlight the climax of stories and to deliver crucial information in narratives (Glock, 1986; Larson, 1978). Empirical evidence further supports these observations. In a study by Wade and Clark (1993), participants first watched videotaped dialogues and then were asked to recount what happened in the videos to listeners. Half of the participants were instructed to give accurate accounts, and the other half were asked to recount as amusingly as possible. Participants who were asked to entertain produced more direct speech than those participants who were asked to be accurate. In order to quantitatively test the assumption that direct speech is more vivid, Groenewold et al. (2014) tested whether direct speech was actually perceived as more lively than indirect speech. Participants listened and rated the liveliness of speech segments with or without direct speech. As predicted, speech fragments that contained direct speech received significantly higher scores for liveliness compared with fragments with indirect speech.

Together, these findings suggest that direct speech is associated with increased vividness or liveliness, explaining why speakers often use it to enrich and dramatize a story. However, on closer examination, the use of direct and indirect speech turns out to be more complicated. People used direct speech more frequently when they told a story enter-

a Correspondence concerning this article should be addressed to Jianan Li, Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, Burgemeester Oudlaan 50, Mandeville Building, PO Box 1738, 3000 DR, Rotterdam, the Netherlands. Email: li@essb.eur.nl

tainingly compared to when they told it accurately (Wade & Clark, 1993). However, under the instruction of being amusing, participants did not use direct speech throughout the whole narration. Instead, they switched between direct and indirect speech (Wade & Clark, 1993). These results led us to hypothesize that the properties of upcoming utterances may play a role in how the language production system selects between these two forms of reported speech. Therefore, the goal of the current study was to explore whether the characteristics of an utterance can affect how it would be reported. We will discuss two factors that are derived from the existing literature.

The first factor is the vividness of non-verbal features accompanying the original utterance, which are incorporated during the macro-planning stages of language production (Levelt, 1993). In macro-planning, the conceptualizer selects the verbal and/or non-verbal information that is expected to achieve the current communicative intention and determines which modalities of expression should be involved (de Ruiter, 2000; Levelt, 1993). Why would narrators include non-verbal information in narratives? One basic premise about narratives is that narrators must tell a story that is worth listening to (Labov, 1982). When conveying non-verbal information, narrators directly demonstrate to others what the event looks like, sounds like or feels like (Clark & Gerrig, 1990) and can further modify or dramatize the voice or gestures of the character to make the narration more engaging (Clark, 2016). Therefore, conveying non-verbal messages is an effective way to create a reportable (Labov, 1982) or tellable (Sacks, 1992) story. We speculated that if the original utterance is accompanied by vivid non-verbal information, participants are more likely to include that non-verbal information and therefore use direct speech instead of indirect speech.

The second factor is the utterance type. Direct speech is constructed as a main clause, and it has a rather loose grammatical structure (Wilkinson et al., 2010). The to-be-reported content is directly attached to the reporting word (e.g., *say*), without any restrictions (e.g., *Neil said: "Tea? Sure!"*). However, indirect speech is constructed as a subordinate clause and must include all the obligatory constituents of a full sentence (Mayes, 1990). As a result of this constraint, some constructions cannot occur in indirect speech (e.g., **Neil said that tea? Sure.*). These constructions are called Main Clause Phenomena (MCP) (Banfield, 1973; Green, 1976): constructions that are grammatical in main clauses, but ungrammatical or much less acceptable in subordinate clauses (Green, 1976). MCP include discourse particles (e.g., "Well", "OK"), rhetorical questions (e.g., "You don't know?"), tag questions (e.g., "See, you don't ask me things like that, do you?"), truncations (e.g., "Tea? Sure."), vocatives (e.g., "John!") and exclamations (e.g., "Gosh!") (Holt, 1996; Mayes, 1990). We hypothesized that if the to-be-reported utterance can be considered one of the Main Clause Phenomena, the reporter would probably use direct speech instead of indirect speech.

Previous studies have shown that people use indirect speech to deliver background information and use direct speech to highlight the peak in a narrative (Holt, 1996). However, no studies have investigated whether there are utterance-level reasons for using direct and indirect speech in a narrative context. Answering this question is important for at least three reasons. First, the fact that people shift back and forth between direct and indirect speech indicates that there might be utterance-level reasons for using one or the other. However, to the best of our knowledge, no research has investigated this question empirically. Second, the current study investigates factors that influence structure choices during language production. How the language production system makes the decision on utterance structures has been a crucial question in the field of language production (Bock & Warren, 1985; Solomon & Pearlmutter, 2004). Previous studies have shown that the final form of an utterance is constrained by many factors, such as the accessibility of concepts and qualities of the visual environment (Bock et al., 1992; Montag & MacDonald, 2014). Our study aims to further explore whether non-verbal information and the structure of the to-be-reported utterance can influence the choice between direct and indirect speech. Investigating factors that shape the speaker's choice between these two reporting styles helps create a more comprehensive understanding of the processes involved in language production, given that direct and indirect speech are an essential part of everyday communication (Clark, 2016). Third, the decision regarding utterance forms has been considered as a mechanism of grammatical encoding stage in the formulator (Levelt, 1993). As described before, the conceptualizer selects information according to the communicative goal and decides in which modality this information shall be expressed. If we find that non-verbal information plays a role in deciding which reporting method to use, we can provide tentative evidence that at least part of the final form (i.e., the utterance is constructed as direct or indirect speech) is constrained at an earlier stage: the macro-planning stage.

We conducted the current study based on the considerations described above. We provided participants with four movie clips and asked them to recount those clips. They watched one clip at a time and started to recount immediately after watching. We analyzed both the dialogues in the movies and participants' reconstructions. This approach allows us to examine how the properties of to-be-reported utterances influence the form in which they are reported. The in principle accepted stage 1 manuscript was registered at https://osf.io/8stng/?view_only=597f32fb58ae4000bdbba45c30532f6e. No data collection and analyses were performed prior to the registration.

Method

Prior Power Analysis

We conducted a pilot study with N = 23 participants to estimate the power of these three factors: (a) utterance type, (b) voice, and (c) facial expressions. The expected effect sizes and parameter estimates for the predictors were based on the data from a pilot study in which we predicted the type of speech from this set of predictors for 23 students with an average of 48 observations per participants (range: 15-77). Following the methodology described below, participants were asked to complete four narrative tasks, in which they produced an average of 48 reported speech tokens. As predicted, utterances with vivid voice and vivid facial ex-

pressions were more likely to be reported in direct speech. Also, utterances that belonged to the class of Main Clause Phenomena were more likely to be reported using direct speech. We ran a power analysis in R using the MLPowSim program by Browne et al. (2009) for a logistic regression model to estimate the number of participants and items. This priori power analysis showed that for the three predictors a power > 0.80 could be achieved with 50 participants with 250 observations per participant, 100 participants with 150 observations per participant, 150 participants with 100 observations per participant, or 250 participants with fewer than 100 observations per participant. It is difficult to control the number of utterances a participant produces due to the nature of the narration production task. In order to ensure we would have enough observations, we set out to collect a maximum of 250 participants. Given the large amount of work on transcribing and coding, sequential analyses were carried out along with the data collection. Sequential analyses allow us to conduct a well-powered study while providing the possibility of collecting fewer participants. The spending function developed was used to calculate the adjusted alpha level (Reboussin et al., 2000). This spending function does not require an equal number of participants between each interim analyses. We decided to perform the first and the second interim analyses after collecting 80 (about one third of the maximum sample size) and 160 (about two-thirds of the maximum sample size) valid participants. The adjusted alpha boundaries for the first and second interim analyses were 0.016 and 0.032, respectively (Reboussin et al., 2000). If the *p* values of the three predictors were all smaller than 0.016 in the first interim analyses, data collection would be terminated. Otherwise, data for another group of valid 80 participants would be collected. If the *p* values of the three predictors in the second interim analyses all fell below 0.032, data collection would be terminated. If not, a final valid 90 participants would be collected. All materials can be found online (https://osf.io/8stng/?view_only=597f32fb58ae4000bdbba45c30532f6e).

Participants

Utterance rating task. 22 participants (12 females, mean age = 18.59 years, aged 18–21 years) were recruited for the rating task. Participants were reimbursed with 0.75–hour course credit.

Narrative task. The first interim analyses showed that utterance type had a significant influence on the use of direct and indirect speech. The vividness of voice and facial expressions did not have an effect. Therefore, the second interim analysis was performed according to the preregistered plan. The results showed that utterance type and the vividness of voice influenced the choice between direct and indirect speech. We did not observe any effect of the vividness of facial expressions. Therefore, a final 90 valid participants were recruited, which resulted in a total of 250 English native speakers (117 females, 7 others, mean age = 31.71 years, aged 18–50 years) recruited from Prolific, an online participants recruitment platform. They were paid £ 4.38 for their participation. All participants signed an informed consent form prior to participation to give consent for audio and video recording. This study was approved by the Ethics Committee of Psychology at the Erasmus University Rotterdam.

Materials

Four movie clips of approximately three minutes each, taken from "Breakfast at Tiffany's" (3:01), "A Beautiful Mind" (3:03), "Dead Poets Society" (2:51) and "Diner" (2:50), were used in the experiment. The clip "Breakfast at Tiffany's" portrayed a conversation between three characters: two young people and a shop assistant at a jewelry store. The clip "A Beautiful Mind" portrayed a conversation between two characters: a woman and her husband who was in a psychiatric hospital. The clip "Dead Poets Society" portrayed a conversation between a teacher and a student who visited the teacher to ask for advice. In the clip "Diner", a male and a female character argued about the arrangement of records. All movie clips can be easily understood without background information. We selected clips with only two or three characters because too many characters might make it difficult for participants to remember "who said what", which is important in our study. We chose clips that focus more on talk than on action because of our study's focus on reported speech.

Procedure

Utterance rating task. Dialogues from the four movie clips were transcribed. Then, the transcripts were segmented into utterances. The separation procedure was performed by two coders following conventional sentence boundaries and intonation contour. Sentence fragments, repetitions, and incomplete sentences were considered as separate utterances. Lexical fillers, such as "well", "I mean", "you know", and "let us see" were treated as separate utterances if they occurred at the beginning or end of another utterance. If they occurred within an utterance, they were treated as being part of that utterance (Dijkstra et al., 2004; Lyons et al., 1994). After segmentation, these utterances were rated on three dimensions: vividness of voice (continuous), vividness of facial expressions (continuous), and utterance type (categorical).

Ten participants were instructed to rate the vividness of voice. Another ten participants were instructed to rate the vividness of the facial expressions. Each participant finished the task individually in a sound-attenuated room. After seated in front of a computer, they were handed a pencil and a paper rating scale with the to-be rated utterances on it. To facilitate ratings, movie clips were segmented into short pieces that lasted approximately five seconds. For the participants who rated the vividness of voice, they were asked to pay attention to the character's voice. Specifically, they were instructed to answer the question "How vivid do you find the voice of the character while producing this utterance" and indicate their answers on a five-point scale ranging from "not vivid at all" to "highly vivid." The rating procedure was the same for the facial expressions with the only difference being that participants were instructed to focus on the character's facial expressions.

Two trained judges naive to the purpose of the experiment coded the utterances from the movies as "one" if the utterance belonged to the class of Main Clause Phenom-

ena, and with "zero" if it did not. The inter-rater reliability with Kappa coefficient was 0.89, which indicated a relatively high agreement between two coders (Landis & Koch, 1977). Disagreements between the two coders were discussed and resolved before later analyses.

Narrative task. Participants were asked to finish the task in a quiet and non-distracting environment. Overall, participants were asked to finish four narrative tasks. They were first instructed to watch one movie clip carefully so that they could provide a detailed account of what happened in the movie. The movie was shown on a computer screen. After viewing each clip, they immediately began to recount. To induce elaborate narrations, we asked participants to retell the clip as if they were telling the story to someone who is not watching. Upon completion of the retelling of one clip, participants took a rest for two minutes before they started to watch and recount the next movie clip. The order of presentation of the movie clips was counterbalanced across participants. All narrations were videotaped. The whole procedure lasted approximately 40 minutes.

Analysis

Exclusion criteria. Participants whose narrations were not recorded because of a recording device malfunction were excluded from the analysis. Narrations that did not contain direct or indirect speech were also excluded. In total, the data from 38 participants were excluded due to the device malfunction and 180 narrations were excluded because no reported speech was included.

Transcription and coding procedure. All recordings obtained from the narrative task were transcribed verbatim for coding. The coding procedure consisted of two steps. In step one, two trained coders categorized each reported speech from participants' narrations as either direct speech or indirect speech. Three grammatical criteria were used for distinguishing direct versus indirect speech. The first one was the deictic words. The deictic words (e.g., I/she; this/that; here/there) in indirect speech (e.g., He said that he thought it would be very smart.) were paraphrased according to the current speaking situation while the deictic words in direct speech (e.g., He said: "I think it would be very smart.") were the same as in the reported situation. The second one was the verb tense. Like deictic words, the verb tense in indirect speech (e.g., She said that she didn't know.) should be adjusted to the current reporting context while the verb tense in direct speech (e.g., She said: "I don't know.") remained unchanged (Li, 1986). The last one was the absence/presence of the complementizer "that"[1]. In indirect speech (e.g., She said that there's no William Parcher.), the reported content was introduced by "that" while there was no complementizer in direct speech (e.g., She said: "There's no William Parcher."). There were 89 utterances that could not be classified by the above-mentioned criteria, the coders listened to the recording for speaker's intonation. If there was any change in the speaker's voice compared to her/his normal voice, this utterance was coded as direct speech. Otherwise, it was treated as indirect speech (Nordqvist, 2001; Wade & Clark, 1993).

In step two, these two judges identified the utterance from the movie dialogue to which the reported speech corresponded. If an utterance from the movie was reported using direct speech, a value of "one" was assigned to that utterance. If this utterance was reported using indirect speech, a value of "zero" was assigned. All utterances were coded by two coders individually. Kappa coefficients were computed to assess the agreement between coders. In step one we achieved a substantial interrater reliability with a Kappa coefficient of 0.81. In step two we achieved an interrater reliability with a Kappa coefficient of 0.87. Coding disagreements were resolved by a discussion between coders before analyses.

Data analysis and results: Mixed effects logistic regression model. The data were analyzed using a mixed effects logistic regression model with the generalized linear mixed model function in R (Bates et al., 2015). The dependent measure was a categorical variable coding whether an utterance was reported in direct speech or indirect speech. Fixed effects included the independent variables: the vividness of voice, the vividness of facial expressions and the utterance type. We also included random intercepts for participants and items. The analyses revealed a significant effect of the vividness of voice (β = 0.51, 95% *CI* [0.12; 0.90], *SE* = 0.20, *Z* = 2.60, $p < 0.01$, odds ratio = 1.67, 95% *CI* [1.13; 2.46]), which means participants were 1.67 times more likely to use direct speech with one point increase (e.g., from 3 to 4) on the vividness scale. The utterance type had a main effect (β = 1.32, 95% *CI* [0.80; 1.83], *SE* = 0.26, *Z* = 5.02, $p < 0.001$, odds ratio = 3.73, 95% *CI* [2.23; 6.24]). Utterances that belonged to the Main Clause Phenomena were 3.73 times more likely to be reported in direct speech. There was no significant effect of the vividness of facial expressions (β = -0.11, 95% *CI* [-0.50; 0.29], *SE* = 0.20, *Z* = -0.53, $p > 0.05$, odds ratio = 0.90, 95% *CI* [0.60; 1.34]). Table 1 summarizes the model.

Discussion

The current study aimed to investigate utterance-related factors that influence people's use of direct and indirect speech in a narrative task. Participants were asked to watch and retell short movie clips. The results showed that utterances accompanied by vivid voice were more likely to be reported in direct speech. The vividness of facial expressions

1 The complementizer "that" can sometimes be omitted in indirect speech. The criterion "absence/presence" of "that" alone is not enough to determine whether an utterance is direct or indirect speech. Therefore, we will take this criterion into account only when the deictic terms and verb tenses are the same in both direct and indirect speech. In most cases, direct speech and indirect speech can be differentiated by deictic terms and verb tenses.

Table 1. Effects of the Vividness of Voice, the Vividness Facial Expressions and Utterance Type on the Use of Direct and Indirect speech.

	Estimate	*SE*	*Z*	*p*	95% CI	
					Lower bound	Upper bound
Fixed effects						
Intercept	-1.72	0.60	-2.88	0.004	-2.89	-0.55
Voice	0.51	0.20	2.60	0.009	0.12	0.90
Utterance type	1.32	0.26	5.02	<0.001	0.80	1.83
Facial expressions	-0.11	0.20	-0.53	0.597	-0.50	0.29
Random effects	Variance	*SD*				
Participant	3.61	1.90				
Item	1.75	1.32				

did not affect the choice between direct and indirect speech. The utterance type had an influence on the use of direct and indirect speech. Utterances that belonged to the Main Clause Phenomena were more frequently reported in direct speech than in indirect speech. Taken together, this experiment showed that both the non-verbal information accompanying the original utterances and the structures of the original utterances have impacts on how likely utterances will be reported directly.

Existing evidence shows that the rates of direct speech in communication are influenced by the aims and contexts of communication. People produce relatively more direct speech for an amusement purpose and in less formal contexts (Koppen et al., 2019; Wade & Clark, 1993). What remains unclear is why people shift between direct and indirect speech on an utterance level.

The present study expands on previous studies in that we found that the choice of direct and indirect speech can be partially explained by utterance level reasons. First, people are more prone to report directly when the original utterances are accompanied by vivid non-verbal information, specifically, by vivid voice. This finding is consistent with the view of the demonstration theory (Clark, 2016). According to Clark (2016), direct speech is an act of demonstration that mainly relies on auditory, visual, and tactile knowledge of physical scenes. Direct speech is associated with a frequent use of demonstrations from both auditory and visual channels, whereas indirect speech is associated with a less frequent use of demonstrations (Blackwell et al., 2015; Stec et al., 2016). Direct speech, unlike indirect speech, is capable of conveying non-verbal information that accompanied previous utterances. This property of direct speech makes it a better candidate when people wish to deliver the non-verbal aspects of the original utterances in narrations than does indirect speech.

As mentioned earlier, even though direct speech makes stories more vivid and involving, people do not use direct speech throughout the whole narration. Actually, only using direct speech in a narration will likely impose an extra cognitive load on listeners (Köder et al., 2015), given that it requires them to constantly change vantage point to comprehend the story. Therefore, direct speech occurs more frequently at the climax of a story (Mayes, 1990). In this study, we found that if the original utterances contain vivid non-verbal information, then participants are more likely to convey the non-verbal information along with the verbal information to enhance the story. However, there is an important caveat when interpreting this result. Our finding can only reveal part of the picture. We found that the utterance with a more vivid voice will be reported more often in direct speech than in indirect speech in a narrative context. This result might not hold for other contexts such as a writing task in which no non-verbal information is involved or a courtroom testimony setting where the main function of direct speech is evidentiality (Chaemsaithong, 2017). The decision between direct and indirect speech is highly flexible and is subject to be influenced by contextual factors. It will be an interesting topic for future studies to investigate factors that account for the use of direct and indirect speech in various other settings.

Contrary to our hypothesis, we did not observe the effect of the vividness of facial expressions on the use direct and indirect speech. We propose two, not mutually exclusive, explanations for this null result. First, it is possible that the effect size of facial expressions is too small to be detected. Given the large sample size in this study, we would expect to detect an influence of the vividness of facial expressions if there is a medium to large effect size. Direct speech is a selective depiction of original utterances (Clark, 2016). This means that not every aspect from the original utterances will be conveyed. Empirical evidence shows that differences exist in the use of non-verbal information from difference modalities. For example, Stec et al. (2016) found that character's intonation and facial expressions occurred more frequently than gestures in direct quotations. In addition, speakers used multimodal depictions when quoting others, whereas self-quotations were more often accompanied by depiction of one modality (Stec et al., 2017). These results are in line with Clark's (2016) view that people selectively depict non-verbal information from previous utterances. The second explanation for the null result is that the monologue setting we created might make it difficult to detect the effect of facial expressions. Existing evidence shows that facial portrayals happen more often in dialogue conditions (face-to-face and telephone communication) than in a monologue condition (Bavelas et al., 2014). It is possible

that the effect of the vividness of facial expressions will be more significant in a dialogue setting, but this is something that could be examined in future studies.

In accordance with our prediction, utterance type also plays a role in deciding how likely an utterance will be reported in direct speech or indirect speech. An utterance that belongs to the Main Clause Phenomena is more likely to be reported in direct speech. This is due to the fact that direct speech has a relative loose sentence structure. The quoted content can be directly placed after the quoting verbs (i.e., say) without any restrictions. Utterances that belong to the Main Clause Phenomena are grammatically correct in direct speech but are incorrect or less acceptable in indirect speech. Therefore, people are more likely to convey them in the form of direct speech. Our finding falls in line with work from Mayes (1990), who also found that direct speech is used when the structures are grammatically incorrect in indirect speech. Due to the relatively lower grammatical complexity of direct speech, people with language deficits benefit from the use of direct speech. For example, aphasic people were found to use direct speech more often than normal people (Groenewold et al., 2013).

The decision between direct and indirect speech can be explained by a current language production model. The language production theory proposed by Levelt (1993) proposed that the production of language can be divided into several subprocesses. The conceptualizer and grammar encoder are of relevance with the current study. The conceptualizer orders information to be expressed to achieve communication goals. The syntactic structure of selected information will be later determined in the grammar encoder (Levelt, 1993). Levelt's model is targeted at the production of verbal messages. It therefore does not explain how non-verbal information is produced. Therefore, this model was extended later by researchers to accommodate the production of non-verbal information such as gestures. It is proposed that the conceptualizer not only selects information whose expression will fulfill the communication goal but also decides in which channel or modality information shall be expressed (de Ruiter, 2000). Returning to the production of direct and indirect speech, if the conceptualizer selects to convey non-verbal information, the utterance will more likely be in the form of direct speech, given that indirect speech is not capable of delivering non-verbal information.

Limitations

There are a few limitations to our findings that may limit the generalizability of the results. First of all, as mentioned earlier, the monologue setting used in this experiment might not be powerful enough to detect the effect of the vividness of facial expressions on the use of direct and indirect speech. The rate of demonstrations in a conversation is sensitive to speaking contexts. Demonstration is an act of communication that is designed for others to directly experience the depicted event. Therefore, the absence of an interlocutor has been observed to significantly reduce the frequency of direct speech (Bavelas et al., 2014). If we could increase the rate of direct speech, we might be more likely to detect the effect of facial expressions. Future studies could evaluate the effects of non-verbal information, especially facial expressions, in a dialogue context or a more interactive context.

The second limitation is that we only examined two types of non-verbal information in this study. Except for voice and facial expressions, gestures, gazes, even the lips, and nose movement can be depicted in direct speech (Cooperrider & Núñez, 2012). It will be interesting to examine the effects of non-verbal information from other modalities. Our intuition is that other non-verbal information also contributes to the decision between direct and indirect speech. Future studies could design experiments that are more sensitive to detect the effects of non-verbal information from other modalities.

In summary, the current findings improve our understanding in that we found that the use of reported speech is more complicated than we already knew. Except for the aim of reporting and reporting contexts, utterance-level factors account for the use of direct and indirect speech as well. Both the vividness of non-verbal information that accompanying the original utterances and the structures of the original utterances have an influence on in which form the utterances will be reported.

Author Contributions

Contributed to conception and design: JL, KD, and RAZ
Contributed to acquisition of data: JL
Contributed to analysis and interpretation of data: JL, JJ, and RAZ
Drafted and/or revised the article: JL, JJ, KD, and RAZ
Approved the submitted version for publication: JL, JJ, KD, and RAZ

Acknowledgments

This research was supported by the scholarship from the China Scholarship Council (201706750007).

Competing Interests

The authors have no competing interests to declare.

Data Accessibility Statement

The stage 1 report, all experimental materials, participant data and analysis scripts (R scripts) can be found on this paper's project page on the OSF (https://osf.io/8stng/?view_only=597f32fb58ae4000bdbba45c30532f6e).

Submitted: October 25, 2021 PDT, Accepted: March 08, 2022 PDT

References

Banfield, A. (1973). Narrative style and the grammar of direct and indirect speech. *Foundations of Language*, *10*(1), 1–39.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bavelas, J., Gerwing, J., & Healing, S. (2014). Effect of dialogue on demonstrations: Direct quotations, facial portrayals, hand gestures, and figurative references. *Discourse Processes*, *51*(8), 619–655. https://doi.org/10.1080/0163853x.2014.883730

Blackwell, N. L., Perlman, M., & Fox Tree, J. E. (2015). Quotation as a multimodal construction. *Journal of Pragmatics*, *81*, 1–7. https://doi.org/10.1016/j.pragma.2015.03.004

Bock, J. K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, *99*(1), 150–171. https://doi.org/10.1037/0033-295x.99.1.150

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*(1), 47–67. https://doi.org/10.1016/0010-0277(85)90023-x

Browne, W. J., Steele, F., & Golalizadeh, M. (2009). *MLPowSim*. http://www.bristol.ac.uk/cmm/software/mlpowsim/

Chaemsaithong, K. (2017). Speech reporting in courtroom opening statements. *Journal of Pragmatics*, *119*, 1–14. https://doi.org/10.1016/j.pragma.2017.08.003

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123*(3), 324–347. https://doi.org/10.1037/rev0000026

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language*, *66*(4), 764–805. https://doi.org/10.2307/414729

Cooperrider, K., & Núñez, R. (2012). Nose-pointing: Notes on a facial gesture of Papua New Guinea. *Gesture*, *12*(2), 103–129. https://doi.org/10.1075/gest.12.2.01coo

Coulmas, F. (Ed.). (1986). *Direct and indirect speech*. De Gruyter Mouton. https://doi.org/10.1515/9783110871968

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge University Press.

Dijkstra, K., Bourgeois, M. S., Allen, R. S., & Burgio, L. D. (2004). Conversational coherence: Discourse analysis of older adults with and without dementia. *Journal of Neurolinguistics*, *17*(4), 263–283. https://doi.org/10.1016/s0911-6044(03)00048-4

Glock, N. (1986). The use of reported speech in Sacamaccan discourse. In G. Huttar & K. Gregerson (Eds.), *Pragmatics in Non-Western Perspective* (pp. 35–61). Summer Institute of Linguistics.

Green, G. M. (1976). Main clause phenomena in subordinate clauses. *Language*, *52*(2), 382–397. https://doi.org/10.2307/412566

Groenewold, R., Bastiaanse, R., & Huiskes, M. (2013). Direct speech constructions in aphasic Dutch narratives. *Aphasiology*, *27*(5), 546–567. https://doi.org/10.1080/02687038.2012.742484

Groenewold, R., Bastiaanse, R., Nickels, L., & Huiskes, M. (2014). Perceived liveliness and speech comprehensibility in aphasia: The effects of direct speech in auditory narratives. *International Journal of Language & Communication Disorders*, *49*(4), 486–497. https://doi.org/10.1111/1460-6984

Holt, E. (1996). Reporting on talk: The use of direct reported speech in conversation. *Research on Language & Social Interaction*, *29*(3), 219–245. https://doi.org/10.1207/s15327973rlsi2903_2

Köder, F., Maier, E., & Hendriks, P. (2015). Perspective shift increases processing effort of pronouns: A comparison between direct and indirect speech. *Language, Cognition and Neuroscience*, *30*(8), 940–946. https://doi.org/10.1080/23273798.2015.1047460

Koppen, K., Ernestus, M., & van Mulken, M. (2019). The influence of social distance on speech behavior: Formality variation in casual speech. *Corpus Linguistics and Linguistic Theory*, *15*(1), 139–165. https://doi.org/10.1515/cllt-2016-0056

Labov, W. (1982). Speech Actions and Reactions in Personal Narrative. In D. Tannen (Ed.), *Analyzing Discourse: Text and Talk* (pp. 219–247). Georgetown University Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Larson, L. (1978). *The functions of reported speech in discourse*. The Summer Institute of Linguistics.

Levelt, W. J. M. (1993). *Speaking: From intention to articulation*. MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001

Li, C. (1986). Direct and indirect speech: A functional study. In C. Coulmas (Ed.), *Direct and indirect speech* (pp. 29–45). De Gruyter Mouton.

Lyons, K., Kemper, S., LaBarge, E., Ferraro, F. R., Balota, D., & Storandt, M. (1994). Oral language and Alzheimer's disease: A reduction in syntactic complexity. *Aging, Neuropsychology, and Cognition*, *1*(4), 271–281. https://doi.org/10.1080/13825589408256581

Macaulay, R. K. S. (1987). Polyphonic monologues. *IPRA Papers in Pragmatics*, *1*(2), 1–34. https://doi.org/10.1075/iprapip.1.2.01mac

Mayes, P. (1990). Quotation in spoken English. *Studies in Language*, *14*(2), 325–363. https://doi.org/10.1075/sl.14.2.04may

Montag, J. L., & MacDonald, M. C. (2014). Visual salience modulates structure choice in relative clause production. *Language and Speech*, *57*(2), 163–180. https://doi.org/10.1177/0023830913495656

Nordqvist, A. (2001). The use of direct and indirect speech by 1½-to 4-year-olds. *Psychology of Language and Communication*, *5*(1), 57–66.

Reboussin, D. M., DeMets, D. L., Kim, K., & Lan, K. K. G. (2000). Computations for group sequential boundaries using the Lan-DeMets spending function method. *Controlled Clinical Trials*, *21*(3), 190–207. https://doi.org/10.1016/s0197-2456(00)00057-x

Sacks, H. (1992). *Lectures on conversation*. Basil Blackwell.

Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology*, *49*(1), 1–46. https://doi.org/10.1016/j.cogpsych.2003.10.001

Stec, K., Huiskes, M., & Redeker, G. (2016). Multimodal quotation: Role shift practices in spoken narratives. *Journal of Pragmatics*, *104*, 1–17. https://doi.org/10.1016/j.pragma.2016.07.008

Stec, K., Huiskes, M., Wieling, M., & Redeker, G. (2017). Multimodal character viewpoint in quoted dialogue sequences. *Glossa: A Journal of General Linguistics*, *2*(1). https://doi.org/10.5334/gjgl.255

Wade, E., & Clark, H. H. (1993). Reproduction and demonstration in quotations. *Journal of Memory and Language*, *32*(6), 805–819. https://doi.org/10.1006/jmla.1993.1040

Wierzbicka, A. (1974). The semantics of direct and indirect discourse. *Paper in Linguistics*, *7*(3–4), 267–307. https://doi.org/10.1080/08351817409370375

Wilkinson, R., Beeke, S., & Maxim, J. (2010). Formulating actions and events with limited linguistic resources: Enactment and iconicity in agrammatic aphasic talk. *Research on Language & Social Interaction*, *43*(1), 57–84. https://doi.org/10.1080/08351810903471506

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/33631-the-influence-of-utterance-related-factors-on-the-use-of-direct-and-indirect-speech/attachment/85286.docx?auth_token=uAQLrT9d-b356VUZTui0