Designing Referential Descriptions for Children, Young Adults, and Computers: A Comprehensive Examination of Talker Informativity

Julie Bannon (bannonj@mcmaster.ca)
Department of Psychology, Neuroscience and Behaviour, McMaster University

Raheleh Saryazdi (raheleh.saryazdi@mail.utoronto.ca)
Department of Psychology, University of Toronto

Craig G. Chambers (craig.chambers@utoronto.ca)
Department of Psychology, University of Toronto

Abstract

Research on referential communication has explored talkers' ability to tailor descriptions for the current context. The present study examines this issue alongside talker adaptations for different addressees. Participants were asked to provide a child, adult, or computer with instructions to select and move objects on a display. Each target object was either unique or accompanied by a same-category competitor. Targets in the latter condition could be differentiated with either a modifier or subordinate term. In addition to examining speech onset latencies, we analyzed referential descriptions for informational adequacy (just enough, underinformative, overinformative), noun type (basic-level or subordinate), and incidence/type of modifiers. The most noticeable effects were observed when addressing children, with participants using more basic terms and more modifiers (particularly color). These results reveal the spontaneous adaptation of referential strategies according to audience type, providing evidence for models of language in which speakers actively consider addressees' needs and cognitive abilities.

Keywords: referential communication; audience design; informativity

Introduction

Effective communication requires speakers to design expressions that are both accurate and easily interpretable. A key challenge for speakers is to further ensure that the information conveyed is appropriate for their intended addressee. In the current research, we address this issue by examining how speakers adjust their referential expressions depending on perceived capabilities of the listener.

It is well established that people make adjustments to their speech depending on their target audience. However, research in this area has traditionally focused on the various acoustic-level modulations that speakers make, especially when addressing young children. For example, when communicating with infants, speakers typically vary their vocal affect, pitch, and vowel articulations (e.g., Fernald & Simon, 1984; Kitamura & Burnham, 2003; Kuhl et al., 1997). Further, evidence suggests that speakers often slow their rate of speech and hyperarticulate during error correction when speaking to a *computer* (e.g., Oviatt et al.,1998; Stent et al., 2008). Given recent technological advancements that allow speech communication with interactive systems such as smart home devices, it is important to consider how people treat such systems as a communicative partner. Interestingly, there is evidence to suggest that speech adaptations made to interactive systems are not always different than adaptations made to human partners. For example, Burnham and colleagues (2010) found that speakers produce similar hyperarticulations of corner vowels when addressing both infants and computers. This suggests that speakers make adjustments they believe meet the needs of their audience, regardless of whether their listener is another person or an artificial communication system. These acoustic level changes in response to an addressee reflect spontaneous adaptions to a listener's perceived abilities. However, it is also important to consider the lexical adaptations speakers make, especially when comparing across audiences with different information processing capabilities, such as children and computer systems.

The informational adequacy of a speaker's expressions is an important aspect of communicative efficiency. Speakers must consider their addressee's level of knowledge in order to produce descriptions that can be easily understood. To date, there has been considerable debate regarding the appropriate level of information to include in referential descriptions. According to Grice's Maxim of Quantity (1975), speakers should provide only the precise amount of information needed to convey an intended message. In the context of referential communication, this means that speakers should only include additional information such as reference to object properties in cases where there is potential for ambiguity (e.g., saying 'the *red* bowl' when there are two bowls in a display). However, a number of studies have shown that speakers include pre-nominal modifiers in their descriptions even when there is no ambiguity in the visual scene (e.g., when all objects belong to separate categories and could easily be differentiated; e.g., Rubio-Fernandez, 2016; Tarenskeen et al., 2015). This behavior results in technically overinformative referential expressions, in which the speaker includes more information than is necessary to identify the target. Although such expressions are often considered to be infelicitous according to Grice's Maxim of Quantity, it has been argued that overinformative descriptions, especially those that involve color, can be a sign of cooperativeness in communication. Rubio-Fernandez (2019) suggests that including redundant modifiers in referential descriptions can help guide a listener's search for an object in a display. In this way, speakers are being cooperative by providing additional information about the target (see also Rubio-Fernandez, 2016). However, there is conflicting evidence as to whether overinformative descriptions actually help the listener to quickly identify the target. Although some studies suggest that they are facilitative (e.g., Arts et al., 2011; Tourtouri et al., 2019), others have found that overinformative expressions can impede comprehension (e.g., Engelhardt et al., 2011). Most research on this topic however has focused on the young adult population.

Research on children's perception of informational content suggests that school-aged children are sensitive to the inclusion of redundant information. For example, Davies and Katsos (2010) found that children judged overinformative references to be infelicitous (although this was not the case when a binary scale was used). However, the overinformative descriptions presented included state (e.g., *the closed umbrella*) or size modifiers (e.g., *the thin nail*), which are not typically used in cases where there is no potential for ambiguity. Instead, evidence from the adult literature suggests speakers are more likely to use color in overinformative descriptions (e.g., Tarenskeen et al., 2015), rather than state or size modifiers. It has also been shown that younger children aged 6-7 years demonstrate a preference for objects that are described with redundant color modifiers; however, this preference is not found in older children aged 9-10 years old (Koolen et al., 2016). Given the limited vocabulary of younger children, speakers might use redundant modifiers to a greater extent as a way to help with referential identification. To accomplish this, speakers may rely more on color terms, which not only capture attention but are also simple in terms of vocabulary. Yet another question is whether similar patterns are observed when designing referential expressions for a computer addressee. The existing evidence suggests that the incidence of overinformative expressions is greater when speakers are addressing a human partner versus a computer (Maes et al., 2007). However, there is also some evidence suggesting that speakers may use more words in their descriptions when communicating with artificial agents (e.g., Kriz et al., 2010). It is important to note that there are also similarities between how speakers address computers and humans. In a recent study using similar paradigm, we observed that both younger and older adults used color modifiers the most with an imagined younger and older adult addressee as well as a computer (Saryazdi et al., 2019). This evidence indicates that speakers may perceive the cognitive capabilities of an artificial system to be similar to those of a human partner.

In addition to using modifiers to identify a target, speakers often have a choice in the type of noun they select for a particular object. For example, many real-world objects can be effectively referred to using either subordinate or basic-level terms (e.g., Dalmatian vs. dog). To date, there has been limited formal investigation into the use of subordinate terms. However, this is an emerging area of research, with a growing number of recent studies exploring the role of subordinate terms in referential communication (e.g., Degen et al., 2020; Frisson & Murphy, 2019; Saryazdi et al., 2019). Degen and colleagues (2020) found that, although people prefer to use basic-level terms overall, they increase their use of subordinate terms when a more specific label is necessary for target identification. This is in line with the evidence that listeners prefer subordinate terms to be used in situations where there are two same-category objects in a display (Frisson & Murphy, 2019). These results suggest that speakers are spontaneously sensitive to the specificity of subordinate terms and expect them to differentiate between two objects of the same category. However, the use of subordinates may not be appropriate in all contexts. When communicating with children who face significant vocabulary constraints, it might be beneficial to differentiate objects using a basic-level term combined with an adjective in place of a subordinate term. On the other hand, subordinate terms may be unproblematic when communicating with a computer, which presumably has an adult-like vocabulary. It is therefore possible that speakers will show different strategies in the use of object labels as well as modifying adjectives in tailoring messages for an addressee.

Along with variation in lexical content, speakers also show differences in speech onset latencies depending on the properties of the visual scene. For example, previous research has found that speakers are slower to initiate their speech when there are more objects present in the display (e.g., Gatt et al., 2017). Onset latencies are typically considered to reflect the planning time of an utterance. In the current context, we consider whether adaptations to the different addresses, in addition to the potential for ambiguity in the visual display, will impact the time speakers take to plan the referential expressions.

Present Study

In the present study, we investigate how speakers design referential expressions depending on the perceived processing abilities of their listener. To accomplish this, we examine speakers' referential choices when communicating with a school-aged child, younger adult, or a computer. Given the existing evidence that speakers make vocal adaptations to both child and computer addressees, we measure whether similar patterns are observed with informational content. One major component in designing utterances for different audiences is the information assumed to be available to the audience. This includes considering the linguistic knowledge of a communicative partner. School-aged children are expected to have a limited working vocabulary compared with young adults and with automated dialogue systems, which can be programmed to have a large lexicon. The current research seeks to determine how speakers use this information about the listener to guide their referential choices. Further, we use speech onset latencies as a way to probe the effort exerted in planning referential expressions as a function of context and addressee.

In the experiment, participants were led to believe they were interacting live with a member of each particular addressee group. Specifically, they were asked to play a game with human players (child or young adult) in an adjoining room, or with the computer. Their task was to provide instructions about how to move objects on a visual display. The objects in this study were selected so they could be accurately identified by a basic-level name (e.g., *'dog'*), or by a specific subordinate term denoting a member of a particular category (e.g., *'Dalmatian'*). The presence of a same-category competitor was varied across trials to investigate how people form referring expressions when it is necessary to include either additional information or use a subordinate term compared with when a basic-level term is sufficient.

In line with previous research, we expect that individuals will use more modifiers when there is a contrast object present. However, we expect that modifier use will vary with different addressees. Because children have limited vocabulary, we predict that speakers will overspecify more frequently when addressing this group in order to draw attention to visual features. Greater use of basic-level terms is also likely to occur with the child addressees. In contrast, we expect that the use of subordinate terms will increase with the computer addressee because computers might be perceived as having a greater lexicon. In addition, and consistent with previous research, we expect color to be the most common modifier used across all groups.

Method

Participants

The final sample included 24 young adult participants (*M* = 19.33 years old, *SD* = 1.49) who were recruited from the University of Toronto Mississauga undergraduate community, and received partial course credit or $10.00 for their participation. Thirteen additional participants were dropped because they expressed suspicion that there was not another player in the next room during debriefing. One participant was also dropped due to an error in the experiment file. All participants learned English before the age of five and considered it as their dominant language.

Materials

The present study employed a game-like paradigm in which participants provided an instruction for a second player who was either a child (age 6), a younger adult (university student), or a computer equipped with speech recognition software. Each display included an image of the other 'player' on top. A workspace below each image showed four objects scattered on the left side, and a 2 X 2 grid on the right side indicating the four possible locations (see Figure 1). The target object in each display was either unique (no contrast) or accompanied by a same-category competitor object (contrast). In the contrast condition, successful identification of the target could be achieved using either a subordinate (e.g., Dalmatian) or a modified basic-level term (e.g., spotted dog).

Each participant completed 3 blocks of trials, one for each potential addressee. Each block consisted of 6 critical trials (3 in each contrast condition) and 6 filler trials, for a total of 36 trials (18 critical). The presence of a same-category competitor, as well as the order of the addressee group, was counterbalanced across lists. Age-of-acquisition norms were used to ensure that both basic-level and subordinate labels for all critical items are learned by the age of 6, and would be perceived as appropriate for a school-aged child (although these subordinate terms are known to young children, the average adult speaker may still be more likely to spontaneously avoid them when addressing a child). A microphone was placed next to the participant's screen to record their instructions. Participants were told the microphone was connected to the other player's computer. The experiment was designed and implemented with Experiment Builder software (SR Research, ON).

Procedure

Participants were told that they would be interacting with other players who were located in an adjoining room, as well as with a computer equipped with voice recognition software. After obtaining consent to participate, they were shown an example of their display prior to the start of the experiment. Participants were told that the image of each addressee at the top of the display was their current partner, with a clipart microphone representing the computer addressee.

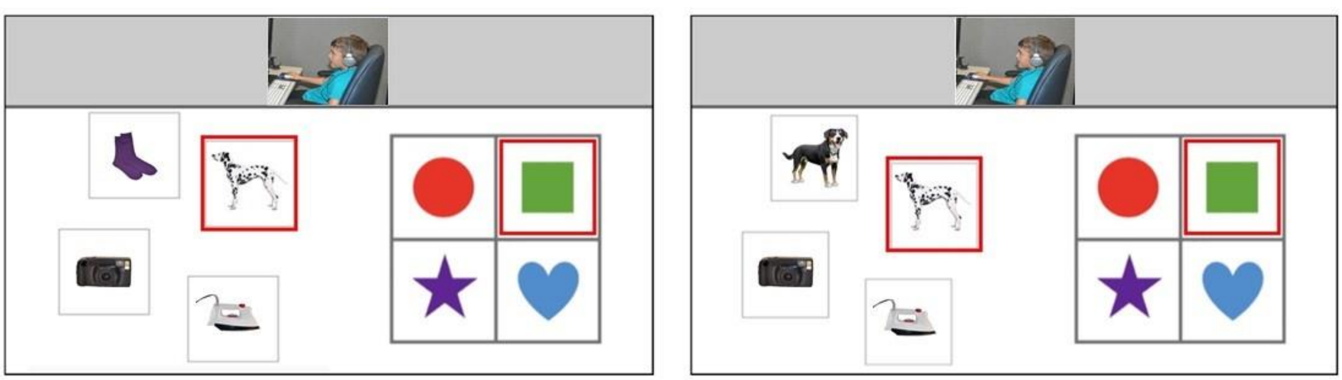

Figure 1. Example experiment display (left: no contrast condition, right: contrast condition)

Prior to initiation of the task, a second experimenter in the adjoining room gave verbal confirmation that the other player was ready to begin. At the start of each trial, a red box appeared around the target image, accompanied by a sound, to indicate to the participant which object was to be moved. After a 2 s delay, a second red box appeared around the target location. Participants were asked to provide an instruction to the other player telling them to move the target object to the location in the form of "click on the X and move it to the Y". After providing the instruction, the participant saw a green arrow appear to the right of the image of the addressee, indicating that the player understood the instruction and they were moving on to the next trial. The movement of the target object was not, however, shown on the participant's screen. If the participant did not give enough information to identify the object, they would see a red question mark appear to the left of the image of the addressee, indicating that the player requested clarification. This was only used on trials in which participants provided only a basic-level term to identify the target when a same-category competitor was present. This manipulation was included to ensure the task appeared realistic to the participant, as a basic-level term in the contrast condition would be insufficient to identify the target. The participant's computer was in fact controlled by the experimenter, who was separated from the participant by an opaque divider.

In between each block, the experimenter left the room to notify the other player that they were ready to begin the next section of the experiment. Similarly, the experimenter in the other room would come in and check if the participant was ready. For the computer addressee, the participant was told the program would load automatically, and the covert experimenter waited approximately one minute before starting the trials.

Coding Procedures

The description content for each trial was transcribed and coded by a research assistant who was blind to the condition and the addressee. The noun phrase for each target object was coded as either including a modifier or not. Each modifier was additionally coded as one of four categories: color, size, location, or other. The label for each object was coded as being either a basic-level or subordinate term. After this, conditions by trial were added to the transcriptions to allow each description to be coded as to whether it provided just enough information to identify the target, or too much or too little information. In the no contrast condition, any additional modification of the noun was considered overinformative. It is important to note that the use of subordinate terms in this condition could be considered as overinformative because a basic-level term would suffice; however, this possibility is often not discussed in previous studies of referential expressions. Thus, in the present study we consider these descriptions to be sufficient even when there is no contrasting alternative present. In the contrast condition, either a bare subordinate noun, or a modified basic-level term was considered to be sufficient. Modification of a subordinate term, or multiple descriptors used alongside a basic-level term were considered overinformative in the contrast condition. Finally, the speech onset times were measured from the appearance of the red box indicating which object the participant was to identify.

Results

All analyses were conducted with R statistical package version 3.6.1 (R Core Team, 2019). Linear mixed effect analyses were performed using lme4 package version 1.1-21 (Bates et al., 2015) and statistical significance was assessed with the lmerTest package version 3.1-0 (Kuznetsova et al., 2017). Each model included the fixed effect of contrast (contrast = 1, no contrast = -1), and addressee group, using young as the reference group (first addressee comparison: young = 1, child = -1; second addressee comparison: young = 1; computer = -1), as well as their interactions. All models included random intercept terms for participant and item, and by-participant and by-item slopes for contrast, addressee groups, and their interactions.

First, we examined differences in speech planning by analyzing the time it took for participants to initiate their utterance. As expected, speech onsets were longer for the contrast (M = 3.17 s, SD = 1.35) condition compared to the no contrast (M = 2.94 s, SD = 0.83) condition, but this difference did not reach significance (p > .05). There were also no differences observed as a function of addressee type or its interaction with contrast.

Next, we examined informational adequacy in terms of whether participants provided "just enough", "overspecified", or "underspecified" descriptions. Figure 2 shows the breakdown of information level as a function of contrast and addressee type. For the analysis, we conducted a logistic mixed effect model in which we compared the incidence of descriptions with "just enough" information with non-

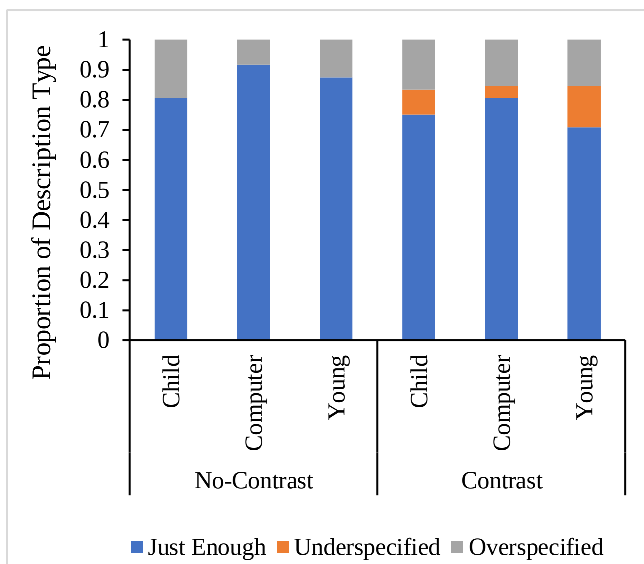

Figure 2. Informational adequacy by condition

optimal descriptions (both underspecified and overspecified descriptions were treated as non-optimal in this analysis). The results revealed a main effect of contrast condition with greater incidence of optimal descriptions in the no contrast vs. contrast condition (β = -0.54, *SE* = 0.16, *Z* = -3.29, *p* = .001), and an effect of young-computer addressee group with more optimal descriptions provided to the computer compared to the young addressee (β = -0.52, *SE* = 0.24, *Z* = -2.19, *p* = .029). No other effects reached significance.

Recall that we also coded responses in terms of whether participants used basic-level or subordinate terms, as each target object was chosen so it could be differentiated at either level. The analysis examining whether participants used a basic term or not (subordinate) showed a significant effect of the contrast manipulation, with more basic terms being used in the no contrast than the contrast condition (β = -0.68, *SE* = 0.18, *Z* = -3.75, *p* < .001). In addition, participants were more likely to provide basic terms when providing instructions to a child addressee (β = -0.43, *SE* = 0.18, *Z* = -2.33, *p* = .02) than a same-age peer (see Figure 3).

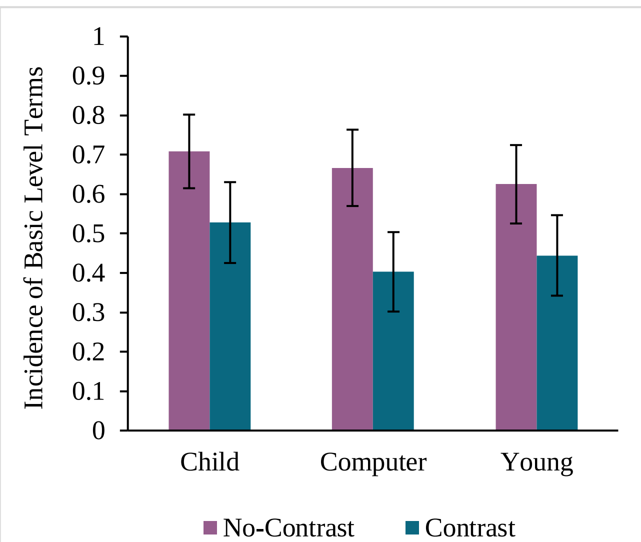

Figure 3. Use of basic-level terms by condition

We also investigated the incidence and the type of modifier used by participants. Specifically, we examined whether or not participants varied modifier use as a function of contrast and addressee type. Once again, the analysis revealed that participants were more likely to use modifiers in the contrast condition as a way to differentiate the target from the same-category object (β = 1.45, *SE* = 0.22, *Z* = 6.54, *p* < .001). In addition, participants were also more likely to provide modifiers when addressing a child than a younger adult (β = -0.48, *SE* = 0.22, *Z* = -2.21, *p* = .027, see Figure 4).

Finally, consistent with previous research, participants were overall more likely to use color modifiers compared to other types of modifiers. This was particularly more prevalent with the child addressee, but also quite common with the computer addressee (see Table 1).

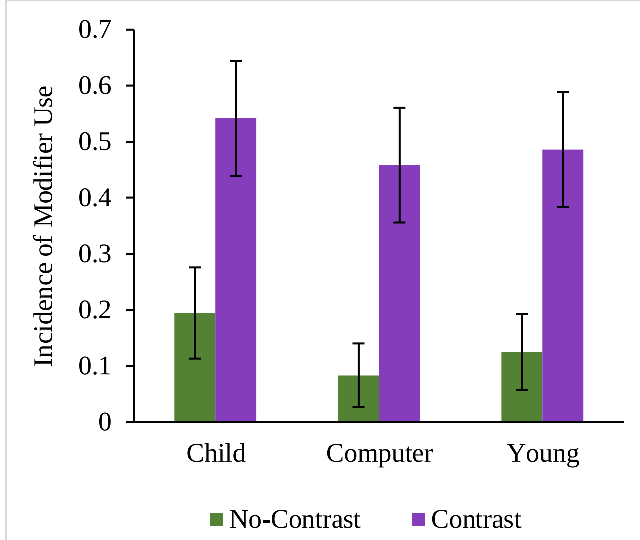

Figure 4. Modifier use by condition

Table 1. The percentage of different modifiers types

Audience	Color	Size	Location	Other
Child	67	7	2	24
Computer	60	5	2	33
Young	49	4	6	41

Discussion

This study investigated how speakers adjust the informational content of their referring expressions depending on both the referential context and the particular addressee (child, young adult, or computer). Participants provided instructions to their communicative partner to move objects on a visual display. The target objects were either unique in the display (no contrast) or accompanied by a same-category competitor object (contrast). First, we examined whether participants provided “just the right amount” of information to identify the target, avoiding the production of over- or under-informative descriptions. Results revealed that speakers were more likely to provide optimal descriptions in the no contrast condition. Interestingly, the most optimal descriptions were used with the computer addressee compared to the young adult. This could indicate that speakers perceived the computer to be less likely to rely on or be able to interpret modifying adjectives to locate a target. Because a computer would perhaps not tag an object’s properties in the same way as a human partner, speakers may have reasoned that modifying adjectives would be less useful to guide attention towards a target. It is also possible that speakers perceive the computer to be less proficient at the segmentation of continuous speech, and opt to use fewer words in their descriptions to reduce the processing load for the computer addressee.

In addition to the informational adequacy of the referential expressions, we analyzed the incidence of subordinate compared with basic-level terms. As expected, more basic-level terms were used in the no contrast condition, in which a basic term was sufficient to distinguish the target. Basic-level terms were also preferred when addressing the child group, indicating that speakers spontaneously shift to simple terminology when addressing children compared to younger adults or computers. These results suggest that participants were implicitly aware of the abilities of their addressee and adjusted their lexical choices to suit the needs of their current partner.

Finally, we examined the frequency of modifier use among participants across all expressions. As expected, participants used more modifiers when there was a same-category competitor present in the display. This resulted in more overinformative descriptions produced in the contrast condition. Of particular interest, however, was whether naming patterns varied as a function of audience type. Participants used more modifiers when addressing a school-aged child than another young adult. Further, fewer modifying adjectives were used when speaking to the computer. In terms of the type of information provided, the most distinctive addressee type was the child group. Specifically, although color modifiers were preferred overall, color was used at a slightly higher rate with children compared to the other two addressee types. This may reflect inferences made by speakers regarding vocabulary constraints in children, leading them to use more color descriptions to efficiently draw children's attention to distinguishing features of the target (this is also consistent with the finding that speakers use more basic-level terms with children). Given that color words are typically learned very early in life, speakers may be trying to simplify their terminology when communicating with a child, despite the fact the subordinate labels for each object would likely be familiar to a school-aged child. Interestingly, we did not find a difference in speech onset latencies across the three different addressees, suggesting the observed adjustments to lexical choices are made with relative ease, and as a result do not require additional planning time.

Overall, the current results suggest a pattern of communication in which speakers are sensitive to both the referential context and the processing abilities of the listener to guide their lexical choices and informativity in communication. The results support a cooperative account of overinformative descriptions (Rubio-Fernandez, 2019), in which speakers sometimes include redundant modifiers to facilitate comprehension for their addressee. Specifically, when addressing children, basic-level terms and perceptually simple forms of modification may be useful to facilitate efficient object identification. Further, the similar onset times between addressee conditions indicate that these adjustments are automatic, and do not place increased planning demands on the speaker.

Acknowledgements

This research was funded by a fellowship to Raheleh Saryazdi and a grant to Craig G. Chambers from the Natural Sciences and Engineering Research Council (Canada). We thank Chris Klammer for his assistance with data collection.

References

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, *43*(1), 361-374.

Bates, D., Maechler, B., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.

Burnham, D., Joeffry, S., & Rice, L. (2010). Computer-and human-directed speech before and after correction. *Proceedings of SST*, 13-17, Melbourne, Australia.

Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations?. *Lingua*, *120*(8), 1956-1972.

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., & Goodman, N. D. (2020). When redundancy is useful: A Bayesian approach to "overinformative" referring expressions. *Psychological Review*.

Engelhardt, P. E., Demiral, Ş. B., & Ferreira, F. (2011). Over-specified referring expressions impair comprehension: An ERP study. *Brain and Cognition*, *77*(2), 304-314.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, *20*(1), 104.

Frisson, S., & Murphy, G. L. (2019). Maxim of quantity and presupposition in understanding object labels. *Language, Cognition and Neuroscience*, 1-10.

Gatt, A., Krahmer, E., Van Deemter, K., & van Gompel, R. P. (2017). Reference production as search: The impact of domain size on the production of distinguishing descriptions. *Cognitive Science*, *41*, 1457-1492.

Grice, H. P. (1975). Logic and conversation. In *Speech Acts* (pp. 41-58). Brill.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, *4*(1), 85-110.

Koolen, R., Krahmer, E., & Swerts, M. (2016). Developmental Changes in Children's Processing of Redundant Modifiers in Definite Object Descriptions. *Frontiers in Psychology*, *7*, 1900.

Kriz, S., Anderson, G., & Trafton, J. G. (2010, March). Robot-directed speech: using language to assess first-time users' conceptualizations of a robot. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 267-274). IEEE.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., ... & Lacerda, F. (1997). Cross-language analysis of phonetic units in

language addressed to infants. *Science*, *277*(5326), 684-686.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13).
Maes, A., Marcelis, P., & Verheyen, F. (2007). Referential collaboration with computers. *Anaphors in Text: Cognitive, Formal and Applied Approaches to Anaphoric Reference*, *86*, 49.
Oviatt, S., Bernard, J., & Levow, G. A. (1998). Linguistic adaptations during spoken and multimodal error resolution. *Language and Speech*, *41*(3-4), 419-442.
R Core Team. (2019). R: A language and environment for statistical computing (Version 3.6.1) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.
Rubio-Fernández, P. (2016). How redundant are redundant color adjectives? An efficiency-based analysis of color overspecification. *Frontiers in Psychology*, *7*, 153.
Rubio-Fernandez, P. (2019). Overinformative speakers are cooperative: Revisiting the Gricean Maxim of Quantity. *Cognitive Science*, *43*(11), e12797.
Saryazdi, R., Bannon, J., & Chambers, C. G. (2019). Age-related differences in referential production: A multiple-measures study. *Psychology and Aging*, *34*(6), 791.
Stent, A. J., Huffman, M. K., & Brennan, S. E. (2008). Adapting speaking after evidence of misrecognition: Local and global hyperarticulation. *Speech Communication*, *50*(3), 163-178.
Tarenskeen, S., Broersma, M., & Geurts, B. (2015). Overspecification of color, pattern, and size: salience, absoluteness, and consistency. *Frontiers in Psychology*, *6*, 1703.
Tourtouri, E. N., Delogu, F., Sikos, L., & Crocker, M. W. (2019). Rational over-specification in visually-situated comprehension and production. *Journal of Cultural Cognitive Science*, *3*(2), 175-202.