Bilingualism: Language and Cognition

cambridge.org/bil

Bilingual referential choice in cognitively demanding situations

Carla Contemori[1] and Iva Ivanova[2]

[1]University of Texas at El Paso and [2]University of Texas at El Paso

Research Article

Cite this article: Contemori C, Ivanova I (2021). Bilingual referential choice in cognitively demanding situations. *Bilingualism: Language and Cognition* **24**, 83–95. https://doi.org/10.1017/S1366728920000176

Received: 5 April 2019
Revised: 29 January 2020
Accepted: 9 February 2020
First published online: 4 March 2020

Keywords:
L2 language production; non-native speakers; pronoun production; perspective-taking; privileged ground; common ground; dual task; working memory load

Address for correspondence:
Carla Contemori,
E-mail: ccontemori@utep.edu

Abstract

Under the Interface Hypothesis, bilinguals' non-nativelike referential choices may be influenced by the increased cognitive demands and less automatic processing of bilingual production. We test this hypothesis by comparing pronoun production in the L2 of nonbalanced Spanish–English bilinguals to that of English monolinguals in two cognitively challenging contexts. In Experiment 1, both monolinguals and bilinguals produced more explicit references when part of the information was unavailable to their addressee (privileged ground) than when all information was shared (common ground), evidencing audience design. In Experiment 2, verbal load led to more unspecified references than visual load and no load (an effect statistically indistinguishable between groups but numerically driven by the monolingual group). While bilinguals produced overall more pronouns than monolinguals in both experiments, there was no indication that bilinguals' referential choice was disproportionally affected by increased cognitive demand, contrary to the predictions of the Interface Hypothesis.

Introduction

Reference is a fundamental aspect of communication that allows speakers to refer to entities in the surrounding world. Monolingual speakers are flexible at making choices about how to refer to discourse entities (e.g., Arnold, 2010; 2016, for a review). For example, English speakers prefer more explicit forms (full noun phrases such as *the woman*) when introducing new entities in the discourse or referring to an entity that has not been mentioned recently, but tend to use attenuated forms (pronouns such as *she*) when the referent is the topic of conversation (e.g., Arnold & Griffin, 2007).

In contrast, non-balanced but highly-proficient bilinguals tend to overuse pronominal forms in a second language (L2) when it is a non-null subject language such as English, even when it leads to ambiguity (Contemori & Dussias, 2016). According to the Interface Hypothesis, bilinguals' non-nativelike referential choice, in a second (L2) and even in a first language (L1), is caused by a difficulty in computing interface structures, in this case between syntax and pragmatics, which is a cognitively more costly operation than computations within a single domain (Sorace, 2011). Further, bilinguals' interface difficulty is proposed to come from less detailed knowledge of or access to relevant computational constraints, less automatic processing (Clahsen & Felser, 2006; Contemori, Pozzan, Galinsky & Dussias, 2018; Pozzan & Trueswell, 2017; Sorace, 2011), and fewer available cognitive resources (Abutalebi, 2008; Sorace, 2011). The Interface Hypothesis then predicts that cognitively demanding situations would disproportionally affect bilinguals' referential choices relative to those of monolinguals. In this study, we test this prediction by examining the production of pronouns of nonbalanced but relatively proficient Spanish–English bilinguals in their L2, in privileged versus common ground (Experiment 1) and under verbal and visual memory load versus no load (Experiment 2).

Referential choice in monolinguals

The referential choices of monolingual native speakers seem to depend on referent accessibility in the speaker's own discourse model as well as on considerations about what the listener can or cannot interpret, also known as audience design. First, research across different languages has shown that the more salient an entity is in the preceding discourse, the more likely speakers are to choose a reduced form such as a pronoun to refer to it (e.g., Ariel, 1990; Arnold, 1998). The presence of a second referent in the preceding discourse also has an impact on referential choice, suggesting that referential choices are determined by saliency influencing referent accessibility for SPEAKERS (e.g., Arnold & Griffin, 2007). For example, when two referents are present in context (e.g., a girl and a boy in (1)), English speakers prefer to use a more explicit referring expression (i.e., a noun phrase (NP)) to refer to a salient antecedent (*the girl* in

© Cambridge University Press 2020

CAMBRIDGE UNIVERSITY PRESS

(1)), even if the use of the pronoun *she* would be felicitous and would not lead to ambiguity (Arnold & Griffin, 2007; Fukumura & Van Gompel, 2012; 2015).

(1) The girl was arguing with a boy. The girl left the room.

According to a discourse-oriented account of referential choice, this preference is related to the competition between two similar entities in the speaker's discourse model (Arnold & Griffin, 2007; Bard & Aylett, 2005). When two referents are mentioned, the speaker's attention is distributed between them, and this decreases the activation of the more salient antecedent in the speaker's discourse model. As a consequence, the speaker prefers to produce an explicit referring expression such as a noun phrase instead of a reduced form such as a pronoun.

Another prominent determinant of referential choice is audience design. On a listener-oriented account, speakers would choose pronominal forms only after evaluating whether such forms make it possible for the listener to identify the intended referent, which involves evaluating the referent's salience in the LISTENER'S discourse model (Ariel, 1990; Chafe, 1976; Clark & Marshall, 1981; Givón, 1983; Gundel, Hedberg & Zacharski, 1993; Prince, 1985). This view is grounded in the Gricean Maxim of Quantity (Grice, 1975), according to which speakers should be as informative as is required (but not more informative than is required). Supporting evidence comes from the so called "gender effect": Speakers are less likely to choose pronouns to refer to a salient entity when another salient entity in the previous discourse has the same gender as the pronoun (Arnold, Eisenband, Brown-Schmidt & Trueswell, 2000; Arnold, 1999; Francik & Clark, 1985; Karmiloff-Smith, 1985: see also Brown-Schmidt & Tanenhaus, 2006; Sedivy, 2003).

The way referential choice is influenced by discourse accessibility in the speaker's or presumed listener's discourse model are not mutually exclusive and may determine reference production in different contexts or at different processing stages (Ariel, 1990; Bard & Aylett, 2005; Gundel et al., 1993; Horton & Keysar, 1996). For example, Hendriks, Koster, and Hoeks (2014) propose that reference production involves two steps, first choosing a form preferred by the grammar and second, determining if this form would be interpretable by the listener as intended, or if another form must be chosen. Importantly, cognitive effort is associated with the second step. For example, it may be contingent on sufficient processing speed (Van Rij, van Rijn & Hendriks, 2010). In a referential communication task, Horton and Keysar (1996) showed that speakers did not take into account common ground in their referring expressions (e.g., *small circle* versus *circle*) when they were under time pressure to complete their utterances. This result is consistent with the two-step account in suggesting that audience design does play a role for reference production but only at a later stage of processing. Importantly, it also suggests that audience design is cognitively costly (and may be compromised in a cognitively demanding situation such as speaking under time pressure).

Further general evidence that reference production is cognitively costly is that speakers under increased cognitive load or with reduced cognitive abilities produce more unspecified references, which should be more economical for speakers (Hendriks et al., 2014; Vogels et al., 2015; but see Rosa & Arnold, 2011). For example, Vogels et al. conducted a picture-description task in which native speakers of Dutch produced story continuations to a confederate, referring to antecedents that were either discourse-salient or non-salient for the addressee. In one block, speakers had to keep a previously presented word in their working memory during picture description, and in another block they described pictures without a secondary task. Vogels et al. found that participants were more likely to use reduced forms (pronouns) under working memory load than under no load. The authors hypothesized that a working memory load impairs speakers' ability to calculate referent salience in the process of building discourse representations, and that results in the adoption of the easier for speakers pronominal forms (see also Hendriks et al., 2014, for a similar account).

Referential choice in bilinguals

Adult learners of a second language have a full set of referential forms in their L1 and may experience difficulties learning the use of referential expressions in the L2 (Montrul, 2004; Montrul & Rodríguez Louro, 2006; Rothman, 2007; 2008; 2009; Sorace & Filiaci, 2006; Tsimpli, Sorace, Heycock & Filiaci, 2004). A large body of research concerning the acquisition of null subject languages (e.g., Spanish, Italian, Greek, and Turkish) has demonstrated that acquiring the biases of use and interpretation of referring expressions in the L2 is a vulnerable domain. In null subject languages like Italian, Spanish and Greek, null pronouns are the most reduced referential form used to refer to an antecedent that is in the speaker's focus of attention, as shown in (2). However, in null subject languages there is also the possibility to use explicit pronouns or full NPs as shown in (3) and (4). Note that in null subject languages explicit pronouns tend to signal a topic shift (e.g., Carminati, 2002).

(2) Anthony$_i$ fue de vacaciones con Simón$_j$. *pro$_i$* Disfrutó mucho de la playa.
Anthony$_i$ went on vacation with Simon$_j$. *pro$_i$* enjoyed the beach very much.

(3) Anthony$_i$ fue de vacaciones con Simón$_j$. Él$_j$ disfrutó mucho de la playa.
Anthony$_i$ went on vacation with Simon$_j$. He$_j$ enjoyed the beach very much.

(4) Anthony fue de vacaciones con Simón. Anthony/Simon disfrutó mucho de la playa.
Anthony went on vacation with Simon. Anthony/Simon enjoyed the beach very much.

Learners of a null subject language tend to overproduce explicit pronouns in subject position, a pattern that does not impair communication but can be perceived as redundant by native speakers (Belletti, Bennati & Sorace, 2007; Montrul & Rodríguez Louro, 2006; Rothman, 2007; 2008, 2009; Sorace & Filiaci, 2006). Interestingly, such patterns are not limited to adult L2 learners but have been shown in a number of bilingual populations, including child bilinguals (Argyri & Sorace, 2007; Haznedar, 2010; Sorace, Serratrice, Filiaci & Baldo, 2009), L1 speakers in situation of attrition (Tsimpli et al., 2004), and heritage speakers (Montrul, 2004, 2006; Keating, VanPatten & Jegerski, 2011)[1].

Even in non-null subject languages such as English, bilinguals show non-nativelike patterns of pronoun production. For

[1]Note that the existing literature on referential choice in bilingual speakers focuses on late L2 learners with varying degrees of proficiency in the L2. In the present paper, the term bilingual is used to identify speakers of two languages, regardless of age of acquisition of the L2 and proficiency level. In our study, age of L2 acquisition did not modulate the results of either experiment, and an objective measure of grammar proficiency (MELICET) used as a continuous variable had only a marginally significant effect in Experiment 2.

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

example, highly proficient L2 speakers overproduced pronouns in their L2, in contexts in which native speakers would rather chose a NP (Contemori & Dussias, 2016). The study of Contemori and Dussias (2016) contained two picture-description experiments in English in which participants read a context sentence and gave a description of a picture depicting either one or two characters (of similar or different gender). When only one character was mentioned in the preceding discourse, English monolinguals chose a pronoun to refer to the salient antecedent more often than a NP. When two characters with either same or different gender were presented in the preceding discourse, English monolinguals preferred to use a NP to refer to the most salient antecedent, in line with previous studies (e.g., Arnold & Griffin, 2007). The same pattern was observed for highly proficient Spanish–English bilinguals, but bilinguals produced significantly more pronouns than monolinguals across all conditions. This was the case even when the use of a pronoun led to ambiguity because the two characters in the preceding discourse had the same gender. Contemori and Dussias (2016) observed that bilinguals' overproduction of pronouns was not related to an inability to consider the listener's perspective, as shown by a separate picture description task that required multiple topic shifts. Also, bilinguals' overproduction of pronouns did not seem related to their working memory and inhibitory control abilities (measured with an Ospan and a Flanker task, respectively). Still, Contemori and Dussias (2016) could not completely discard the role of cognitive capacity in bilinguals' referential choice.

Indeed, according to the Interface Hypothesis (Sorace, 2011), residual indeterminacy in bilingual referential choice across different languages and proficiency levels is due to an increased need for cognitive resources when computing interface structures between syntax and pragmatics. This is because computation of interface structures has an increased cognitive cost relative to the computation of structures in a single domain (e.g., Burkhardt, 2005; Piñango & Burkhardt, 2005). Further, especially in an L2, bilinguals have a less detailed knowledge of the relevant computational constraints or less automatic access to this knowledge (Abutalebi, 2008; Clahsen & Felser, 2006; Contemori et al., 2018; Pozzan & Trueswell, 2017). In addition, bilinguals have fewer available cognitive resources relative to monolinguals: Speaking in an L2 is cognitively effortful (e.g., Abutalebi, 2008), and even speaking one's L1 requires cognitive resources to prevent interference from the non-target language (Green, 1998; Kroll, Dussias, Bogulski & Valdés-Kroff, 2012). The need to manage conflict between the two languages may cause bilinguals to use their verbal working memory less efficiently (Luo, Craik, Moreno & Bialystok, 2013), which may result in a difficulty calculating discourse prominence (e.g., Contemori, Asiri, Perea-Irigoyen, in press), in turn leading to non-target choice of referring expressions.

Aims and predictions

The present study tests the ability of the Interface Hypothesis to explain bilingual referential choice. To this aim, we compare the production of pronouns by Spanish–English bilinguals in their L2 to that of English monolinguals in two English picture-description experiments that involve conditions with increased cognitive demand. The Interface Hypothesis proposes that bilinguals produce non-nativelike references because bilingual language production may require extra cognitive resources, in turn leaving fewer resources for the cognitively-demanding reference production. Based on these assumptions, we predict that bilingual referential choice would be even less nativelike in cognitively demanding situations because such situations will further deplete bilinguals' limited cognitive resources. Thus, the cognitively demanding conditions in our experiments should disproportionally affect the referential choices of bilinguals relative to those of monolinguals.

In both experiments, participants heard two context sentences (e.g., *A man sat at breakfast with a woman. The man was quite full.*) and saw a picture depicting the situation. They then read the word *Subsequently* from the computer screen and proceeded to describe to the experimenter a picture that continued the narrative of the context sentences (e.g., a picture of a woman getting up and holding a plate). All pictures depicted two people in everyday situations; the people could be referred to with a NP (e.g., *the woman*) or a pronoun (e.g., *she*). After hearing participants' descriptions, the experimenter selected the described pictures among a separate set of pictures.

Experiment 1 employed a common and privileged ground manipulation (Fukumura & van Gompel, 2012; Vogels et al., 2015). In the common ground condition, the two context sentences were heard by both speaker (participant) and listener (experimenter). In the privileged ground condition, the second context sentence was heard only by the speaker (participant). Sensitivity to this manipulation would be demonstrated by production of more explicit references such as noun phrases (and hence fewer pronouns) in privileged ground than in common ground. In common ground, no specific adaptation to the listener is needed because the listener has all the information and the discourse model is shared between speaker and listener. Conversely, in privileged ground more explicit references would help the listener bridge the first context sentence with the picture description, but determining the referring expressions optimally suited to the listener's discourse model would be cognitively costly (Hendricks et al., 2014; Vogels et al., 2015). The Interface Hypothesis then predicts that bilinguals tested in their L2 would be less sensitive, or not sensitive at all, to the ground manipulation and show less difference in pronoun production between privileged and common ground than monolinguals (or, a statistical interaction between the factors Language group and Ground condition). This is because the increased cognitive load of speaking an L2 would prevent bilinguals from the cognitively costly adoption of the listener's perspective in privileged ground.

Speakers' sensitivity to the privileged versus common ground manipulation was shown in the study of Vogels et al. (2015), whose participants produced overall more pronouns in common ground (Experiment 2) than in privileged ground (Experiment 1). We note that this manipulation had no statistical effect in the study of Fukumura and Van Gompel (2012), although there was a numerical trend for more pronouns in common than in privileged ground.

Experiment 2 directly manipulated cognitive load by including dual task conditions in which participants had to keep previously presented material in their working memory while completing the picture description task. Participants kept in memory five digits in a verbal load condition, and the positions of five squares presented on a 3 x 3 grid in a visual load condition. In a baseline no-load condition, participants kept nothing in memory during picture description and only afterwards saw five digits that they had to reproduce immediately.

The verbal load manipulation should affect referential choice such that participants produce more pronouns under verbal load

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

than under no load (Vogels et al., 2015). This is because pronouns place less processing burden on the production system (Almor, 1999; Hendriks et al., 2014; Kuijper, Hartman & Hendriks, 2015); resource depletion under load would increase the need for economy of computation and decrease speaker's abilities to compute the referent's salience in their own discourse model, possibly also a proxy for the listener's (Vogels et al., 2015). But crucially, under the assumptions of the Interface Hypothesis, we predict that referential specificity should be disproportionally affected in bilinguals under verbal load. Thus, bilinguals should show a larger effect of verbal load than monolinguals in their pronoun production (or, a statistical interaction between the factors Language group and Load condition but reflecting a different pattern than in Experiment 1). Conversely, the visual load condition should not affect the production of referring expressions because visual working memory should play no role in a verbal domain (e.g., Baddeley, 1986; Just & Carpenter, 1992) and specifically in production planning at the discourse level. Thus, both bilinguals and monolinguals should perform comparably on the visual load condition and no load conditions.

Experiment 1

Method

Participants

Twenty-one English monolingual speakers (mean age: 21 years; SD: 3.2) and forty-four Spanish–English bilinguals (mean age: 21 years; SD: 3.5) participated in the study. Monolingual and bilingual speakers were undergraduate students at a large American University and received course credits for their participation. Participants were recruited from a Spanish–English bilingual community at the U.S.–Mexico border. Therefore, monolingual speakers living in the area had some exposure to Spanish. We ensured that the monolingual participants in our experiments did not have any childhood exposure to Spanish and did not indicate any knowledge of a second language. Five additional monolingual participants were discarded from analyses because they reported childhood exposure to Spanish.

Bilingual participants acquired Spanish first in the family, and English as an L2 in early or late childhood. Bilingual participants' proficiency in English grammar (most relevant here in comparison to other standardized proficiency measures such as productive vocabulary) was tested with a subsection of the *Michigan English Language Institute College English Test* (MELICET). It includes 50 multiple-choice fill-in-the-blank questions divided into two sections, 30 individual sentences with one blank in each, and a scientific passage about the psychological influences of color, containing 20 blanks. The MELICET scores for the bilingual participants ranged between 20 and 48, indicating intermediate to high proficiency in English. Additionally, bilingual participants completed a language history questionnaire (Marian, Blumenfeld & Kaushanskaya, 2007). Bilinguals' language characteristics are summarized in Table 1.

Method

Materials and procedure

We used the picture description task from Vogels et al. (2015), translated into English. Participants described simple everyday scenes with two people to the experimenter (a trained research assistant who played the part of the addressee). On each trial, participants saw on the computer screen two pictures of a male and a female character engaging in a daily activity, and described the second (target) picture to the addressee (see Figure 1). At the beginning of the experiment, participants were instructed to give a description of each target picture to the addressee so that the addressee could select the corresponding picture in a picture book. During the experiment, the addressee could not communicate with participants except by giving a sign that she had selected the picture from the picture book and she was ready for the following trial. At that point, participants could continue with the following trial by pressing the spacebar on the computer keyboard.

Before describing the target picture (Figure 1, Panel B), participants were presented with a context picture (Figure 1, Panel A) that appeared on the screen with a sentence written underneath it that participants had to read aloud (e.g., *A man sat at breakfast with a woman*). This sentence was followed by a second prerecorded context sentence (e.g., *The man was quite full*). The two sentences, as shown in Figure 1, Panel A, provided a verbal context for the target picture. Participants would then see the target picture on the screen (Figure 1, Panel B), read the word *Subsequently* written underneath it, and continued the sentence by describing the target picture. Participants had seven seconds from the onset of the target picture to give their description. After this time had elapsed, the pictures disappeared from the screen and participants pressed the spacebar to continue with the following trial.

The crucial manipulation was the presentation of the second context sentence (e.g., *The man was quite full*). This second sentence was heard either by both the participant (speaker) and the experimenter (addressee) through the loudspeakers (common ground condition), or only by the participant through headphones (privileged ground condition).

There were a total of sixteen experimental items (defined as the two context sentences together with the context and target pictures). Each participant saw eight items in the common ground condition and eight items in the privileged ground condition. Each participant saw each item only once, but items occurred equally often in the two ground conditions across participants. Experimental items always contained one male and one female character and thus never gave rise to potential ambiguity. In all experimental items, the characters were referred to as *the woman* and *the man* or *the girl* and *the boy* in the context sentences (this was not the case for filler items, see below). Additionally, for each of the sixteen items there were two versions of each target picture: one in which the male character performed the action and was visually salient and one in which the female character performed the action and was visually salient. Therefore, the expected reference in participants' descriptions was to the first-mentioned character in the first context sentence on half of the trials, and to the second-mentioned character in the first context sentence on the other half. In addition, the gender of the salient character (i.e., the one that the speaker was expected to mention first in the description) and this character's position on the picture (left/right and foreground/background) were counterbalanced across items.

Four lists in a Latin-square design were created by crossing ground condition (common ground, privileged ground) and target-picture salient entity (first-mentioned, second-mentioned). Each list contained sixteen experimental items (eight in privileged ground and eight in common ground). Each list contained an additional 20 filler items, portraying similar everyday scenes as

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

Table 1. Language characteristics of bilingual participants in Experiments 1 and 2.

	Experiment 1		Experiment 2	
	Spanish (L1)	English (L2)	Spanish (L1)	English (L2)
Age of exposure	1.8 (1.8)	5.4 (3.3)	1.6 (1.6)	5.2 (4.9)
Length of residence in a country where the language is primarily spoken	15 (9.5)	17.9 (7.2)	16.5 (8.2)	14 (8.3)
Average daily exposure (%)	48.5 (17.4)	58.1 (17.2)	36 (27)	64 (24)
Speaking (% average daily)	42.3 (24.7)	58.1 (23.8)	49 (36)	51 (34)
Reading (% average daily)	27.2 (21.9)	74.7 (21.2)	34 (26)	62 (21)
Average proficiency				
Speaking Proficiency (1-10)	8 (1.5)	8.6 (1)		8.3 (5.2)
Comprehension Proficiency (1-10)	8.5 (1.1)	8.7 (1)	9.1 (1)	8 (2.8)
Reading Proficiency (1-10)	7.4 (1.2)	8.6 (1)	8.5 (1.6)	7.2 (2.5)
MELICET (50 total)	-	38 (6)	-	38 (5)

Note: Standard deviation is given in parentheses.

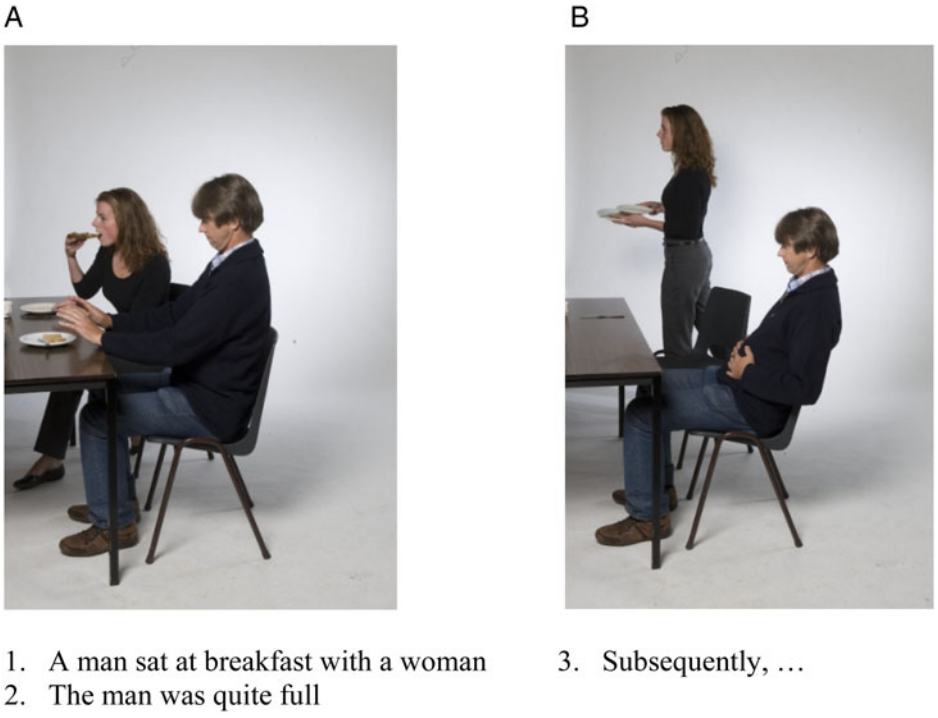


Fig. 1. Example of the picture-description task used in Experiments 1 and 2. Figure 1, Panel A: Context picture and sentences. Figure 1, Panel B: Target picture.

the experimental items. The filler pictures represented either only one character or two characters of the same gender; in the latter case, the characters were referred to in the context sentences with specific names such as *the actress* or *the writer*.

The ground manipulation was implemented within-subjects, and blocked. That is, half of the experimental and filler items were presented in a common ground block and the other half in a privileged ground block, with their order of presentation counterbalanced across participants. When half of the trials were completed, participants were given written instructions that the task would change. Participants completed two additional practice items at the beginning of each block, to ensure that they had understood the instructions.

Coding and data analysis

The speakers' descriptions were coded by a research assistant unaware of the purpose of the study. The coding was then checked by the first author, and any disagreement was resolved with discussion.

In total, there were 1056 descriptions, 704 for the bilingual group and 352 for the monolingual group. All the references to the non-salient character in the picture were discarded. We also discarded descriptions that did not contain either a NP or a third person pronoun (i.e., cases of ellipsis, third person plural pronoun "they", references to objects). A total of 76 (7.2%) descriptions were discarded (42 or 6.0% for the bilinguals and 34 or 9.7% for the monolinguals). We analyzed the remaining

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

descriptions which always contained either a pronoun or a NP referring to the salient character presented in the picture.

The factors manipulated in this task were ground condition (privileged ground versus common ground) and language group (monolinguals versus bilinguals). The dependent measure was the number of pronouns produced. We used Logistic Mixed-effects Regression Modeling (LMER; Jaeger, 2008) to analyze the number of pronouns in participants' descriptions (pronouns were coded as 1 and NPs were coded as 0). The fixed effects in the main model were Ground (privileged ground, common ground), Language group (monolinguals, bilinguals), and their interaction. A higher number of bilingual than monolingual participants was tested in Experiment 1 to allow greater power in investigating relationships between pronoun production and language history characteristics. Subsequent LMER models analyzed the number of pronouns produced by the bilingual group alone, as a function of their MELICET scores and age of first exposure to English. The fixed predictors in each of these models were Ground, one of the language history variables as a continuous predictor, and their interaction. Additional models explored how the pronouns produced in each ground condition by the two groups were affected by whether the salient (target) character on the target picture was the first- or second-mentioned entity in the first context sentence. These analyses are reported in the Appendix since they are outside of the main goals of this study.

All reported models used the *glmer* function in the lmerTest package (version 2.0-33, lme4 version 1.1-13) in R (version 3.4.1). The models had the maximal random-effects structure justified by the design. To aid convergence, the number of iterations the model computed was increased to 100,000. If a full random-effects model did not converge, the model was simplified by removal of random-effects correlations as a first step, and then step-wise removal of the random effect accounting for least variance, with the restriction that random slopes were removed before random intercepts.

Results

Figure 2 presents violin plots of the proportion of pronouns produced by monolinguals and bilinguals in common and privileged ground, out of the total number of NPs and pronouns produced.

The results of the LMER models are reported in Table 2. Participants produced overall more pronouns in privileged ground (36.7%) than in common ground (24.5%; a significant effect of Ground). Further, bilinguals produced overall more pronouns (37.2%) than monolinguals (17.4%; a significant effect of Language group). Importantly, the effect of ground was similar across the two groups (there was no interaction between Ground and Language group). Further, the violin plots in Figure 2 reveal that pronoun production follows a bimodal distribution, and effects seem driven by more versus fewer participants producing only pronouns (or, respectively, fewer versus more participants producing only noun phrases).

In the models testing the effects of language history variables on bilinguals' pronoun production in the two ground conditions, there were no effects or interactions involving MELICET scores or age of first exposure to English.

Discussion

Experiment 1 revealed that participants produced overall fewer pronouns in privileged ground than in common ground, but this effect did not differ between monolinguals and L2-speaking bilinguals. This finding is inconsistent with the claim of the Interface Hypothesis (Sorace, 2011) that cognitive resources and slower speed of processing are fundamental factors influencing referential choice in bilinguals. According to the predictions of the Interface Hypothesis, bilinguals should have shown less sensitivity than monolinguals to the perspective-taking needs of privileged ground because perspective taking is cognitively demanding.

However, while bilinguals and monolinguals used a similar perspective-taking strategy when selecting referring expressions, bilinguals used overall more pronouns than monolinguals (resulting in more individuals using pronouns exclusively), a non-nativelike aspect of reference production. This result is in line with that of Contemori and Dussias (2016) who found that Spanish–English bilinguals in their L2 produced more pronouns than monolinguals when reference maintenance was required in different discourse contexts. The findings of our Experiment 1 undermine the cognitive demands of perspective taking as a possible explanation for the non-nativelike pronoun production of the bilinguals in our study.

Our test of the Interface Hypothesis in Experiment 1 was built on the assumption that at least monolingual speakers would be sensitive to the common versus privileged ground manipulation, in line with Vogels et al. (2015). This assumption was warranted: Monolinguals *and* bilinguals showed sensitivity to the needs of their listeners by using more explicit references when part of the context was not available to the listeners. Thus, our results demonstrate that both monolingual and bilingual speakers consider the addressee's discourse model, using audience design. Our results diverge from those of Fukumura and van Gompel (2012), who found that monolingual English speakers used a similar number of pronouns in common and privileged ground, indicating that speakers used their own discourse model when choosing a referring expression. We are unsure of the reasons of this divergence, beyond differences in materials and analyses (LMER in our study and that of Vogels et al., and ANOVA in the study of Fukumura & van Gompel, 2012). Most important for our purposes here is that our results are consistent with the study of Vogels et al. (2015), whose materials we used directly. We also note that Fukumura and van Gompel did observe a numerical trend towards fewer pronouns in the privileged (33%) than in the common ground condition (37%) (an effect approaching significance in their by-item analyses).

In a second experiment, we put the Interface Hypothesis to another test. In Experiment 2, bilinguals and monolinguals performed the picture description task under added cognitive load.

Experiment 2

In Experiment 2, we implemented a second manipulation in which referential choice is tested in a cognitively effortful context. We used the same picture description task as in Experiment 1, but in which all the information was in common ground. We additionally created three load conditions: (i) a verbal load condition, in which participants had to remember five digits while performing the picture description; (ii) a visual load condition, in which participants had to remember the position of five squares on a grid while performing the picture description; and (iii) a no load control condition, in which participants only performed the picture description (followed by immediate recall of five digits).

As demonstrated in previous studies, verbal memory load can affect how speakers choose between attenuated and explicit

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

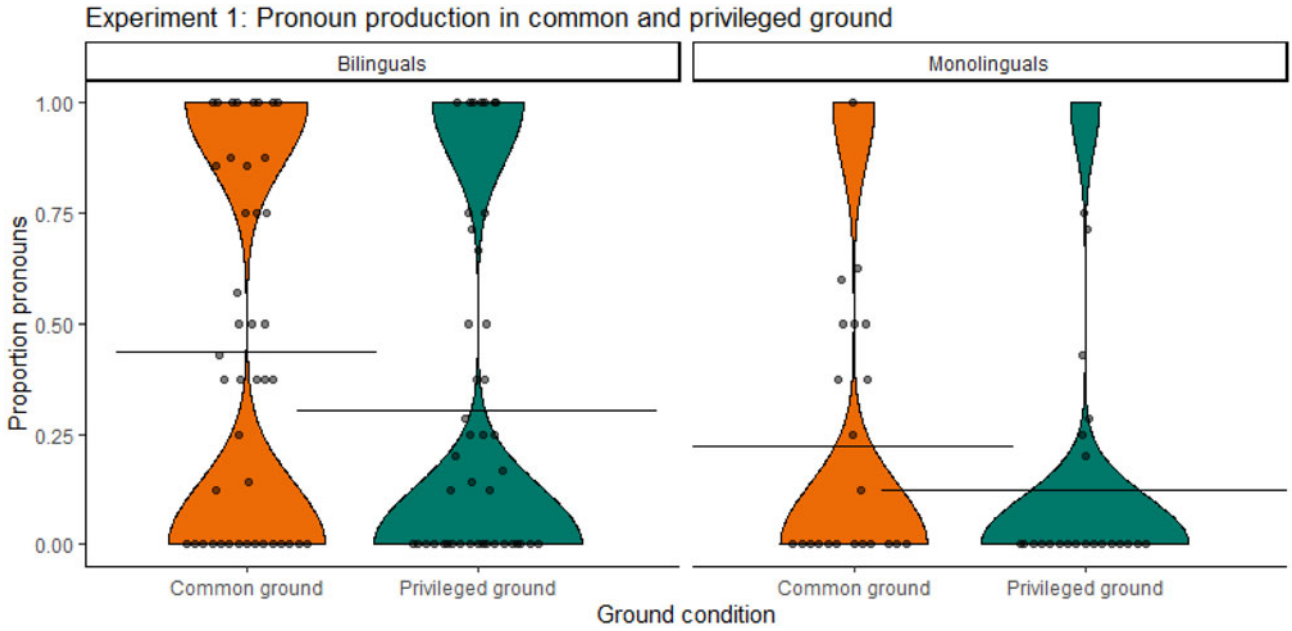


Fig. 2. Violin plot of the proportions of pronouns produced by bilinguals and monolinguals in common and privileged ground in Experiment 1. Dots reflect individual participants' condition means, and horizontal lines reflect group condition means.

Table 2. LMER results in Experiment 1.

Models	Predictors	*Estimate*	*SE*	*z*	*p*
Main model[a]					
	Ground	1.10	.21	5.27	< .001
	Language group	-2.34	.94	-2.48	.01
	Ground x Language group	.07	.46	.14	.89
Models with language history predictors[b]					
MELICET scores[c]	MELICET	.004	.09	.04	.97
	MELICET x Ground	.02	.07	.25	.80
Age of first exposure to English[c]	Age Eng. exposure	-.05	.17	-.28	.78
	Age Eng. exposure x Ground	.20	.14	1.40	.16

Note: In this and subsequent tables, dark grey rows indicate significant effects, and light grey rows indicate marginal effects.
[a]This model (the first that converged) had only by-subject and by-items random intercepts.
[b]The number of participants included in these models differed because of missing data.
[c]N = 43

expressions (e.g., Arnold, Bennetto & Diehl, 2009; Vogels et al., 2015). In the study of Vogels et al. (2015), verbal load affected speakers' production of referring expressions, such that participants used pronouns more often under verbal load than under no load. Vogels et al. suggested that the verbal memory load affected speakers' assumptions about the referent's accessibility according to the listener's perspective (implying that at least some of the speakers' referential choices are made based on the listener's model of the discourse).

In Experiment 2, we expected that monolingual participants would increase their use of pronouns in the verbal load condition compared to the no load condition, in line with Vogels et al. (2015). Regarding visual load, we did not expect that it would influence the choice of referring expressions, to the extent that verbal and visual working memory resources are assumed to be separate (e.g., Baddeley, 1986). Hence, we did not predict differences between the visual load and no load condition in monolinguals. Note, however, that this is the first study (to our knowledge) investigating the contribution of visual memory to referential choice.

Most importantly, we predict based on the Interface Hypothesis that the verbal memory condition would differentially impact the referential choice of bilinguals in comparison to monolinguals. That is, bilinguals should show a larger effect of verbal load (i.e., produce disproportionally more pronouns under verbal load than

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

under no load), relative to monolinguals (while pronoun production under visual load should not differ between the two groups). The verbal load effect should manifest as an interaction between Load condition and Group (but reflecting a different pattern than predicted for Experiment 1). On the other hand, if bilinguals' nonnativelike choice of referring expressions is not related to limited cognitive capacity (in accordance with the findings of Experiment 1 and *contra* the Interface Hypothesis), verbal load should impact bilinguals' pronoun production similarly to monolinguals. If so, we should observe a main effect of Load (with more pronouns produced overall under verbal load than under no load and visual load). Lastly, following the findings of Experiment 1, we may predict a main effect of group such that bilinguals produce overall more pronouns than monolinguals. However, such globally higher pronoun production by bilinguals (i.e., a non-nativelike pattern) would be inconsistent with the predictions of the Interface Hypothesis if it is unaffected by verbal load.

Method

Participants

Seventy-eight Spanish–English bilinguals (22 males; mean age: 21.4; SD; 4.3) and fifty-four monolingual speakers (19 males; mean age: 20.9; SD; 3.1) participated in the experiment. One additional bilingual participant was discarded due to a coding error. Participants were recruited at the same American University as in Experiment 1 and were living in a Spanish–English community at the border between the US and Mexico at the time of testing. Monolingual participants indicated that English was their only language and did not report childhood exposure to a second language.

Bilinguals had acquired Spanish (L1) in the family and English in early or late childhood, as in Experiment 1. Bilingual participants completed the same proficiency test as in Experiment 1 (MELICET mean score: 38.7; SD: 5.4; range: 25/50-47/50). T-tests comparing the MELICET scores of bilinguals across Experiment 1 and 2 showed no significant differences [$t(117) = 0.12$; $p = 0.9$), indicating that the two experiments tested bilinguals with comparable English proficiency scores. The characteristics of the bilingual participants who took part in Experiment 2 are summarized in Table 1.

Materials and procedure

Experiment 2 employed the same picture description task as Experiment 1, with the following exceptions. First, all the items were presented in common ground. Second, each participant performed the picture description task in three load conditions: verbal load, visual load and no load.

In the verbal load condition, participants saw five digits on the screen for 1000 ms before the picture was presented, as shown in Figure 3. Participants were instructed to keep the five digits in memory while they were describing the picture. After completing the picture description, participants had to type in the five digits on the numeric keypad on the keyboard. When they were ready to continue, they pressed the space bar to move on to the following trial.

The visual load condition used a similar presentation, except that participants were presented with 3 x 3 grids in which five squares were highlighted in red. After seeing a grid for 1000 ms, participants gave a picture description while keeping in memory the position of the highlighted squares on the grid. After completing the description, participants highlighted five squares in an empty 3 x 3 grid that was shown on the screen by using the numeric keypad (the 1 to 9 digits on the keypad mapped onto the grid positions by moving first to the right and then down). Participants then pressed the space bar to continue to the next trial.

In the No Load condition, participants were presented with five digits at the end of each trial (after they had already described the picture) for 1000 ms, and had to enter the digits on the screen immediately after seeing them, as shown in Figure 3. In this condition, no information was kept active in working memory during the description of the picture.

The cognitive load manipulation was implemented within participants, and blocked. That is, the task was divided in three blocks, one for each load condition. Each block contained six experimental items, and six to seven fillers. Before starting each block, participants were given instructions and completed two practice items. The order of the blocks was counterbalanced across participants.

Coding and data analysis

We adopted the same coding criteria as in Experiment 1. A total of 101 (4.7%) descriptions were discarded from analyses (70 or 5.6% for the bilinguals and 31 or 3.5% for the monolinguals). We analyzed the remaining descriptions which always contained either a pronoun or a NP referring to the salient character presented in the picture.

We coded accuracy in the memory task by scoring the number of digits and grid positions that participants entered correctly on each trial. Accuracy was scored regardless of the order of presentation of the digits or grid positions, with 5 being the most accurate score (i.e., all numbers or grid positions were correct) and 0 being the least accurate score (i.e., none of the digits or grid positions entered were correct).

Two LMER models compared cognitive load effects on pronoun production (a pronoun was coded as 1 and a noun phrase as 0). The first model compared the effects of verbal load to the no load baseline. The predictors in this model were Cognitive load (verbal load, coded as 0.5, and no load, coded as -0.5; visual load was coded as 0), Language group (centered around the mean: bilingual, coded as -0.33, and monolingual, coded as 0.67) and Memory accuracy as a continuous predictor (centered around the mean), as well as their interactions. The second model was identical except the Cognitive load predictor compared visual load against no load (visual load was coded as 0.5 and no load was coded as -0.5; verbal load was coded as 0). As in Experiment 1, additional models analyzed bilinguals' pronoun production as a function of their MELICET scores and age of first exposure to English. These models had Load, one language history variable, and their interaction as fixed predictors. The Load predictor compared verbal load to no load. (Since these models showed no interactions with Load, we also ran models with each language history variable as a single fixed predictor; these models showed identical effects of the language history variables to the ones reported here.) As in Experiment 1, we ran additional models using Referent as a factor, to examine if participants' referential choice was influenced by which referent was cued by the target picture. The model results are reported in the Appendix.

Results

Figure 4 presents violin plots of the percentage of pronouns produced by the two groups in the three Cognitive load conditions

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

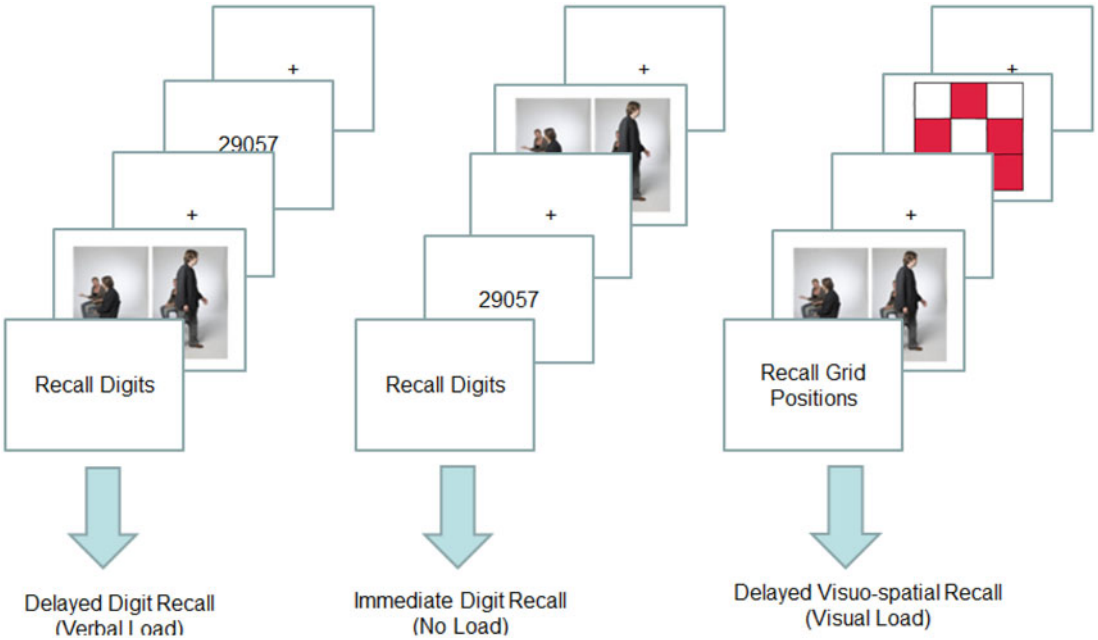


Fig. 3. Example of the three memory load conditions in Experiment 2.

(verbal load, visual load and no load) out of the total number of NPs and pronouns produced.

The results of the LMER models are reported in Table 3. The model targeting effects of verbal load indicated that, overall, participants produced slightly more pronouns under verbal load (43.8%) than under no load (39.6%; a marginal effect of Cognitive load). In addition, although there was no statistical interaction between Language group and Cognitive load, the numerical trend on Figure 4 suggests that it was monolinguals who produced on average more pronouns under load than under no load, while for bilinguals there was no difference. As in Experiment 1, bilinguals produced overall more pronouns (48.0%) than monolinguals (32.2%; a significant effect of Language group). No other effects or interactions were significant.

The models including language history variables indicated that participants with higher objective proficiency (i.e., higher MELICET scores) tended to produce slightly fewer pronouns than bilinguals with lower objective proficiency (a marginal effect of MELICET scores, $p = .09$). No other effects in these models were significant.

Discussion

In Experiment 2, participants produced on average slightly more pronouns under verbal load than without (while visual load had no effect), but the increase in pronoun production under verbal load was statistically comparable between bilinguals and monolinguals (and numerically seemed driven by monolinguals and even absent for bilinguals). The fact that bilinguals did not produce disproportionally more pronouns under verbal load relative to monolinguals further undermines the assumptions based on the Interface Hypothesis (Sorace, 2011) that the referential choice of bilinguals should be disproportionally affected by verbal load relative to monolinguals. Still, as in Experiment 1, bilinguals produced more pronouns overall than monolinguals. These results do not support the hypothesis that limited cognitive resources can be a fundamental factor explaining our bilinguals' nonnativelike pattern of referential choice. (We propose explanations of this pattern in the General Discussion.)

We found a marginal effect of verbal memory load compared to a no load condition, with (monolingual) participants producing slightly more pronouns when verbal memory was loaded. These results are comparable to those of Vogels et al. (2015), who used a verbal memory load manipulation with Dutch speakers during the same picture description task. Despite differences with the present study (for example, their verbal load consisted in tracking whether one of two monosyllabic content words was the same as the one presented on the previous trial), the effect of load in Vogels et al.'s data was also relatively small.

Concerning the visual load manipulation, we did not find a difference between the number of pronouns produced under visual load and no load. In our experimental design, the visual information presented to participants did not seem to interfere with the language production system and referential choice in particular[2]. This result supports our prediction that visual working memory would not be recruited – or would be recruited to a lesser degree

[2]As pointed out by an anonymous reviewer, the visual condition could be considered verbal in that participants could remember the display as numbers, rather than as a visual pattern.

To check for this possibility, we debriefed participants at the end of the session, asking about what strategy they were using to remember the grid positions and giving them a four-option answer. However, as indicated by the percentages of responses below, the majority of our participants reported that they used a "visual" strategy:

1) "I remembered how the grid positions looked on the screen": monolinguals=0.49; bilinguals=0.55;
2) "I tried to see if the grid positions looked like a shape or a thing": monolinguals=0.22; bilinguals=0.20;
3) "I converted the grid positions into numbered positions": monolinguals= 0.20; bilinguals=0.15;
4) "Other": monolinguals=0.09; bilinguals=0.10.

An independent-sample t-test demonstrated no difference between the number of monolingual and bilingual participants who remembered the display as numbers ($t(130)=0.497$, $p=.6$).

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

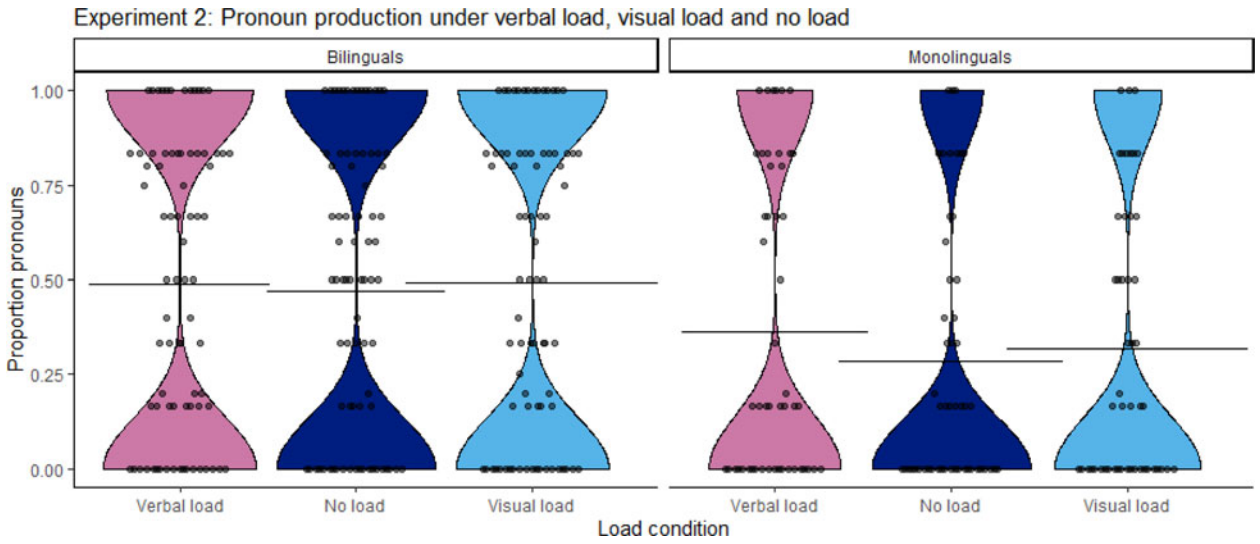


Fig. 4. Violin plot of the proportions of pronouns produced by bilinguals and monolinguals under verbal load, visual load and no load in Experiment 2. Dots reflect individual participants' condition means, and horizontal lines reflect group condition means.

than verbal working memory – during referential choice (see Baddeley, 1986).

Note that in Experiment 2 recall accuracy was comparable across bilinguals and monolinguals (there was no interaction between Memory accuracy and Language group). Additionally, recall accuracy did not seem to influence the production of referring expressions overall. This result suggests that differences in pronoun production between bilinguals and monolinguals in our study were not attributable to differential recall success between the two groups.

General Discussion

This study tests the claim of the Interface Hypothesis that bilinguals exhibit non-nativelike referential choices because reference production is cognitively effortful, and bilingual language processing may be less automatic or efficient and demand extra cognitive capacity (Sorace, 2011). We thus aimed to evaluate the relative importance of cognitive resources during the production of referring expressions in bilingual and monolingual speakers. In two experiments, we compared the production of referring expressions in cognitively effortful circumstances of Spanish–English bilinguals speaking in their intermediate-to-highly proficient L2 (English) to that of English monolinguals.

Experiment 1 used a common versus privileged ground manipulation under the assumption that referential choice in privileged ground is more cognitively effortful because it makes it necessary to consider the listener's perspective. Experiment 1 thus tested the prediction that bilinguals' referential choice would show less or no audience design by producing disproportionally more reduced referential forms (pronouns) when the preceding discourse information was only partially shared with the listener (privileged ground). Even though bilinguals produced significantly more pronouns than monolinguals overall, thus showing a non-nativelike pattern, both groups showed a similar-sized sensitivity to privileged ground. This result demonstrates that speakers are sensitive to their listeners' needs and engage in audience design, but it is inconsistent with the prediction of the Interface Hypothesis.

In Experiment 2, we used a cognitive load manipulation in a dual task context to further test the prediction that, in a cognitively challenging situation, bilinguals may produce disproportionally more pronouns than monolinguals. Consistent with Vogels et al. (2015), our analyses revealed an increase in pronoun use when participants had to keep five digits in working memory (verbal load condition) relative to when they did not (no load condition), which was not significantly modulated by group (and, if anything, the numerical trend suggests that it was driven entirely by the MONOLINGUAL group). Keeping in memory the position of five squares on a 3 x 3 grid (visual load condition) had no effect. Further, bilinguals produced more pronouns than monolinguals across all conditions, showing the same non-nativelike pattern as in Experiment 1.

In sum, we assumed that the privileged ground condition (Experiment 1) and the verbal load condition (Experiment 2) are more effortful in terms of cognitive resources, and we reasoned that a disproportionate tendency of bilinguals to produce pronouns in these conditions relative to monolinguals would support the Interface Hypothesis. Both experiments showed sensitivity to the ground and load manipulations, but the effects of these manipulations were comparable across monolinguals and intermediate-to-high proficiency bilinguals. Thus, the results of this study do not support the initial predictions based on the Interface Hypothesis.

Interestingly, an aspect of our results is in line with the findings of Contemori and Dussias (2016). These authors found that Spanish–English bilinguals tested in a picture description task without cognitively effortful conditions produced more pronouns overall than English monolinguals. We note that, in Contemori and Dussias's (2016) study, the bilingual participants were all highly proficient in their L2 English, while in the present study we had a wider range of proficiency levels as measured by a standardized assessment of grammar knowledge (MELICET). Additionally, Contemori and Dussias found no relationship between two cognitive control measures (Flanker task and the Ospan task) and bilinguals' referring expressions. The results of the present study thus provide further evidence that Spanish–English bilinguals tested in English tend to be less explicit than

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

Table 3. LMER results in Experiment 2.

Model	Predictors	*Estimate*	*SE*	*z*	*p*
Effects of verbal load[a]					
	Load	.31	.16	1.95	.052
	Language group	1.29	.43	3.05	.002
	Memory accuracy	-.04	.09	-.40	.69
	Load x Language group	-.37	.34	-1.10	.27
	Load x Memory accuracy	.21	.24	.89	.37
	Language group x Memory accuracy	.17	.20	.85	.40
	Load x Language group x Memory accuracy	-.55	.51	-1.08	.28
Effects of visual load					
	Load	.10	.21	.48	.63
	Language group	1.38	.43	3.18	.002
	Memory accuracy	-.10	.08	-1.19	.23
	Load x Language group	-.08	.43	.20	.85
	Load x Memory accuracy	.31	.25	1.25	.21
	Language group x Memory accuracy	.22	.18	1.20	.23
	Load x Language group x Memory accuracy	-.35	.53	-.66	.51
Models of verbal load vs. no load with language history predictors[b]					
MELICET scores[c]	MELICET	-.08	.05	-1.72	.09
	MELICET x Load	.03	.05	.60	.55
Age of first exposure[d] to English	Age Eng. exposure	-.08	.06	-1.38	.17
	Age Eng. exposure x Load	-.08	.06	-1.33	.19

[a]This model had only random intercepts for subjects and items.
[b]The number of participants included in these models differed because of missing data.
[c]N = 75
[d]N = 74

English monolinguals when producing referring expressions (regardless of their exact level of proficiency given it is at least intermediate). They also suggest that factors other than increased cognitive demands may be at the heart of this behavior, although more work is needed to test other types of cognitive demand.

What could explain, then, the overall higher use of pronouns in Spanish–English bilinguals during picture description compared to English monolinguals? We consider here two possible explanations: (1) a possible lack of bilinguals' exposure to referential use in English; and (2) the possibility that bilinguals may have experienced cross-linguistic interference (see also Sorace, 2011).

The hypothesis that bilingual participants lack sufficient exposure to referential use in English (Hypothesis 1) is consistent with evidence that experience, including the quality and the quantity of the input received by bilingual speakers, may have an impact on referential choice (Sorace & Serratrice, 2009). Bilinguals typically have less exposure to their second language than monolingual speakers. The importance of exposure may be crucial in the processing of discourse information, as in the case of referential choice investigated here. For example, as shown by a large body of studies looking at priming effects in production, the occurrence of a word or a particular syntactic construction is a significant predictor for its subsequent use. Although priming studies on monolinguals and bilinguals have focused mainly on lexical and syntactic production, recent research has looked at the effects of priming on discourse production in bilingual populations as a precursor of language change (Travis, Torres-Cacoullos & Kidd, 2017). It is thus possible that bilinguals may have less experience with how referential forms are used in English as a result of less exposure to the L2. However, we note that, if exposure to the L2 alone is the cause of the production pattern observed in bilinguals, we should have found a more robust effect of proficiency and age of exposure in both of our experiments. The bilinguals we tested ranged from intermediate to highly proficient in their L2 English (as measured with a grammar test). However, when an effect of proficiency emerged in our analyses (Experiment 2), it only approached significance (and age of first exposure to English had no effect in either experiment). Additionally, our bilinguals ranged on proficiency more than the bilinguals tested in Contemori and Dussias (2016), but in both studies bilinguals used overall more pronouns than English monolinguals. In sum, it is possible that there is an effect of exposure on discourse production. This factor should be systematically investigated in future research on bilingual pronoun production, and the priming methodology may be a promising technique to explore this issue.

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

Concerning the cross-linguistic interference hypothesis (Hypothesis (2 above)), Spanish and English (the two languages of the bilinguals tested here) present some important differences based on the sets of pronominal forms available and their use. Specifically, Spanish–English bilinguals may produce more pronouns when two referents are presented in the previous context because they may fail to suppress the assumption from their L1 Spanish that overt pronouns signal a topic-shift (e.g., Contemori & Dussias, 2016). According to this hypothesis, bilinguals use forms which are explicit in their L1 but not explicit enough in their L2 (because in English the more explicit form is the NP). Although some studies on referential choice across-languages have excluded the possibility of cross-linguistic interference (e.g., Chamorro, Sorace & Sturt, 2016; Contemori & Dussias, 2016; Sorace, 2011), other studies have interpreted the non-nativelike production of referring expressions in the L2 as the result of cross-linguistic differences (Montrul, 2004; Montrul & Rodríguez Louro, 2006; Rothman, 2007; 2008; 2009; Sorace & Filiaci, 2006; Tsimpli et al., 2004). In the present study, we did not test a group of bilinguals in which the L1 and L2 have the same set of referring expressions and similar production biases. Hence, we cannot exclude that the production of referring expressions may be susceptible to L1 interference even with bilinguals with early exposure and relatively high proficiency in the L2. Future research should address this question, testing referential choice in bilingual groups that have different language pairs.

We further propose that bilinguals' difficulty with referring expressions may be due in part to a difficulty evaluating the saliency of the referents in discourse (Contemori et al., in press). As suggested by Cunnings (2017) for comprehension, L2 speakers may be more susceptible to retrieval interference during sentence processing, which could result in non-target interpretation of pronouns. During comprehension of referring expressions, a memory retrieval operation triggers the retrieval of the potential antecedents presented in the discourse. According to Cunnings (2017), L2 learners may experience higher competition between two salient antecedents in comparison to monolingual speakers, resulting in non-target interpretation (see also Contemori et al., in press). This account could be adopted to interpret the production results presented in the current Experiments 1 and 2. If bilinguals experience a difficulty evaluating the saliency of the referents presented in the preceding discourse, they may adopt an underspecified reference more often than monolingual speakers. Nevertheless, more research is needed to pin down the underlying causes of the observed non-nativelike reference production in bilinguals, and cross-linguistic interference and input should be investigated further.

Conclusion

In the present study, we tested the impact of increased cognitive demand on bilingual referential choice. The results of two experiments indicated that bilinguals' choice of referring expressions remained non-nativelike (bilinguals produced more pronouns overall than monolinguals) but was not differentially affected in cognitively effortful contexts relative to monolinguals, contrary to some of the predictions derived from the Interface Hypothesis (Sorace, 2011).

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728920000176

Acknowledgements. We thank Catherine Householder, Jennifer Young, Che Juárez and Zacnite Garcia for data collection and coding, Laurel E. Brehm for teaching us to make violin plots, and the audiences of the Architectures and Mechanisms for Language Processing conference and the Annual CUNY Conference on Human Sentence Processing for helpful comments and suggestions.

References

Abutalebi J (2008) Neural aspects of second language representation and language control. *Acta Psychologica* **128**, 466–78.

Almor A (1999) Noun-phrase anaphora and focus: the informational load hypothesis. *Psychological Review* **106**, 748–765.

Argyri E and Sorace A (2007) Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition* **10**, 79–99.

Ariel M (1990) *Accessing noun phrase antecedents*. New York, NY: Routledge.

Arnold JE (2016) Explicit and emergent mechanisms of information status. *TopICS* **8**, 722–736.

Arnold J (2010) How Speakers Refer: The Role of Accessibility. *Language and Linguistics Compass* **4**, 187–203.

Arnold JE (1999) *Marking salience: The similarity of topic and focus*. Unpublished manuscript, University of Pennsylvania.

Arnold JE (1998) *Reference Form and Discourse Patterns*. Dissertation, Stanford University.

Arnold JE, Eisenband JG, Brown-Schmidt S and Trueswell JC (2000) The immediate use of gender information: eyetracking evidence of the time-course of pronoun resolution. *Cognition* **76**, 13–26.

Arnold JE, Bennetto L and Diehl JJ (2009) Reference Production in Young Speakers with and without Autism: Effects of Discourse Status and Processing Constraints. *Cognition* **110**, 131–146.

Arnold J and Griffin ZM (2007) The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language* **56**(4), 521–536.

Baddeley AD (1986) *Working Memory*. Oxford, UK: Oxford Univ. Press.

Bard, EG and Aylett MP (2005) Referential form, word duration, and modeling the listener in spoken dialogue. In Trueswell JC and MK Tanenhaus (eds), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press, pp. 173–191.

Belletti A, Bennati E and Sorace A (2007) Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language and Linguistic Theory* **25**, 657–689.

Brown-Schmidt S and Tanenhaus MK (2006) Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language* **54**, 592–609.

Burkhardt P (2005) *The syntax-discourse interface: Representing and interpreting dependency*. Amsterdam/Philadelphia: John Benjamins.

Carminati MN (2002) *The processing of Italian subject pronouns* (Unpublished doctoral dissertation). University of Massachusetts, Amherst.

Chafe WL (1976) Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Charles NL (ed), *Subject and Topic*. New York: Academic Press, pp. 27–55.

Chamorro G, Sorace A and Sturt P (2016) What is the source of L1 attrition? The effects of recent re-exposure on Spanish speakers under L1 attrition. *Bilingualism: Language and Cognition* **9**(3), 520–532.

Clahsen H and Felser C (2006) How native-like is non-native language processing? *Trends in Cognitive Sciences* **10**(12), 564–570.

Clark HH, and Marshall CR (1981) Definite reference and mutual knowledge. In Joshe AK, B Webber and IA Sag (eds), *Elements of discourse understanding*. Cambridge: Cambridge University Press.

Contemori C, Asiri O and Perea Irigoyen ED (in press) Anaphora resolution in L2 English: an analysis of discourse complexity and cross-linguistic interference. *Studies in Second Language Acquisition*.

Contemori C, Pozzan L, Galinsky P and Dussias G (2018) When actions and looks don't line up: The contribution of referential and prosodic information in the processing of PP ambiguities in bilinguals. *Linguistic Approaches to Bilingualism*. DOI: https://doi.org/10.1075/lab.18001.con

Contemori C and Dussias G (2016) Referential choice in a second language: evidence for a listener-oriented approach. *Language, Cognition and Neuroscience* **31**(10), 1257–1272.

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press

Cunnings I (2017) Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition* **20**(4), 659–678.

Francik EP and Clark HH (1985) How to make requests that overcome obstacles to compliance. *Journal of Memory and Language* **24**, 560–568.

Fukumura K and van Gompel RPG (2012) Producing Pronouns and Definite Noun Phrases: Do Speakers Use the Addressee's Discourse Model? *Cognitive Science* **36**(7), 1289–1311.

Fukumura K and van Gompel RPG (2015) Effects of order of mention and grammatical role on anaphor resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **41**(2), 501–525.

Givón T (1983) Topic continuity in discourse: An introduction. In Givón T (ed), *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins Publishing, pp. 1–42.

Green DW (1998) Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* **1**, 67–81.

Gundel JK, Hedberg N and Zacharski R (1993) Cognitive status and the form of referring expressions in discourse. *Language* **69**, 274–307.

Haznedar B (2010) Transfer at the syntax-pragmatics interface: Pronominal subjects in bilingual Turkish. *Second Language Research* **26**, 355–378.

Hendriks P, Koster C and Hoeks JCJ (2014) Referential choice across the lifespan: why children and elderly adults produce ambiguous pronouns. *Language, Cognition and Neuroscience* **29**(4), 391–407.

Horton WS and Keysar B (1996) When do speakers take into account common ground? *Cognition* **59**, 91–117.

Jaeger TF (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* **59**, 434–446.

Just MA and Carpenter PA (1992) A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* **98**, 122–149.

Karmiloff-Smith A (1985) Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes* **1**, 61–85.

Keating G, VanPatten B and Jegerski J (2011) Who was walking on the beach: Anaphora resolution in Spanish heritage speakers and adult second language learners. *Studies in Second Language Acquisition* **33**, 193–221.

Kroll JF, Dussias PE, Bogulski CA, and Valdés Kroff JR (2012) Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In Ross B (ed), *The Psychology of Learning and Motivation*, Volume **56**. Academic Press, pp. 229–273.

Kuijper S, Hartman CA and Hendriks P (2015) Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns. *PLoS ONE* **10**(7).

Luo L, Craik FIM, Moreno S and Bialystok E (2013) Bilingualism interacts with domain in a working memory task: Evidence from aging. *Psychology and Aging* **28**(1), 28–34.

Marian V, Blumenfeld HK and Kaushanskaya M (2007) The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research* **50**, 940–967.

Montrul S, and Rodríguez Louro C (2006) Beyond the syntax of the null subject parameter: A look at the discourse-pragmatic distribution of null and overt subjects by L2 learners of Spanish. In Torrens V and L Escobar (eds), *The acquisition of syntax in romance languages*. Amsterdam: Benjamins, pp. 401–418.

Montrul S (2004) Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition* **7**, 125–142.

Montrul S (2006) Incomplete acquisition as a feature of bilingual and L2 grammars. In Slabakova R, SA Montrul and P Prévost (eds), *Inquiries in linguistic development: In honor of Lydia White*. Amsterdam: Benjamins, pp. 335–359.

Piñango MM, and Burkhardt P (2005) Pronominal Interpretation and the Syntax-Discourse Interface: Real-time Comprehension and Neurological Properties. In Branco A, T McEnery and R Mitkov (eds), *Anaphora Processing: linguistic, cognitive and computational modeling*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 221–238.

Pozzan L and Trueswell JC (2017) Second language processing and revision of garden-path sentences: a visual word study. *Bilingualism: Language and Cognition* **19**, 636–643.

Prince EF (1985) Fancy syntax and shared knowledge. *Journal of Pragmatics* **9**(1), 65–81.

Rosa EC and Arnold JE (2011) The role of attention in choice of referring expressions. In Proceedings of PRE-Cogsci: Bridging the gap between computational, empirical and theoretical approaches to reference, Boston.

Rothman J (2007) Pragmatic Solutions for Syntactic Problems: Understanding Some L2 Syntactic Errors in Terms of Pragmatic Deficits. In Baauw S, F Dirjkoningen and M Pinto (eds), *Romance languages and linguistic theory 2005*. Amsterdam: John Benjamins, pp. 299–320.

Rothman J (2008) How pragmatically odd! Interface delays and pronominal subject distribution in L2 Spanish. *Studies in Hispanic and Lusophone Linguistics* **1**, 317–339.

Rothman J (2009) Pragmatic deficits with syntactic consequences? L2 pronominal subjects and the syntax-pragmatics interface. *Journal of Pragmatics* **41**, 951–973.

Sedivy JC (2003) Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* **32**, 3–23.

Sorace A (2011) Pinning down the concept of 'interface' in bilingualism. *Linguistic Approaches to Bilingualism* **1**, 1–33.

Sorace A and Filiaci F (2006) Anaphora resolution in near-native speakers of Italian. *Second Language Research* **22**(3), 339–368.

Sorace A and Serratrice L (2009) Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism* **13**, 195–210.

Sorace A, Serratrice L, Filiaci F and Baldo M (2009) Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua* **119**, 460–477.

Travis C, Torres Cacoullos R and Kidd E (2017) Cross-language priming: A view from bilingual speech. *Bilingualism: Language and Cognition* **20**(2), 283–298.

Tsimpli I, Sorace A, Heycock C and Filiaci F (2004) First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism* **8**, 257–277.

Van Rij J, Van Rijn H and Hendriks P (2010) Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language* **37**(3), 731–66.

Vogels J, Krahmer E and Maes A (2015) How cognitive load influences speakers' choice of referring expressions. *Cognitive Science* **39**(6), 1396–1418.

https://doi.org/10.1017/S1366728920000176 Published online by Cambridge University Press