

ARTICLE

Identifying bilingual reference profiles: a cluster-analysis approach to reference production among Greek–Italian bilingual children

Margreet Vogelzang^{1,2}  and Jacopo Torregrossa³ 

¹Theoretical and Applied Linguistics, University of Cambridge, Cambridge, UK; ²School of Psychology, Newcastle University, Newcastle upon Tyne, UK and ³Department of Romance Languages, Goethe University Frankfurt, Frankfurt, Germany

Corresponding author: Margreet Vogelzang; Email: mv498@cam.ac.uk

(Received 16 October 2023; Revised 25 July 2024; Accepted 26 July 2024)

Abstract

Empirical studies on bilingual children's reference production have often focussed on comparisons with monolingual peers. In this study, we introduce the concept of 'reference profiles': Speakers may exhibit similar or different behaviours in reference production, independently of whether they belong to a specific group (e.g., monolinguals or bilinguals) or whether their production adheres to some norm. Thirty-seven Greek–Italian bilingual children ($M_{age} = 9;4$, range 7;10–11;6) performed narrative retelling tasks in both of their languages, as well as vocabulary tasks and various cognitive tasks. The results show that the children had a good mastery of reference (i.e., appropriately using null pronouns, full pronouns or full nouns) in both of their languages. Using cluster analyses, we identified two distinct reference profiles. Further investigation showed that these profiles differed in both their sustained attention and in the use of overspecified referring expressions in contexts where reference to the same referent was maintained. These results are interpreted in light of current cognitive theories of (bilingual) reference processing and emphasise the potential of reference profiles for the study of other domains beyond bilingual reference production.

Keywords: accessibility; bilingualism; cluster analyses; cognitive skills; reference production

1. Introduction

Bilingual children's language production abilities have been investigated from different perspectives over the last 20 years. However, many previous studies are based on an (unintended) normativity bias. That is, more often than not they report group comparisons between monolinguals and bilinguals, with monolinguals serving as the 'control group' (Rothman et al., 2023). As a result, any difference between bilinguals and monolinguals is interpreted in terms of a bilingual effect (both in typically



developing and in non-typically developing samples, e.g., Hambly et al., 2013; Serratrice & Hervé, 2015; Vender et al., 2018). This kind of analysis hides that monolinguals may exhibit the same variation as observed among bilinguals in certain aspects of language production and also hides that some (sub-groups of) bilinguals may behave similar to monolinguals. The same argument could be applied to language comprehension.

In recent years, more and more researchers have emphasised the individual variation that can be found amongst bi- or multilingual adults and children focussing, for example, on the development of questionnaires to capture this variation (De Cat et al., 2023; Wigdorowitz et al., 2022), arguing for embracing heterogeneity in bilingualism (Prévost & Tuller, 2021) and suggesting methodological alternatives (Rothman et al., 2023). Although some of these suggested alternatives still include group-based comparisons, for example, between groups of heritage language speakers, the first alternative Rothman et al. (2023) propose is deemed most interesting for the current investigation: having no control/comparison group. In light of this, the present study aims to introduce the concept of a 'profile': Speakers may exhibit similar or different behaviours in language production, independently of whether they belong to a specific group (e.g., monolinguals vs. bilinguals) or their language production adheres to some criteria of appropriateness, which is in turn typically based on monolingual samples.

In the current study, we will focus specifically on the production of referring expressions (REs, henceforth), and therefore investigate 'reference profiles'. We operationalise this concept in terms of groups of speakers with similar reference-production behaviour, as identified here by means of a cluster analysis. The data-driven, 'bottom-up approach' used in this study allows us to identify clusters of speakers based on their reference-production behaviour without the need for a control/comparison group. As an additional advantage, this methodological approach allows us to investigate multiple morphosyntactic forms and discourse functions at once, which is rare in research on reference production. Once clusters have been identified, we will examine in how far the speakers in a cluster share similar cognitive (based on measures of executive function [EF] and theory of mind [ToM]) and language-experience profiles. This way, we aim to shed new light on the relationship between reference production and linguistic and cognitive variables. In the following sections, we will discuss the use of REs in discourse, the role of cognitive abilities in RE production, and RE production in bilinguals in turn.

1.1. Referring expressions and their discourse functions

Reference production is a complex task that requires keeping track of discourse referents and selecting discourse-appropriate REs to refer to them. Whether a RE is appropriate depends on the status of the corresponding referent in discourse. For example, referents that are *maintained* in discourse tend to be more accessible (or prominent, see von Heusinger & Schumacher, 2019) than referents that are *newly introduced* or *reintroduced* in discourse after a hiatus. At the level of linguistic encoding, more accessible referents can be referred to with more reduced REs (e.g., pronouns or clitics), whereas less accessible ones may require (indefinite or

In the present study, we will examine in how far speakers differ from each other in how they map discourse functions onto the use of REs. The accessibility marking hierarchy will serve as some indication of (universal) pragmatic appropriateness, but we will not go as far as to (numerically) compare bilingual children's productions to any norms or control group (based on the considerations in [Section 1](#)).

1.2. Role of cognitive abilities in reference production

The production of REs is a complex task from both the linguistic and cognitive point of view ([Hendriks, 2016](#)). A referent's discourse accessibility corresponds to a mental representation of this referent in the speaker's mind. The task of the speaker is to map this mental representation onto the use of a RE ([Kibrik, 2011](#); [Torregrossa et al., 2019](#)). Importantly, the referent's mental representation interacts with the speakers' individual cognitive profiles, for example, EFs and ToM ([Hendriks, 2016](#)). This interaction motivates (at least part of) the observed inter-individual variation in reference production. Most of the studies considered in this section concern children. If a study refers to adult speakers, we will mention this explicitly.

EFs are a multifaceted construct consisting of several components ([Miyake et al., 2000](#)), including 'inhibition of prepotent responses, shifting mental sets, monitoring and regulating performance, updating task demands, goal maintenance, planning, working memory, and cognitive flexibility, among others' ([McCabe et al., 2010](#), p. 222). Updating refers to an individual's ability to remember a previous stimulus, link a current stimulus to a previous one and continuously integrate 'new' information into 'old' one ([Carriedo et al., 2016](#)). In reference production, referents should be kept in memory as discourse unfolds ([Swets et al., 2007](#); [Van Rij et al., 2013](#); [Vogels et al., 2015](#)) and a referent's current mention needs to be linked to its previous mention while new information about this referent is continuously integrated. Research has shown that children with better updating abilities use more appropriate references ([Whitely & Colozzo, 2013](#)).

Some studies have specifically found a relationship between limited updating abilities and the use of overspecified REs in bilingual children (e.g., [Torregrossa et al., 2021](#)). [Torregrossa](#) and colleagues interpreted this use of overspecified REs in terms of speakers' reduced ability to link a referent's current mention to its previous mention under limited updating abilities. This would lead to the production of REs encoding low degrees of accessibility in contexts that would allow for the use of more reduced REs ([De Cat, 2015](#)).

The use of overspecified REs may also be related to speakers' limited ability to keep referents in memory, which is also one of the functions of updating. As a result, the accessibility of a referent decays more rapidly than when the speaker has a high working memory capacity ([Hendriks & Vogelzang, 2020](#); [Vogelzang et al., 2021](#) for a computational model related to the comprehension of null pronouns). This low accessibility would correspond to the use of explicit REs ([Rosa & Arnold, 2011](#) on adult speakers), in compliance with the accessibility marking hierarchy described in [Section 1.1](#). Alternatively, however, if the speaker struggles to keep referents in memory and therefore fails to retrieve the full noun, an underspecified form may be used instead.

EFs are also related to sustained attention ('goal maintenance' in the quote by [McCabe et al., 2010](#) above), intended as the ability to maintain attention to a stimulus

(a referent) for an extended period of time (Fisher & Kloos, 2016). In the domain of reference production, less sustained attention may mean that a referent is less accessible in a speaker's mind (its accessibility fades away more easily than when attention is high). This would lead to a similar prediction as with low WM capacity, namely the use of overspecified REs.

All the above-mentioned studies investigated how variation in EFs affects the production of REs in referent maintenance contexts. However, different patterns may be observed when reintroducing a referent. For example, Van Rij et al. (2013) show that adult speakers with lower WM capacity tend to use underspecified REs when reintroducing a referent; they may use a pronoun to refer to an antecedent in object position, although pronouns show the tendency to refer to subject antecedents (see also Serratrice & De Cat, 2020 for similar results with bilingual children tested in their societal language English). This is because under limited WM capacity, speakers' mental representations of a referent's accessibility are driven more by factors such as recency of mention than the actual computation of discourse-cues (e.g., syntactic position of the antecedent; e.g., Almor et al., 1999; Van Rij et al., 2013). These results show that variation in EFs may affect reference production in different ways, depending on the discourse function to be expressed. Finally, it should be noted that some other studies have found no effects of WM on pronoun production or interpretation (for adults, see Arnold et al., 2018; for children, Kuijper et al., 2021).

Another relevant component of EFs is shifting mental sets, which refers to the ability to switch behavioural responses according to a certain context (Diamond, 2013). In the domain of reference production, this may correspond to the use of variable form–function relationships. For example, Bamberg (1987) has shown that young children often stick to a thematic subject strategy when telling narratives, whereby they tend to use reduced REs only in association with the main character. This may be due to their still developing cognitive flexibility. Cognitive flexibility may also be involved in the mapping between discourse functions and syntactic positions. For example, some children may tend to maintain reference in subject position, whereas others may maintain reference in both subject and object position, which leads to 'more "dynamic" patterns of switches (from subject to non-subject and vice versa)' (Bongartz & Torregrossa, 2020, p. 255).

The considerations discussed thus far concern how cognitive variables interact with the representation of a referent's accessibility in the speaker's mind. However, reference production is also a pragmatic process whereby a speaker selects a form that can be recovered (i.e., interpreted correctly) by the listener. ToM may be crucially involved in this process. ToM is a broad cognitive construct comprising several sub-skills, including the ability to grasp that others' beliefs and mental representations may not be the same as one's own. This ability to keep track of others' beliefs and mental states is related to reference production: speakers need to know what listeners think/know when determining whether a RE will be understood correctly. Children with developing ToM may take a more self-centred perspective (Hendriks, 2016). Therefore, development of ToM abilities is predicted to be associated with a decrease in the use of underspecified pronouns. Indeed, ToM has been found to be positively associated with the use of full nouns for referents with low accessibility (Kuijper et al., 2015 on production; Kuijper et al., 2021 on comprehension).

The present study will take into account all of these different types of cognitive abilities (updating, cognitive flexibility, ToM) and investigate how they relate to bilingual reference production.

1.3. Bilingualism and reference production

We have thus far established that reference production involves the interaction between cognitive and linguistic factors. Therefore, it is particularly interesting to analyse it from the perspective of bilingual language acquisition, given that bilingual children may exhibit a variety of linguistic profiles both intra-individually – since, for instance, they may have a different competence in each of their languages – and inter-individually – since, for instance, they may differ from each other in their competence in a language.

Indeed, the production of REs varies both within and across bilingual speakers (Montrul, 2004). Both overspecification and underspecification have been attested (Serratrice & Hervé, 2015; Sorace & Serratrice, 2009). Torregrossa et al. (2019, 2021) have argued that the type of reference-production pattern observed is related to the bilingual profile of the participants in terms of language exposure, with overspecification being associated with unbalanced bilinguals tested in the weaker language, and underspecification with balanced bilinguals. Contemori et al. (2024) report similar results in adult bilinguals: overspecified REs were produced in the non-dominant language (Spanish) and underspecified REs in the dominant one (English).

Torregrossa et al. (2019, 2021) proposed a processing account for this phenomenon: overspecification emerges if the syntactic options available in a language (i.e., functional categories such as pronouns) are not proceduralised. As a result, speakers rely on a pragmatic strategy consisting in the repetition of a full noun. Underspecification is shown to be an effect of reduced speed in lexical retrieval. For example, an underspecified null pronoun is produced if children are slower in retrieving the (contextually more appropriate) full noun.

Other studies have proposed a representational account of bilingual children's use of REs. For example, several studies have noticed that bilingual children speaking a combination between a null-subject and clitic-language (such as Italian) and a non-null-subject and non-clitic language (such as English) tend to produce full pronouns in subject and object position in the null-subject and clitic-language in contexts in which the use of null pronouns or clitics would have been more appropriate (Serratrice et al., 2004; Sorace & Serratrice, 2009; Tsimplici & Sorace, 2006). This has been interpreted in terms of cross-linguistic effects from the non-null-subject language to the null-subject one (Torregrossa & Bongartz, 2018), although similar effects can occur when children speak two null-subject languages (e.g., Sorace et al., 2009). Mishina-Mori et al. (2024) analysed reference production by English–Japanese bilingual children, with Japanese being a null-subject language and English a non-null-subject one. Interestingly, they observed the overproduction of full noun phrases in Japanese only in referent reintroduction contexts and concluded that only discourse contexts that require more processing resources (i.e., referent reintroduction vs. referent maintenance, according to their analysis) are more vulnerable to cross-linguistic effects. Cross-linguistic effects have also been found in the production of REs in Italian by Italian–Greek bilingual children. These children used more null pronouns for referent reintroduction than Italian monolinguals, arguably due to cross-linguistic influence from Greek to Italian (Andreou et al., 2023), given that Greek allows for the use of null subjects in referent reintroduction contexts to a greater extent than Italian (Torregrossa et al., 2020).

In addition to cross-linguistic effects, representational accounts may also examine in how far bilingual reference production is affected by development. In other words,

bilinguals' referencing abilities may still be developing as in the case of monolinguals or with a slight delay compared to monolinguals. For example, Serratrice and De Cat (2020) found that bilingual children's tendency to use underspecified REs decreased with increasing WM and language proficiency in English, with both variables strongly correlating with age. This suggests that bilingual children follow the same path of reference acquisition as their monolingual peers. Some of the observed differences in bilingual pronoun production have been attributed to cross-linguistic influence. Because cognitive variables, rather than cross-linguistic influence, are the main focus of the current study, we will investigate two languages with similar – although not identical – reference systems: Greek and Italian (Section 1.1).

It should be considered that overall, all the studies reviewed in this section based on the comparison between monolinguals and bilinguals report quantitative differences between the two groups. However, to the best of our knowledge, no study so far has shown any consistent or systematic violation of the accessibility marking hierarchy by bilinguals, which suggests that bilinguals are in general aware of the pragmatic principles underlying the use of REs in discourse in their two languages (Flores & Rinke, 2020; Torregrossa et al., 2019).

Another important caveat is that most of the studies above have analysed a limited set of REs (e.g., null pronouns, overt subject pronouns or full nouns) within one reference function (introduction, maintenance or reintroduction), without considering which types of reference functions are expressed by other types of REs in the system. There is a gap in the literature when it comes to empirically examining bilinguals' reference management across forms and across functions. Besides not needing a control/comparison group, the cluster analysis we will apply also has the advantage of being able to take into account different REs and reference functions at the same time.

2. The present study

Although first attempts at examining correlations between reference production and cognitive abilities have been made (Section 1.2), as of yet no clear *profiles* of bilingual reference production have been identified. In this study, we aim to identify distinct profiles of use of REs by bilinguals and link these to children's language background and cognitive abilities. In order to provide a comprehensive overview of bilingual children's abilities with regards to reference management, we have formulated three research questions, which will be investigated by means of a narrative retelling task in Greek and Italian.

- *RQ1*. Does bilingual children's reference production adhere to the accessibility marking hierarchy in both of their languages?

To this end, we distinguish three *discourse functions* (following Bamberg, 1987; Fichman & Altman, 2019; Serratrice, 2007; Whitely & Colozzo, 2013): 1) introduction of new referents; 2) maintenance of reference and 3) reintroduction of a previously mentioned referent. We expect bilingual children to produce pragmatically sensible REs in both of their languages (Flores & Rinke, 2020; Torregrossa et al., 2019; see also the considerations at the end of Section 1.3), and thus adhere to the

accessibility hierarchy. Because this is the premise for the later research questions, it needs to be established first.

- RQ2. Can distinct profiles of reference management in bilinguals be identified?

Even if children adhere to the accessibility marking hierarchy in both of their languages, there may still be inter-individual variation, especially because REs are flexible and oftentimes there is no one ‘correct’ form (Section 1.1). We will run exploratory cluster analyses to examine whether distinct profiles of reference management can be identified. This data-driven, bottom-up approach allows us to avoid examining children’s reference production based on norms or comparisons with a control group (in line with suggestions by Rothman et al., 2023).

- RQ3. How do these profiles relate to children’s language background and cognitive abilities?

If distinct profiles of reference management can be identified, these will be examined further in relation to children’s language background and cognitive abilities. Based on the background literature, we make different predictions for referent maintenance and referent reintroduction. For referent introduction, no specific predictions emerged from the literature so the analyses will be exploratory. For referent maintenance, we predict children with more limited updating and sustained attention abilities to use explicit forms (e.g., full nouns) in contexts that would allow for the use of reduced forms (e.g., null pronouns or clitics) (in line with Torregrossa et al., 2021, see also De Cat, 2015). However, note that some effects in the opposite direction have been found in that production of underspecified REs increases with lower WM (Serratrice & De Cat, 2020).

For referent reintroduction, we predict children with lower updating abilities to use more reduced forms in contexts where the use of explicit forms would be more appropriate (in line with Lehmkuhle & Lindgren, 2024; Serratrice & De Cat, 2020; Van Rij et al., 2013). Furthermore, we predict children with lower cognitive flexibility to stick to certain referencing strategies to a greater extent than children with higher cognitive flexibility, for instance, associating the expression of certain discourse functions with specific syntactic positions (Bamberg, 1987; Bongartz & Torregrossa, 2020). Finally, we predict children with lower ToM abilities to use underspecified reduced forms (Kuijper et al., 2015). These tendencies should be observed in both of the bilinguals’ languages.

3. Methods

3.1. Participants

Thirty-seven children (*Mean age* = 9;4, *range* 7;10–11;6; two dates of birth unknown²) attending an Italian school in Athens (Greece) participated in the research. In the school, Italian was the main medium of instruction for both content

²These children’s date of birth was unknown as the questionnaire was not filled out by their parents; we coded these children’s ages as missing values. The two children were still included in the dataset to get a maximally large sample of children.

and language learning with a total number of 24 h of instruction per week. Greek was taught as an additional language for about 5 h per week. Besides attending this school, no additional recruitment/inclusion/exclusion criteria were used. Both the parents and the teachers reported that none of the participants had previously identified speech, hearing or visual impairments. Most of the children ($n = 20$) were simultaneous bilinguals – being exposed to Greek and Italian from birth; 15 children were successive bilinguals: six first exposed to Italian from age 3 (upon entrance in kindergarten), seven first exposed to Italian from age 6 (upon entrance in primary school) and two first exposed to Greek from age 6. We did not receive the questionnaire from two parents. Three additional children participated but they did not complete the narrative retelling task in both of their languages and were therefore excluded from the analyses. Before conducting the study, the parents provided written informed consent and we reminded the children that they did not have to take part in the activity if they did not want to.

3.2. Tasks

The children performed a narrative retelling task in both Greek and Italian, as well as vocabulary (to establish language dominance), Wisconsin Card Sorting, ToM and 2-back tasks. Examples of the materials used in each task can be found in the Supplementary Material.

The children performed a narrative retelling task, namely the *Edmonton Narrative Norms Instrument* (ENNI; Schneider et al., 2005), in both Greek and Italian. The task contains a set of 13 pictures that present a story about four characters (stories ‘A3 – airplane’ and ‘B3 – balloon’ were used). Children listened to a recording of the story, and were then asked to retell it. The retelling mode was used in order to make sure that all children understood the story plot. Otwinowska et al. (2020) showed that the retelling mode benefits the complexity of the story structure but does not affect children’s use of specific structures (e.g., specific REs).

An expressive vocabulary task was done by the children in both languages (Greek and Italian). The lexical items were drawn from the *Word Finding Vocabulary Test* (Renfrew, 1995). We administered the task as a series of 50 pictures of objects (taken from the internet, see Supplementary Material) appearing on a PowerPoint presentation. The children got one point for each object that was named correctly either right after the presentation of the image or after providing a semantic cue (using the same cues for all participants). We assigned 0.5 points to those items that were correctly named after providing a phonemic cue (i.e., the first syllable of the corresponding word) (Greek *mean* = 35.6, *range* = 7–45; Italian *mean* = 34.3, *range* = 20–48). Based on this, their dominant language was identified as the language in which the child scored the highest: this language was used as the language of administration for the cognitive tests described below (25 children were dominant in Greek and 12 children were dominant in Italian).

The 2-back task (Kirchner, 1958) is a task requiring WM and updating. In this task, children saw changing slides with a single digit and were asked to press a button when the current digit was the same as the digit two slides back. The task contained 60 trials, 20 of which should have been responded to with a button press and 40 of which should have been ignored. A-prime scores (Zhang & Mueller, 2005) were calculated based on the number of correct and incorrect responses (hits and false

alarms) ($mean = 0.81$, $range = 0.55-0.97$). Instructions for this task are verbal, but the task itself relies solely on (language-independent) digits.

The Wisconsin Card Sorting Test (WCST, henceforth; Kongs et al., 2000) was used to measure cognitive flexibility in a relatively language-independent manner. Children were presented with four cards that contain symbols varying in shape, colour and number. They were asked to match another, novel, card to one of the four cards, without receiving instructions what to match it on. Thus, children had to figure out the underlying rule through trial and error. However, the underlying rule changed throughout the task. Perseverative errors occurred when there was a rule change, but the child continued with the same response strategy as before. Such errors were thus an indication of a failure to inhibit a response and switch to a different response strategy, reflecting cognitive flexibility. Perseverative errors are reported as a percentage out of the total of 128 trials ($mean = 14.9$, $range = 0-32$; lower scores indicate better performance). Once a correct rule is found, 'failure to maintain set errors' reflected cases in which children failed to continue with the same, correct, strategy and thus changed their strategy before this was appropriate ($mean = 2.8$, $range = 0-8$; lower scores indicate better performance). This can be taken to indicate distractibility or absence of continued attention (Figuroa & Youmans, 2013; Miles et al., 2021). One child did not perform the test because s/he was absent in the corresponding session.

Finally, ToM was tested through a task using silent videos of characters performing actions in a social situation (Devine & Hughes, 2013). The children were asked questions related to how the characters' mental states motivated their actions. Six questions were asked about the videos; four targeting first order ToM and two targeting second order ToM. Children were asked the questions in their dominant language and could answer in any language they wanted. Responses were scored as 0 (incorrect or irrelevant), 1 (factually correct answers with no explicit reference to false belief) or 2 (correct answers with explicit reference to false belief), following the coding procedure in Devine and Hughes (2013). For each child, we considered the total response score ($mean = 5.7$, $range = 1-10$; max score 12).

3.3. Procedure

Children were tested individually in a quiet room in three test sessions. The second session was conducted at least 1 week after the first one, whereas the third session between 1 day and 1 week after the second one. In the first session, we administered the narrative task and the vocabulary task either in Greek or Italian. In the second session, we administered the narrative and the vocabulary task in the other language, counterbalancing the order of the languages between the two test sessions. We made sure that half of the children retold the story A3 in Italian and half of the children did it in Greek. The same holds for story B3.

The narrative task was administered in the form of a PowerPoint presentation. The children listened to a pre-recorded version of the story (either A3 or B3) on headphones, while looking at a series of pictures. The story was told by a female voice. Then, they had to retell the story to the experimenter, while looking at the same series of pictures previously shown. The experimenter did not have access to the pictures. The children were thus encouraged to provide as many details of the story as possible (Torregrossa & Bongartz, 2018 for methodology). The narrative task was audio-

recorded before being transcribed. The expressive vocabulary task was also administered in the form of a PowerPoint presentation. The experimenter documented whether the children provided a correct answer, a correct answer after a semantic cue, a correct answer after a phonemic cue, a wrong answer or no answer, using a scoring sheet. Children's answers were also audio-recorded, which allowed us to recheck each scoring sheet, correcting any possible coding error by the experimenter during the administration of the task. Within each session, we counterbalanced the order of administration of the narrative and the vocabulary task, respectively.

In the third session, we administered the cognitive tasks in the language that was identified as the dominant one based on the administration of the vocabulary tests in the first and second session. We counterbalanced the order of administration of the three tasks. The experimenters were a native speaker of Greek and a native speaker of Italian, who were responsible for the Greek and the Italian session, respectively (with the narrative and the vocabulary task). One of the two experimenters also administered the cognitive tasks, depending on the child's dominant language.

3.4. Coding and statistical analyses

All children were able to perform both of the narrative tasks; the number of clauses they used in their productions as well as the number of words per clause are presented in [Table S2](#) in the Supplementary Material. Clauses were defined based on the occurrence of a finite or nonfinite verb. For each clause, we coded all REs referring to animate referents or the main inanimate referent (the little airplane in story A3 and the balloon in story B3). REs were coded based on their morphosyntactic form and their discourse function. Morphosyntactic forms which were not sufficiently frequent in the dataset (relative pronouns, possessive adjective/pronouns, demonstratives, quantifiers) were excluded; this affected 11.66% of the data. Three thousand two hundred sixty-four REs, of six forms, remained: indefinite full nouns, definite full nouns, full pronouns, clitics, nulls and nonfinites (nonfinites only in Italian; see [Section 1.1](#)). Each RE was also coded for its discourse function, which relates to the accessibility of its antecedent in the discourse: 1) *introduction* of new referents (intro); 2) *maintenance* of reference (maintain) and 3) *reintroduction* of a previously mentioned referent (reintro). Some examples of the narrative production data and coding in both Greek and Italian can be seen in [Table 1](#). More details about the coding scheme as well as an example fragment can be found in the Supplementary Material.

All analyses were carried out using R (version 3.6.2; R Core Team, 2019). First, Pearson correlations were calculated with the package *Hmisc*, function *rcorr* (Harrell Jr, 2019) to examine the relation between the additional measures of vocabulary, WM and updating (2-back task), ToM, cognitive flexibility (perseverative errors in the WCST), and sustained attention (failure to maintain set in the WCST). This was done to check whether they reflect different constructs. This is especially relevant for the various cognitive measures, which are theoretically distinct but all require some type of executive function.

Then, we examined the frequency of each morphosyntactic form as relating to each discourse function in Greek and Italian separately. This will provide an overview of whether the REs that the children used are what would be expected based on their function in discourse. In order to statistically confirm the mapping of morphosyntactic forms to discourse functions, we conducted cross-sectional

Table 1. Example of utterances produced in the narrative task with the relevant coding

Participant	Language	Produced utterance	Morphosyntactic form	Discourse function
CH005	Italian	C'era <u>una cagnolina</u> <i>There was a little dog</i>	INDEF full noun	Intro
CH017	Greek	Ήτανε ένα <u>αγόρι</u> <u>καμηλοπάρδαλη</u> <i>It was a giraffe boy</i>	INDEF full noun	Intro
CH011	Italian	e <u>la cagnolina</u> saltò <i>and the little dog jumped</i>	DEF full noun	Reintro
CH017	Greek	<u>Ο καμηλοπάρδαλης</u> <u>χάρηκε</u> <i>The giraffe was happy</i>	DEF full noun	Reintro
CH005	Italian	<u>Lui</u> si affrettò <i>He hurried</i>	Full pronoun	Reintro
CH005	Greek	<u>Αυτός</u> έπαιζε <i>He was playing</i>	Full pronoun	Reintro
CH005	Italian	<u>a</u> <u>slegarlo</u> <u>a</u> <u>slegarlo</u> <i>to untie it</i>	Nonfinite Clitic pronoun	Maintain Reintro
CH008	Greek	και <u>το</u> του είπε και <u>το</u> του είπε <i>and to said to him</i>	Null pronoun Clitic pronoun	Maintain Reintro
CH011	Italian	però <u>lo</u> non ha riuscito <i>but didn't succeed</i>	Null pronoun	Maintain

Coded referring expressions on each line are underlined. More examples can be found in the [Supplementary Material](#). The full dataset is provided at https://osf.io/m2kwq/?view_only=7fe2391c5b2f4ff5b64828af0f8d155c.

k-means cluster analyses (one for Greek and one for Italian) with the *kmeans* function from the *factoextra* package (Kassambara & Mundt, 2020; see Hamann & Abed Ibrahim, 2017; Peristeri et al., 2022 for more examples of such cluster analyses). A cluster analysis computationally clusters data points into meaningful subsets (i.e., clusters) in an unsupervised manner, that is, without having any bias as to how the dataset should be split up and how many clusters should be present. *K*-means (MacQueen, 1967) is the name of the clustering algorithm applied; in this algorithm, each observation belongs to the cluster with the nearest mean. This technique was used to examine whether, from the morphosyntactic forms that the children used, their functions could be accurately predicted. The input was the percentage of usage of each morphosyntactic form for each of the three functions (introduction, maintenance, reintroduction) by each child. Thus, there are three data points per child, that is, one per discourse function per child, but each data point is an array with information about the production of each morphosyntactic form (i.e., type of RE). The output was a clustering of these data points. The optimal number of clusters was calculated based on the elbow method (Cui, 2020; using the *factoextra* package, *fviz_nbclust* function, Kassambara & Mundt, 2020). For both languages, the elbow method indicated that the optimal number of clusters was 3. Once clustering is complete, because we know the true function of each data point, we can check the quality of the clustering by calculating its accuracy (i.e., a match between the identified cluster and the actual use of REs for a particular function). However, note that cluster analyses are probabilistic and therefore each execution of the algorithm may lead to slightly different outcomes. In this study, to prevent the possibility of fishing for significant or desirable results, we only considered the clusters identified in the first run of the algorithm.

In the next step, we examined whether distinct profiles of reference management by bilinguals could be identified. To this end, the percentage of usage of each morphosyntactic form at three ‘time points’ (for three functions: introduction, maintenance, reintroduction) for each language was used as the input for a *k*-means cluster analysis (of repeated-measures trajectories, using the *kml3d* package, *kml3d* function, Genolini et al., 2015). The output was a clustering of children based on their production profiles in both languages. For this analysis, the Calinski and Harabasz criterion (Calinski & Harabasz, 1974) indicated that the optimal number of clusters was 2. Again, we only considered the clusters identified in the first run of the algorithm. Because in this case, there is no correct or incorrect production profile, no accuracy measure was calculated. Instead, we examined how the two clusters related to children’s language background and cognitive abilities. For this purpose, independent samples *t*-tests between the groups (clusters) were performed to address RQ3 and, in particular, the hypotheses that 1) the group of children who used more full pronouns or nouns in maintenance contexts have more limited EF abilities (updating or sustained attention, based on respectively the 2-back task and failure to maintain set in the WCST); 2) the group of children who use more null and overt pronouns in reintroduction contexts have more limited updating abilities or lower ToM scores and 3) the group of children who tend to associate discourse functions with certain syntactic positions show lower cognitive flexibility (based on perseverative errors in the WCST). Because of these specific hypotheses, one-sided *t*-tests were used for these tasks/measures, whereas two-sided *t*-tests were used for the vocabulary measures.

4. Results

4.1. Compliance with the accessibility marking hierarchy in both languages

We address RQ1 first. The descriptives of children’s use of REs (Figure 2) show that, in both languages, children used mostly indefinite full nouns for introductions of a new referent. This is as was expected based on the accessibility marking hierarchy, as

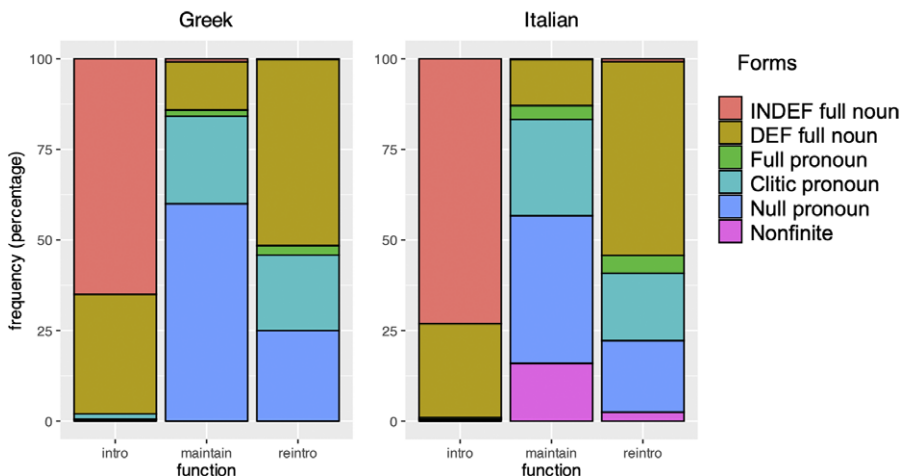


Figure 2. Morphosyntactic forms of referring expressions used for each function in Greek (left) and Italian (right), relating to their accessibility in discourse.

new referents have a low accessibility. Some definite full nouns were also used for this function. In maintenance positions, mostly nulls and clitics were used. Maintenance contexts involve reference to a highly accessible referent, and therefore selection of forms that are to the right of the accessibility hierarchy is appropriate. Note that in Greek, around 20% more null pronouns were used compared to Italian; This may relate to the fact that Greek speakers use null pronouns in contexts in which Italian speakers can use nonfinite forms (see Section 1.1). Finally, in reintroduction contexts the children mostly used definite full nouns. This too is in line with the accessibility marking hierarchy, as reintroductions are found in situations in which reference is made to an antecedent of medium accessibility. Some clitics and null pronouns were also used for reintroductions, which is interesting as these could lead to ambiguous reference.

We used cluster analyses (one for each language) to investigate whether the patterns in Figure 2 are sufficiently different for each discourse function to be identified as a distinct pattern based on each child's production of REs. For Greek, 103 out of 111 data points were classified as introduction, maintenance or reintroduction correctly, showing an accuracy of 92.8% (i.e., a match between the identified cluster and the actual use of REs for a particular function, reflecting very high classification accuracy, Dalmaijer et al., 2022). Four data points were incorrectly identified as reintroductions (these were actually introductions) and four data points were incorrectly identified as maintenance (these were actually reintroductions). For Italian, 104 out of 111 data points were clustered correctly, showing an accuracy of 93.7%. Four data points were incorrectly identified as reintroductions (three were actually introductions and one was maintenance) and three data points were incorrectly identified as maintenance (they were actually reintroductions). We refer to Table S5 in the Supplementary Material for a presentation of the three-cluster output obtained for the datasets of morphosyntactic forms in both languages.

The high classification accuracy shows that the children were very consistent in their production of form–function relationships. The results can further be interpreted as showing that the children mastered the conditions of use of REs in discourse by largely following the accessibility hierarchy, using mostly indefinite full nouns for introductions, null pronouns and clitics for maintenance, and definite full nouns for reintroductions in both languages. There are some small differences between the use of REs in Greek compared to Italian, but for the purposes of the current study it seems that in both languages, children produce largely pragmatically appropriate REs.

4.2. Identifying profiles of reference management in bilinguals

We now address RQ2. Cluster analyses identified two distinct groups of children (A and B; Figure 3 and Table 2), thus reflecting distinct profiles of reference production. Group A contained the majority of the children ($n = 28$), whereas group B ($n = 9$) represented a smaller group of children with different behaviour when it comes to reference production. We will describe the general trends in the clustering data below, although some observations are based on small percentage-based differences.

In introduction contexts, cluster A produced more definite full nouns, but fewer indefinite full nouns compared to cluster B in Greek (*means* of 33.7%

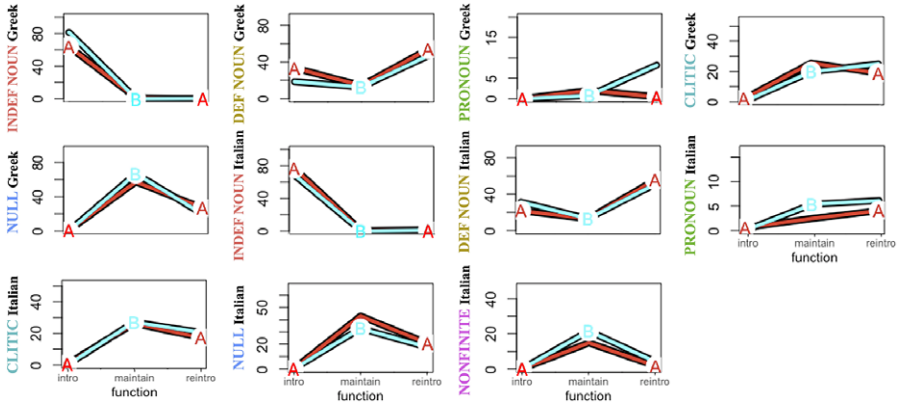


Figure 3. The two-cluster output obtained for the children’s use of morphosyntactic forms in each function (intro, maintenance, reintro) for both Greek and Italian. Note that the scales of the x-axes have been adjusted for each graph to zoom in on the differences between clusters A and B.

Table 2. The two-cluster output obtained for the children’s use of morphosyntactic forms in each function (intro, maintenance, reintro) in Greek (left) and Italian (right)

		Greek		Italian	
		Cluster A	Cluster B	Cluster A	Cluster B
Intro	INDEF full noun	63.8	81.5	77.0	68.8
	DEF full noun	33.7	18.5	21.7	31.2
	Full pronoun	0	0	0.6	0
	Clitic pronoun	1.8	0	0.7	0
	Null pronoun	0.6	0	0	0
	Nonfinite	–	–	0	0
Maintain	INDEF full noun	1.0	0	0.3	0
	DEF full noun	14.2	12.6	13.1	13.1
	Full pronoun	2.0	0.8	2.5	5.4
	Clitic pronoun	25.5	19.8	26.3	27.1
	Null pronoun	57.3	66.8	43.0	32.9
	Nonfinite	–	–	14.8	21.5
Reintro	INDEF full noun	0.2	0.3	0.9	1.0
	DEF full noun	53.9	46.9	55.4	50.9
	Full pronoun	0.5	8.2	4.1	6.1
	Clitic pronoun	18.7	24.9	17.3	20.6
	Null pronoun	26.7	19.6	20.7	17.5
	Nonfinite	–	–	1.6	3.9

vs. 18.5% for definite full nouns and 63.8% vs. 81.5% for indefinite full nouns). Interestingly, the opposite pattern can be observed in Italian, with cluster A producing fewer definite full nouns but more indefinite full nouns compared to cluster B (means of 21.7% vs. 31.2% for definite full nouns and 77.0% vs. 68.8% for indefinite full nouns). Although both forms are used for referent introduction in both languages, it is striking that the groups of children show opposite trends in their two languages. This points towards the possibility that the two groups vary in their language skills in the two languages. This possibility will be investigated in RQ3 (Section 4.3).

In maintenance contexts, some tendencies were observed in only one language. Most notably, children in cluster B used more full pronouns in maintenance position in Italian (5.4% vs. 2.5% in cluster A) and fewer null pronouns (32.9% vs. 43.0%). Furthermore, they used more nonfinite clauses (21.5% vs. 14.8%). Thus, clusters A and B show a preference for different pronominal forms here, although they both use pronominal forms versus full nouns at similar rates. In maintenance contexts, null pronouns are usually pragmatically appropriate because their antecedent should be clear, and thus it could be argued that the use of a full pronoun is unnecessary in maintenance contexts (i.e., it may be an overspecification). In Greek, children in cluster B use more null pronouns in maintenance contexts (66.8% vs. 57.3% in cluster A). In these contexts, they tended to use fewer clitic pronouns (19.8% vs. 25.5% in cluster A). This result seems to stem from a difference in grammatical role of the produced RE (i.e., subject vs. object) rather than any under- or overspecification.

In reintroduction contexts, we can observe some interesting differences between the clusters in both Greek and Italian (even if to a lower extent in Italian). In Greek, cluster B used up to 20% full pronouns (*mean* = 8.2%) in these contexts, whereas cluster A used fewer full pronouns (*max.* 8%, *mean* 0.5%). In contrast, cluster A used more null pronouns (*means* 26.7% vs. 19.6%). Children in cluster B use a greater number of clitics than the children in cluster A (*means* 18.7% vs. 24.9%). In Italian, these patterns are repeated although the differences are much smaller. The use of any pronoun in referent reintroduction could be seen as a risky, as it is potentially ambiguous. Therefore, it is interesting and possibly more straightforward to compare the children's use of noun phrases. Indefinite full nouns were hardly ever used for reintroductions by both groups. However, children in cluster A use more definite full nouns than children in cluster B, which is visible in both Greek and Italian (*mean* = 53.9% in cluster A vs. 46.9% in cluster B in Greek; *mean* = 55.4% in cluster A vs. 50.9% in cluster B in Italian). As full nouns are typically appropriate in reintroduction contexts, this could be interpreted to indicate that cluster A tends to be more accurate in reintroduction contexts than cluster B (but see Section 5 for an interpretation of the overproduction of null subjects by cluster A based on cognitive variables).

4.3. Language background and cognitive abilities

The final research question (RQ3) asked how the profiles identified in RQ2 relate to children's language background and cognitive abilities. Table S4 in the Supplementary Material reports the correlational analyses between the linguistic and cognitive measures considered in the present study. Italian vocabulary was significantly positively correlated with WM and updating (i.e., 2-back task performance) and ToM performance. No other significant correlations were found, and thus it seems that the different cognitive measures captured different constructs.

An overview of the children's scores per cluster is presented in Table 3. Cluster A was on average 4 months older, had higher ToM scores, and lower WCST scores (note that for both perseverative errors and failures to maintain set, lower scores indicate better performance). Of these comparisons, only the attentional measure (failure to maintain set in WCST) reached significance. Note that the two groups of children scored very similarly on their vocabulary measures, and therefore language

Table 3. Mean cognitive and language dominance scores and ages of children in the two clusters

	Cluster A	Cluster B	<i>p</i> -value
Greek vocabulary	35.7	35.5	0.97
Italian vocabulary	33.9	35.5	0.39
2-back (A-prime)	0.81	0.83	0.26
ToM	5.8	5.4	0.36
WCST (perseverative errors)	13.8	18.7	0.10
WCST (failure to maintain set)	2.5	3.9	0.04*
Age (months)	113.3	108.9	0.40

* $p < .05$

proficiency and/or dominance cannot account for the identified profiles of reference production in bilinguals.

5. Discussion

5.1. *The production of discourse-appropriate referring expressions in Greek and Italian*

The first result emerging from the present study is that the children were able to produce appropriate and consistent REs in discourse. In line with the accessibility marking hierarchy, in both languages, they used indefinite and definite full nouns for referent introduction, definite full nouns for referent reintroduction and null pronouns and clitics for referent maintenance (Figure 2). We also noticed some slight differences between Greek and Italian. In Greek, the children used a greater number of null pronouns for referent maintenance than in Italian. This is because Greek does not allow for nonfinite clauses. Note that nonfinites are not frequently examined as REs and one could argue that they should not be considered as such – this may be a confound in the current study.

In addition, children used more null pronouns in referent reintroduction in Greek than in Italian. This is consistent with the tendency observed in the literature that null pronouns in Greek are more inclined to be used in referent reintroduction – for instance, when shifting from object to subject position – than null pronouns in Italian (Andreou et al., 2023; Torregrossa et al., 2020). The cluster analysis reported in Section 4.1 further supported the conclusion that the children were consistent in their use of REs in both of their languages and that they showed different RE production patterns for each discourse function: The algorithm was very accurate in predicting discourse functions based on the frequency of use of one RE or the other. On the whole, these results suggest that the children had a good mastery of reference in both languages and, on top of that, showed some differences in the production of REs between their two languages.

This first result is somewhat surprising in light of previous literature on the acquisition of reference by bilingual children. However, it should be considered that the bilinguals examined in this contribution had a relatively balanced profile compared to the bilinguals considered in other studies, who tend to be dominant in the societal language (Benmamoun et al., 2013). For example, we observed that the children of the present study produced narratives of similar complexity in terms of number of clauses (Table S2 in the Supplementary Material) and their vocabulary scores were very similar across the two languages (Table 3). It is very likely that this balanced profile was related to their literacy exposure: the language which was the

main medium of instruction at school (Italian) was different from the majority language in society (Greek). This suggests that under certain conditions of language and literacy exposure, bilingual/heritage speakers may be able to exhibit a good mastery even of those structures that have been shown to be particularly difficult to acquire, like REs and their conditions of use (Bongartz & Torregrossa, 2020; Rinke & Flores, 2014; Torregrossa et al., 2023). Literacy exposure may be the key reason for the ability shown by the children considered in this study to produce discourse-appropriate REs, given that studies conducted on monolingual adults revealed that literacy exposure has a positive effect on reference production and comprehension (Arnold et al., 2018). However, it is not excluded that other factors may have played a relevant role for the results shown in this study. For example, as mentioned in Section 1.1, Greek and Italian pattern very similarly in the mapping between REs and discourse functions. This means that cross-linguistic effects are less likely to be visible in this group of children compared to other groups of children speaking combinations of languages that are typologically more distant from each other. The possibility of a cross-fertilization between the two languages is not excluded either, in compliance with observed cases of positive cross-linguistic influence in the domain of reference production (Section 1.3).

5.2. Identifying bilingual profiles of reference production

One of the main aims of this contribution was to provide a methodological approach for identifying reference-production profiles. In this way, we attempted to follow suggestions from the literature (Rothman et al., 2023) and overcome the bias in previous studies whereby reference production by bilinguals was compared to the golden standard of monolinguals. We did, however, use the accessibility marking hierarchy to provide some indication of pragmatic appropriateness of a RE. We followed a data-driven, bottom-up approach that allowed us to identify clusters of children based on their reference behaviour. In particular, we explored the potential of cluster analyses, considering the three discourse functions (introduction, maintenance and reintroduction) as ‘time points’. This kind of analysis allowed us to consider different mappings between morphosyntactic forms and discourse functions at the same time, improving on previous studies that have examined one discourse function at a time. This methodology has the potential to be applied to assess variation among other types of speakers beyond bilinguals, also considering other types of linguistic phenomena.

The analysis identified two clusters of children, exhibiting specific features in their reference production. In a nutshell, in introduction contexts cluster A produced more definite full nouns, but fewer indefinite full nouns compared to cluster B in Greek, whereas the opposite pattern was found in Italian. In maintenance contexts, the clusters of children varied from each other in their preference for different pronominal forms (nulls vs. clitics vs. full pronouns). These preferences also varied across the two languages. Finally, the children in cluster B produced more clitics and full pronouns, but fewer null pronouns in reintroduction contexts, which held in both languages.

Because of its unsupervised nature, the analysis identified two clusters which were unbalanced in terms of number of children, with one group consisting of 28 children and the other of 9 children. On the one hand, the observation that one cluster of children had a relatively small size is consistent with our previous consideration that

the children were homogeneous in their mastery of use of REs. This may be in turn related to their being relatively balanced in terms of language proficiency and experience (Section 5.1). On the other hand, the unbalanced sample size of the two groups may limit the power of the statistical analyses to be conducted in the next steps of the analysis (Section 5.3).

Before proceeding to the next section, some shortcomings of cluster analyses are worth pointing out. Firstly, since they are based on a probabilistic approach, re-running the analysis may lead to slightly different clusters than the ones identified in a previous analysis. In this study, we considered the two clusters identified in the first analysis that we ran. We refer to our OSF-materials the possibility to re-run the analysis and compare any new results with the ones reported in this study. Secondly, as stated before, because analyses of the type applied in this contribution are unsupervised, clusters can vary in size. This can limit the power of any follow-up comparisons, as it did in the current study due to the unbalanced cluster sizes. Finally, because all data points are taken into account regardless of their relation, outcomes of cluster analyses can be complex and difficult to interpret. We therefore focused our interpretation of the results mainly on the REs and discourse functions for which we had formulated concrete predictions.

5.3. Language and cognitive variables involved in bilingual reference production

The analysis of how children's cognitive abilities related to their reference profiles was limited by the fact that the two clusters of children were unbalanced in their size, which affected the statistical power of the conducted *t*-tests. However, certain tendencies emerged, which were (at least partly) consistent with our hypotheses related to the connection between reference production and cognitive abilities. Children in cluster B tended to be slightly younger, exhibit lower ToM abilities, make more perseverative errors (based on the WCST) and be less able to maintain set (based on the WCST) than children in cluster A. Children in cluster A had lower updating skills (based on the 2-back task). In particular, the two clusters differed significantly from each other only in their ability to maintain set (based on the WCST; Table 3). We did not find any differences in language proficiency between the clusters.

The primary and only significant result thus concerned children's ability to maintain set. In Section 2, we hypothesised that reduced sustained attention (failure to maintain set in the WCST) may lead to a rapid decay of the accessibility of a referent. This decay would correspond to the use of more explicit REs than required by the discourse context. The children in cluster B tended to have lower sustained attention than the children in cluster A (Table 3). Crucially, the children in cluster B produced a greater number of full pronouns for referent maintenance than the children in cluster A. This pattern was observed only in Italian. In general, full pronouns are redundant in contexts of referent maintenance, where null pronouns or clitics could be used without leading to ambiguity. Therefore, this result supports our hypothesis related to the relationship between lower sustained attention and the production of overspecified, redundant REs. As of yet no study has investigated this relationship. In this sense, this study managed to fill this gap.

Other results show trends in the expected direction, but no significant differences between the two clusters' cognitive measures. Therefore, these trends will only be

mentioned briefly, accompanied by an explicit call for further research with larger sample sizes. First, we predicted children with lower ToM abilities to use underspecified reduced forms for referent reintroduction (Kuijper et al., 2015). Cluster B showed non-significantly lower ToM scores, which could motivate their tendency to produce a smaller number of definite full nouns in referent reintroduction across their two languages. Hendriks (2016) argues that the automatising of ToM in reference production improves with increasing cognitive maturity. The observation that the children in cluster B tended to be younger than the children in cluster A could support this claim. However, note that age did not differ significantly between the groups either, and that the children had a wide age range, which may be a confound in the current study.

Second, we predicted that children with limited updating abilities would use more reduced forms in contexts where the use of explicit forms would be more appropriate in reintroduction contexts (in line with Serratrice & De Cat, 2020; Van Rij et al., 2013). Some evidence for this could be found in the 'risky' use of REs by children in cluster A, who had non-significantly lower updating scores (based on the 2-back task) and also produced more null pronouns for referent reintroduction in both Greek and Italian than the children in cluster B. On the contrary, we found no evidence for our hypothesis related to the connection between reduced updating abilities and the production of overspecified REs in referent maintenance (Torregrossa et al., 2021 for different results). However, it should be noted that the difference in updating abilities between the children in cluster A and the children in cluster B may not be large enough to affect the production of REs in maintenance contexts. The sample considered in Torregrossa et al. (2021) showed a much larger variation in updating abilities than the sample considered here.

Finally, we predicted children with reduced cognitive flexibility to stick to certain referencing strategies to a greater extent than children with higher cognitive flexibility in referent reintroduction. Children in cluster B exhibited non-significantly reduced cognitive flexibility (perseverative errors in the WCST) compared to the children in cluster A, and they also tended to express referent maintenance in subject position (by using null pronouns) in Greek to a greater extent than the children in cluster A, who expressed referent maintenance also in object position (by using clitics). More research is needed to confirm these trends.

5.4. Conclusion

In conclusion, despite the shortcomings of the methodological approach discussed in Section 5.2, these results show the potential of cluster analysis for the understanding of inter-speaker variation in reference production, without the need for comparison to a norm or monolingual control group. The method identified two groups with distinct reference patterns. Although they scored comparably on language background variables, we found that the group of children with decreased sustained attention tended to produce more overspecified REs. Overall, the results show that the children had a good mastery of reference in both of their languages, but different reference profiles could still be identified. This approach may in the future be extended to other types of speakers and to domains beyond reference production.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/langcog.2024.48>.

Data availability statement. Supplementary materials as well as the data and analysis code are available at https://osf.io/m2kwq/?view_only=7fe2391c5b2f4ff5b64828af0f8d155c.

References

- Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S., & Tyler, L. K. (1999). Why do Alzheimer patients have difficulty with pronouns? Working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain and Language*, 67(3), 202–227. <https://doi.org/10.1006/brln.1999.2055>
- Andreou, M., Torregrossa, J., & Bongartz, C. (2023). The use of null subjects by Greek–Italian bilingual children: Identifying cross-linguistic effects. In G. Fotiadou & I. M. Tsimpli (Eds.), *Individual differences in anaphora resolution: Language and cognitive effects* (pp. 166–191). John Benjamins.
- Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of Pragmatics*, 16, 443–463.
- Arnold, J. E., Strangmann, I. M., Hwang, H., Zerkle, S., & Nappa, R. (2018). Linguistic experience affects pronoun interpretation. *Journal of Memory and Language*, 102, 41–54.
- Bamberg, M. (1987). *The acquisition of narratives: Learning to use language*. De Gruyter Mouton. <https://doi.org/10.1515/9783110854190>
- Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, 39(3–4), 129–181. <https://doi.org/10.1515/tl-2013-0009>
- Bongartz, C., & Torregrossa, J. (2020). The effects of balanced biliteracy on Greek–German bilingual children's secondary discourse ability. *International Journal of Bilingual Education and Bilingualism*, 23(8), 948–963. <https://doi.org/10.1080/13670050.2017.1355888>
- Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1–27.
- Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., & Rucián, M. (2016). Development of the updating executive functions: From 7-year-olds to young adults. *Developmental Psychology*, 52(4), 666–678. <https://doi.org/10.1037/dev0000091>
- Contemori, C., Tsuboi, N., & Armendariz Galaviz, A. L. (2024). Referential choice in two languages: The role of language dominance. *International Journal of Bilingualism*, 28(3), 494–512. <https://doi.org/10.1177/13670069231170545>
- Cui, M. (2020). Introduction to the K-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, 1(1), 5–8. <https://doi.org/10.23977/geors.2020.030102>
- Dalmaier, E. S., Nord, C. L., & Astle, D. E. (2022). Statistical power for cluster analysis. *BMC Bioinformatics*, 23(1), 205. <https://doi.org/10.1186/s12859-022-04675-1>
- De Cat, C. (2015). The cognitive underpinning of referential abilities. In L. Serratrice & S. E. M. Allen (Eds.), *The acquisition of reference* (pp. 263–284). John Benjamins Publishing.
- De Cat, C., Kaščelan, D., Prévost, P., Serratrice, L., Tuller, L., & Unsworth, S. (2023). How to quantify bilingual experience? Findings from a Delphi consensus survey. *Bilingualism: Language and Cognition*, 26(1), 112–124. <https://doi.org/10.1017/S1366728922000359>
- Devine, R., & Hughes, C. (2013). Silent films and strange stories: Theory of mind, gender, and social experiences in middle childhood. *Child Development*, 84(3), 989–1003.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Fichman, S., & Altman, C. (2019). Referential cohesion in the narratives of bilingual and monolingual children with typically developing language and with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 62(1), 123–142. https://doi.org/10.1044/2018_JSLHR-L-18-0054
- Figueroa, I. J., & Youmans, R. J. (2013). Failure to maintain set: A measure of distractibility or cognitive flexibility? *Proceedings of the Human Factors and Ergonomics Society*, 57(1), 828–832. <https://doi.org/10.1177/1541931213571180>

- Fisher, A., & Kloos, H. (2016). Development of selective sustained attention: The role of executive functions. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive functions in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 215–237). American Psychological Association. <https://doi.org/10.1037/14797-010>
- Flores, C., & Rinke, E. (2020). Factors constraining subject expression in European Portuguese spoken in Hamburg. A bi-generational corpus investigation. *Discourse*, 26. <https://doi.org/10.4000/discours.10648>
- Genolini, C., Alacoque, X., Sentenac, M., & Arnaud, C. (2015). kml and kml3d: R packages to cluster longitudinal data. *Journal of Statistical Software*, 65(4), 1–34. <http://www.jstatsoft.org/v65/i04/>
- Hamann, C., & Abed Ibrahim, L. (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, 2, 16. <https://doi.org/10.3389/fcomm.2017.00016>
- Hambly, H., Wren, Y., McLeod, S., & Roulstone, S. (2013). The influence of bilingualism on speech production: A systematic review. *International Journal of Language & Communication Disorders*, 48, 1–24. <https://doi.org/10.1111/j.1460-6984.2012.00178.x>
- Harrell Jr, F. E., with contributions from Charles Dupont and many others (2019). *Hmisc: Harrell miscellaneous*. R package version 4.5-0. <https://CRAN.R-project.org/package=Hmisc>
- Hendriks, P. (2016). Cognitive modeling of individual variation in reference production and comprehension. *Frontiers in Psychology*, 7, 1–17. <https://doi.org/10.3389/fpsyg.2016.00506>
- Hendriks, P., & Vogelzang, M. (2020). Pronoun processing and interpretation by L2 learners of Italian: Perspectives from cognitive modelling. *Discours*, 26. <https://doi.org/10.4000/discours.10679>
- Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and visualize the results of multivariate data analyses*. R package version 1.0.7. <https://CRAN.R-project.org/package=factoextra>
- Kibrik, A. (2011). *Reference in discourse*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199215805.001.0001>
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55, 352–358.
- Kongs, S. K., Thompson, L. L., Iverson, G. L., & Heaton, R. K. (2000). *Wisconsin Card Sorting Test – 64 card computerized version*. Psychological Assessment Resources.
- Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, 43(13), 3231–3250. <https://doi.org/10.1016/j.pragma.2011.06.008>
- Kuijper, S. J. M., Hartman, C. A., & 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), 1–18.
- Kuijper, S. J. M., Hartman, C. A., & Hendriks, P. (2021). Children’s pronoun interpretation problems are related to theory of mind and inhibition, but not WM. *Frontiers in Psychology*, 12, 610401. <https://doi.org/10.3389/fpsyg.2021.610401>
- Lehmkuhle, L., & Lindgren, J. (2024). *Referring expressions in the oral narratives of German-speaking children: The influence of local and global accessibility factors and working memory capacity* [Poster presentation]. Child Language Symposium, Newcastle, UK.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (vol. 1, pp. 281–297). University of California Press.
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology*, 24(2), 222–243. <https://doi.org/10.1037/a0017619>
- Miles, S., Howlett, C. A., Berryman, C., Nedeljkovic, M., Moseley, G. L., & Phillipou, A. (2021). Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behavior Research Methods*, 53(5), 2083–2091. <https://doi.org/10.3758/s13428-021-01551-3>
- Mishina-Mori, S., Nakano, Y., Yujobo, Y. J., & Kawanishi, Y. (2024). Is referent reintroduction more vulnerable to crosslinguistic influence? An analysis of referential choice among Japanese–English bilingual children. *Languages*, 9(4), 120. <https://doi.org/10.3390/languages9040120>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morpho-syntactic convergence. *Bilingualism: Language and Cognition*, 7, 125–142.

- Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2020). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 23(9), 1083–1107. <https://doi.org/10.1080/13670050.2018.1434124>
- Peristeri, E., Silleresi, S., & Tsimpli, I. M. (2022). Bilingualism effects on cognition in autistic children are not all-or-nothing: The role of socioeconomic status in intellectual skills in bilingual autistic children. *Autism*, 26(8), 2084–2097. <https://doi.org/10.1177/13623613221075097>
- Prévost, P., & Tuller, L. (2021). Bilingual language development in autism. *Linguistic Approaches to Bilingualism*, 12(1), 1–32. <https://doi.org/10.1075/lab.21018.pre>
- Renfrew, C. (1995). *Word Finding Vocabulary Test*. (4th ed.) Speechmark Publishing.
- Rinke, E., & Flores, C. (2014). Heritage Portuguese bilinguals' morphosyntactic knowledge of clitics. *Bilingualism: Language and Cognition*, 17(4), 681–699.
- Rosa, E. C., & Arnold, J. E. (2011). The role of attention in choice of referring expressions. In K. van Deemter, A. Gatt, R. van Gompel, & E. Krahmer (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1–6). Cognitive Science Society.
- Rothman, J., Bayram, F., DeLuca, V., Di Pisa, G., Duñabeitia, J. A., Gharibi, K., Hao, J., Kolb, N., Kubota, M., Kupisch, T., Laméris, T., Luque, A., van Osch, B., Soares, S. M. P., Prystauka, Y., Tat, D., Tomić, A., Voits, T., & Wulff, S. (2023). Monolingual comparative normativity in bilingualism research is out of “control”: Arguments and alternatives. *Applied Psycholinguistics*, 44(3), 316–329. <https://doi.org/10.1017/S0142716422000315>
- Schneider, P., Dubé, V. R., & Hayward, D. (2005). *The Edmonton narrative norms instrument*. University of Alberta, Faculty of Rehabilitation Medicine. <http://www.rehabresearch.ualberta.ca/enni>
- Serratrice, L. (2007). Referential cohesion in the narratives of bilingual English-Italian children and monolingual peers. *Journal of Pragmatics*, 39, 1058–1087.
- Serratrice, L., & De Cat, C. (2020). Individual differences in the production of referential expressions: The effect of language proficiency, language exposure and executive functions in bilingual and monolingual children. *Bilingualism: Language and Cognition*, 23(2), 371–386. <https://doi.org/10.1017/S1366728918000962>
- Serratrice, L., & Hervé, C. (2015). Referential expressions in bilingual acquisition. In L. Serratrice & S. E. Allen (Eds.), *The acquisition of reference* (pp. 311–333). John Benjamins.
- Serratrice, L., Sorace, A., & Paoli, S. (2004). Cross-linguistic influence at the syntax-pragmatics interface: Subjects and objects in English-Italian bilingual and monolingual acquisition. *Bilingualism: Language and Cognition*, 7(3), 183–205. <https://doi.org/10.1017/S1366728904001610>
- Sorace, A., & 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., & Baldo, M. (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of bilingual children. *Lingua*, 119, 460–477.
- Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, 136(1), 64–81. <https://doi.org/10.1037/0096-3445.136.1.64>
- R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Torregrossa, J., Andreou, M., & Bongartz, C. (2020). Variation in the use and interpretation of null subjects: A view from Greek and Italian. *Glossa: A Journal of General Linguistics*, 5(1): 95. <https://doi.org/10.5334/gjgl.1011>
- Torregrossa, J., Andreou, M., Bongartz, C., & Tsimpli, I. (2021). Bilingual acquisition of reference: The role of language experience, executive functions and cross-linguistic effects. *Bilingualism: Language and Cognition*, 24(4), 694–706. <https://doi.org/10.1017/S1366728920000826>
- Torregrossa, J., & Bongartz, C. (2018). Teasing apart the effects of dominance, transfer, and processing in reference production by German-Italian bilingual adolescents. *Languages*, 3, 36. <https://doi.org/10.3390/languages3030036>
- Torregrossa, J., Bongartz, C., & Tsimpli, I. (2015). Testing accessibility: A cross-linguistic comparison of the syntax of referring expressions. In *Proceedings of the 89th Annual Meeting of the Linguistic Society of America, Portland* (pp. 1–4). <https://doi.org/10.3765/exabs.v0i0.3046>
- Torregrossa, J., Bongartz, C., & Tsimpli, I. M. (2019). Bilingual reference production. A cognitive-computational account. *Linguistic Approaches to Bilingualism*, 9(4–5), 569–599.

- Torregrossa, J., Eisenbeiß, S., & Bongartz, C. (2023). Boosting metalinguistic awareness under dual language activation: Some implications for bilingual education. *Language Learning*, 73, 683–722. <https://doi.org/10.1111/lang.12552>
- Tsimpli, I. M., & Sorace, A. (2006). Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *BUCLD 30: Proceedings of the 30th Annual Boston University Conference on Language Development* (pp. 653–664). Cascadia Press.
- Van Rij, J., Van Rijn, H., & Hendriks, P. (2013). How working memory load influences linguistic processing in adults: A computational model of pronoun interpretation in discourse. *Topics in Cognitive Science*, 5(3), 564–580. <https://doi.org/10.1111/tops.12029M>
- Vender, M., Delfitto, D., & Melloni, C. (2018). Clitic production in bilingual children: When exposure matters. *Languages*, 3(3), 22. <https://doi.org/10.3390/languages3030022>
- Vogels, J., Krahmer, E., & Maes, A. (2015). How cognitive load influences speakers' choice of referring expressions. *Cognitive Science*, 39(6), 1396–1418. <https://doi.org/10.1111/cogs.12205>
- Vogelzang, M., Guasti, M. T., van Rijn, H. and Hendriks, P. (2021). How children process reduced forms: A computational cognitive modeling approach to pronoun processing in discourse. *Cognitive Science*, 45, e12951. <https://doi.org/10.1111/cogs.12951>
- von Heusinger, K., & Schumacher, P. B. (2019). Discourse prominence: Definition and application. *Journal of Pragmatics*, 154, 117–127. <https://doi.org/10.1016/j.pragma.2019.07.025>
- Whitely, C., & Colozzo, P. (2013). Who's who? Memory updating and character reference in children's narratives. *Journal of Speech Language and Hearing Research*, 56(5), 1625–1636. [https://doi.org/10.1044/1092-4388\(2013\)12-0176](https://doi.org/10.1044/1092-4388(2013)12-0176)
- Wigdorowitz, M., Pérez, A. I., & Tsimpli, I. M. (2022). Sociolinguistic context matters: Exploring differences in contextual linguistic diversity in South Africa and England. *International Multilingual Research Journal*, 16(4), 345–364. <https://doi.org/10.1080/19313152.2022.2069416>
- Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, 70, 203–212.

Cite this article: Vogelzang, M., & Torregrossa, J. (2024). Identifying bilingual reference profiles: a cluster-analysis approach to reference production among Greek–Italian bilingual children, *Language and Cognition* 16: 2081–2104. <https://doi.org/10.1017/langcog.2024.48>