

Opportunities and challenges in the analysis of the Multilingual Assessment Instrument for Narratives (MAIN)

First Language

2022, Vol. 42(2) 321–329

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01427237211064695

journals.sagepub.com/home/fla**Cécile De Cat** 

University of Leeds, UK

Abstract

The development of the Multilingual Assessment Instrument for Narratives (MAIN) has no doubt contributed to prompting a renewed interest in children's narratives. This carefully controlled test of narrative abilities elicits a rich set of measures spanning multiple linguistic domains and their interaction, including lexis, morphosyntax, discourse-pragmatics, as well as various aspects of narrative structure, communicative competence, and language use (such as code-switching). It is particularly well suited to the study of discourse cohesion, referential adequacy and informativeness, and of course to the study of narrative structure and richness, and the acquisition of a more formal or literary register. In this commentary article, I reflect on the five empirical papers included in the special issue. I focus on methodological challenges for the analysis of narratives and identify outstanding questions.

Keywords

Literacy, bilingualism, discourse-pragmatics, morphosyntax, narratives

Introduction

Tests from the Language Impairment Testing in Multilingual Settings (LITMUS) battery (Armon-Lotem et al., 2015) have become widely used to assess key aspects of bi/multilingual children's language, across a large (and ever growing) number of languages. As such, it has been a catalyst for greater comparability across studies focusing on typical and atypical language development in bi/multilingual children, but also increasingly in monolingual children.

Corresponding author:

Cécile De Cat, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK.

Email: c.decat@leeds.ac.uk

The advent of one such test, the Multilingual Assessment Instrument for Narratives (MAIN), has no doubt contributed to prompting a renewed interest in children's narratives in recent years. This carefully controlled test of narrative abilities elicits a rich set of measures spanning multiple linguistic domains and their interaction, including lexis, morphosyntax, discourse-pragmatics, as well as various aspects of narrative structure, communicative competence, and language use (such as code-switching). It is particularly well suited to the study of discourse cohesion, referential adequacy and informativeness, and of course to the study of narrative structure and richness, and the acquisition of a more formal or literary register. Its semi-spontaneous nature also allows a relative amount of freedom to the narrator, yielding insights into lexical richness, syntactic complexity, as well as morphosyntactic accuracy and fluency.

The articles in this special issue focus on the following aspects:

The discourse appropriateness of referential expressions in reference introduction (Lindgren et al., 2022) and reference maintenance (Andreou et al., 2022) or in both (Fichman et al., 2022; Hržica & Kuvač Kraljević, 2022);

The impact of cross-linguistic influence on referential choices (Lindgren et al., 2022; Otwinowska et al., 2022);

The impact of cross-linguistic influence and of DLD on morphosyntactic accuracy and referential choices (Andreou et al., 2022; Fichman et al., 2022);

The impact of gender-based ambiguity on the choice of NP versus pronoun (Hržica & Kuvač Kraljević, 2022);

Developmental effects in the encoding of new information by young bilinguals (Lindgren et al., 2022).

A summary of each study was provided in the introduction to the special issue (Gagarina & Bohnacker, 2022). In this commentary, I focus on methodological challenges for the analysis of narratives and identify outstanding questions.

Discourse appropriateness

The significance of results (both quantitative and qualitative) depends on how the data are apprehended, that is, it depends on the operationalisation of the variables of interest. For instance, discourse appropriateness is a multi-faceted phenomenon, which has given rise to a rich literature spanning formal pragmatics (Ariel, 1994; Buring, 2003), processing (Arnold, 2010; Arnold, Eisenband, et al., 2000; Arnold, Losongco, et al., 2000) and cognition (Gundel et al., 1993). How discourse appropriateness is operationalised for the analysis of narrative data requires choosing a set of analytical categories sufficiently represented in the available data, and defining the corresponding diagnostics clearly (so the analysis can be reproduced).

To be discourse-appropriate, the choice of linguistic form needs to match the information status of its referent. The choice affects referent explicitness (i.e. full noun phrase vs pronoun – with the added option of null elements in many languages) and definiteness.

The information status of the referent is determined by two dimensions: ambiguity potential (determined by the givenness and salience of the referent, for example, the recency of previous mention, or the availability of shared information) and discourse function (i.e. topic, focus). The two dimensions are related: focus is typically associated with new information and topichood requires the referent to be part of the common ground between speaker and addressee.

Most studies in this special issue operationalise discourse appropriateness as a binary variable.

Andreou et al. (2022) focused on referent maintenance contexts exclusively, and equated discourse inappropriateness with the use of an indefinite in such contexts. Reduced forms (such as null subjects, clitics and article drop) were automatically considered discourse-appropriate, irrespective of their salience and the presence of competitors. As shown in their Figure 1, there were hardly any indefinites in the data: performance according to the discourse appropriateness criterion was clearly at ceiling. It is unclear whether the same conclusion would have emerged from a more fine-grained analysis (e.g. focusing on the degree of ambiguity of null subjects, clitics and article drop, operationalised as an ordinal variable).

Fichman et al. (2022) focused on the appropriateness of pronoun use (vs full NPs) in introduction and maintenance contexts, and (in Hebrew only) on the appropriateness of definite articles in referent-introduction contexts. Hardly any pronouns were produced in introduction contexts, suggesting adequate levels of discourse competence. In maintenance contexts, pronoun inadequacy was equated with ambiguity ('a listener could not identify the referent') and subdivided in two types: discourse-pragmatics-based versus morphosyntax-based. I will come back below to the overlap between these two categories. The criteria used to determine the discourse-related ambiguity of pronouns were not clearly outlined in the methods section, but appear to have included not only the presence of competitors, but also recency of previous mention, and topichood. The very low number of pronouns (especially in Russian) demands caution in the interpretation of the results, and suggests that the operationalisation of discourse-pragmatic adequacy might not have been optimal for the available data. In the analysis of definiteness distinctions in the Hebrew data, discourse appropriateness is operationalised as a binary variable (assuming indefinites are obligatory for referent introduction and definites for referent maintenance) and article omission is treated as a morphosyntactic and semantic error. All groups appeared to over-use definites in introduction contexts, and it would be interesting to see if this was affected by the centrality of characters. Indefinites were only over-used in maintenance contexts by the BiDLD group, but the high prevalence of article omission in that group complicates the picture (especially as it is not included in Table 7). Indeed, in the discussion, the authors acknowledge that it is not possible to disentangle semantic specificity from discourse appropriateness as the cause for article omission.

Hržica and Kuvač Kraljević (2022) did not focus on discourse appropriateness *per se*, but on the use of pronouns versus full noun phrases. The key criterion in their study was the ambiguity potential of pronouns, especially in terms of gender cues. Pronoun use was examined across referential functions, under the assumption that it is always underinformative in referent-introduction contexts. Introduction and reintroduction contexts were merged into a single analytical category, but the criteria for distinguishing

reintroduction from maintenance were not provided. It is unclear whether this was based on linear distance between mentions, on the presence of an intervening referent, on the discourse status of intervening referents, or on topic discontinuity.

Lindgren et al. (2022) investigated the predictors of linguistic forms used for referent introduction. Three structures were distinguished: labelling (a stand-alone NP or an NP in a presentative copular sentence), canonical argument structure (subject, object), or narrative presentations (genre-specific as in *Once upon a time there was a cat.*). The authors argue that labelling structures do not have a narrative function and that they are merely descriptive. They suggest that labellings have a deictic dimension: deixis is either implied or instantiated with a gesture in the case of stand-alone indefinites, and it is inherent in the case of deictic subjects in 'predicative clausal constructions' such as *There is a cat.* While it is true that the use of deixis is infelicitous in the absence of shared visual information (as per the MAIN protocol), it is important to acknowledge the existential nature of indefinites in labellings, and the associated discourse function: by stating the existence of a referent, the child is introducing it in the common ground. The use of indefinites in labellings is therefore felicitous from a discourse-management point of view, even if it is sub-optimal from a narrative point of view. In that sense, labellings could be considered a precursor of the 'narrative presentations', in which the frame of reference is no longer defined deictically but in relation to a fictional story time.

Across studies, a key challenge in the analysis of narrative data is the measuring of salience and ambiguity. For instance, pronoun use is more likely to be felicitous in maintenance contexts, but it can still result in ambiguity, depending on the distance from antecedent or presence of competitors, as mentioned earlier. Similarly, characters can be introduced with a definite felicitously if their reference can be derived through bridging (Matsui, 1993), as in the case of 'baby birds' if a nest has been mentioned previously. Topichood is another important dimension, as it interacts with definiteness. The cognitive-computational approach of Torregrossa et al. (2018) is promising, for the assessment of the degree of activation of referents in narratives.

Teasing apart discourse-pragmatic from morphosyntactic aspects

Another key challenge in the analysis of narrative data is to tease apart the source of errors in the use of referential expressions. The discourse status of a referent is determined by the interaction of morphosyntactic, pragmatic and prosodic factors, and it is therefore impossible to fully disassociate discourse appropriateness from morphosyntactic accuracy.

Andreou et al. (2022), in their study of referent maintenance, show that DLD and cross-linguistic influence in bilinguals have an impact on the grammaticality of referential expressions, but not on their discourse appropriateness. Their methodology is exemplary in my view. Grammaticality and appropriateness are analysed separately, so that any referential expression is evaluated in both dimensions. In cases where grammaticality affects the evaluation of appropriateness, the assumptions underlying the analysis are clearly laid out and justified. For instance, article drop is considered appropriate (as it always appeared in topic-shift contexts) but ungrammatical in NPs. While the

main analysis focuses on the effect of DLD and cross-linguistic influence on rates of ungrammatical or inappropriate forms, secondary analyses explore the variability (in terms of frequency of occurrence across participant groups) across types of appropriate forms and across types of ungrammatical forms. This affords a fine-grained analysis of the effect of cross-linguistic transfer and DLD.

In Fichman et al.'s (2022) study, there is an overlap between discourse and morpho-syntax in the evaluation of 'adequacy'. The authors attempt to classify errors as either ungrammatical or inappropriate, and this forces them to decide which type of error takes precedence. For instance, gender errors were classed as morpho-syntactically inadequate, but it is not clear whether the level of salience of the referent was taken into account in such cases. In their example (2), for instance, a masculine pronoun is used to refer to a feminine noun in a context featuring two (feminine) nouns. Classifying the error as morphosyntactic obscures the fact that the use of any pronoun in this context leads to ambiguity. The impact of the overlap between the morphosyntactic and discourse-pragmatic dimensions was difficult to assess, however, given the lack of clarity of data presentation (as for example in their Table 6 reporting on pronoun adequacy).

In their account of the over-use of pronominals by Polish-English bilinguals, Otwinowska et al. (2022) argue that the DP model of transfer is more parsimonious than the NP model as it allows the use of a single framework to analyse monolingual and bilingual data. An even more parsimonious account of the overuse of D elements in Polish would, I suggest, *not* involve cross-linguistic transfer affecting syntactic representation: the overuse of pronominal elements in the narratives could be a manifestation of the child intentionally encoding referentiality on discourse-pragmatic grounds. It would stem from the prevalent use of overt elements in English, which would affect the child's evaluation of the need for overt elements in Polish. As suggested by Valian (2020) in her commentary of Polinsky and Scontras (2020), pronoun over-use by heritage speakers might aim to avoid ambiguity (given the variation they experience in the input), rather than to avoid dealing with 'silence'. In the absence of independent evidence for DP transfer, a discourse-pragmatic approach to the over-suppliance of pronouns is to be preferred, at least on grounds of parsimony. The latter approach predicts that interpretive properties of the referent (such as animacy, salience, topichood) affect the use of overt D elements in Polish. Further research will be necessary to determine if this is the case.

Otwinowska et al. (2022) investigate differences in MLUw (Mean Length of Utterances in words) (between bilinguals and monolinguals) in light of the over-use of pronouns by the bilinguals. The relationship between MLUw and informativity is implied but not investigated directly. What is unclear is whether MLUw should be interpreted as an indicator of structural complexity or informativity. If the latter, what does its apprehension *per communication unit* indicate, compared to a global measure across the narrative?

Lindgren et al. (2022) is the only study in which the syntactic structure hosting the referential expression was taken into account as a main factor in the analysis. They adopt a multi-dimensional approach to the use of indefinites for referent introduction, taking into account animacy and syntactic structure, and combining quantitative and qualitative analyses.

Semi-spontaneous or quasi-experimental?

While the context is partly constrained by the pictorial stimulus in this task, it also depends on how the child conceptualises the visual information into a coherent discourse and narrative. If a child fails to see the continuity between pictures and treats them each as individual scenes to describe, their repeated use of indefinites for the same referent is inadequate from a narrative task point of view but not from the point of view of the child's intention – see example (17) in Lindgren et al. (2022) as an illustration. Some of the 4-year-olds in their study needed repeated prompting to engage with the narrative task (a point I come back to below).

Treating referential functions (introduction vs maintenance) as quasi-experimental conditions assumes that the child is cognitively able to distinguish the two. In other words, narrative cohesion is a prerequisite for the evaluation of discourse-pragmatic appropriateness in this task; the *intention* to encode a referent as introduction or maintenance needs to be established. Interesting examples where this appears not to be the case are discussed by Fichman et al. (2022) and by Lindgren et al. (2022).

In future research, it would be interesting to study the use of experimenter prompts as scaffolds for the narrative task in studies aiming to assess discourse-pragmatic appropriateness. It would also be interesting to see if proficiency in young bilinguals has an impact on their ability to engage with the narrative task.

Protocol

The protocol for the MAIN aims to minimise the likelihood that the child will assume shared visual information with their interlocutor. Narratives can be elicited in telling or retelling mode. If the same experimenter tells the story and then listens to its retelling (as was the case in Otwinowska et al.'s 2022 study), this can fundamentally affect shared knowledge as the child can assume familiarity of the experimenter with the characters of the story and with the visual information available. Otwinowska et al. (2022) elicited narratives in both telling and retelling mode for each child, but did not include (telling vs retelling) mode as a factor in their analysis, on the grounds that there was no statistically significant difference in MLUw between the two modes (and no interaction between MLUw and bilingualism). What remains unclear is whether the proportion of DPs with overt D elements was higher in the retelling mode compared with the telling mode. Similarly, the use of deictic forms by the experimenter when prompting the child (as in Lindgren et al.'s 2022 study) could have increased the likelihood that the child could forget to take into account the fact that the experimenter could not see the pictures. See, for example, the prompt in Lindgren et al.'s (2022) example (15), which includes the deictic *da* 'there', implying a shared dimensional space. Note, however, that the MAIN manual lists standardised prompts which mostly do not include deictics. The manipulation of the picture book by the experimenter (for the unfolding of pictures) might also reduce children's ability to assume non-shared perspective, given the shared attention on the picture book implied by the manipulation.

Cognitive cost as a potential source of 'errors'

Hržica and Kuvač Kraljević (2022) propose to use the gender-based ambiguity of pronouns as a proxy for the cognitive complexity of stories involving three characters. In

Croatian, reference encoding in the Baby Goat story is thereby assumed more cognitively demanding than in the Baby Birds story (as the ambiguity potential of pronouns is greater in the former, given the identical gender of characters). Hržica and Kuvač Kraljević (2022) observe that, in the story where all three characters were of different gender, children used pronouns significantly less than adults for referent maintenance. A significant decrease in pronoun use was observed both in children and adults in the story with gender-identical characters, but the difference between stories was of a smaller magnitude in children (compared to adults).

Does the discourse-oriented approach make opposite predictions to the listener-oriented approach with respect to the use of pronouns versus full noun phrases, as claimed by Hržica and Kuvač Kraljević (2022)? The work of Arnold and colleagues (Arnold, 2008; Arnold et al., 2009; Arnold & Griffin, 2007) is presented as representative of the discourse-oriented approach. What Arnold and colleagues argue is that pronoun use depends on thresholds of referent activation (which may vary depending on speakers' cognitive profile), and that it may be determined more by speaker-internal constraints (which influence their discourse representation) than by the computation of the listener's mental state. There is however no claim that, in typically developing children, pronouns are intrinsically 'more difficult to produce than nouns', nor that speakers generally 'prefer specific referential forms'. In fact, evidence from young, typically developing children suggests that their use of pronouns is sensitive to the discourse context (Hickmann & Hendriks, 1999), and hence the activation levels of referents. Hržica and Kuvač Kraljević's (2022) evidence goes in the same direction: it suggests that children *are* sensitive to the presence of gender as a disambiguation cue (as they use pronouns less in the same-gender story), and that they are sensitive to the discourse context (as they only use pronouns for referent maintenance, and not for referent introduction). Children's greater preference for noun phrases in the different-character story (compared with adults) may well be due to the number of characters in the story, as it increases the overall ambiguity potential of pronouns. However, as the number of characters is not manipulated in this study, and the comparison was not based on a systematic review of the literature, further research will be necessary to elucidate this point. An explanation in terms of increased cognitive burden predicts individual differences between children, which should correlate with baseline cognitive measures (indicative of working memory and/or Theory of Mind). Further research will also be needed to demonstrate this is the case.

Individual variation

Many of the studies in this special issue reveal an interesting picture of individual variation, beyond the group results. Otwinowska et al.'s (2022) Figure 3, for instance, features a higher amount of variability in bilinguals (compared to monolinguals). Were Polish exposure and proficiency significant predictors of the use of referential markers in the bilinguals? Otwinowska et al. (2022) suggest that the likelihood of transfer (which they claim consists in the use of the English DP structure in Polish) is 'determined by factors such as quality of input and individual exposure'. However, the effect of Polish exposure is only investigated in relation to MLUw (and reported to be non-significant, at group level). Figure 3 is an invitation to investigate this further, using continuous predictors (i.e. PaBiQ and proficiency scores).

Lindgren et al. (2022) found no difference in performance between the bilinguals' two languages at group level, and no significant impact of language exposure or proficiency,

and advocate the need to go beyond frequencies. They provide an insightful qualitative analysis of individual differences, focusing on children's use of the presentation and labelling constructions, and revealing more advanced narrative abilities in the 6-year-olds (compared with the 4-year-olds).

Power

The narratives elicited by the MAIN are typically quite short. This can create challenges for quantitative analyses, unless a sufficiently large number of participants are included. With limited data, breaking down the analysis into many dependent variables gives a fragmented view and risks missing global effects (unless the two approaches are combined, as in Andreou et al., 2022).

Data exclusion can also increase the risk of unintentional bias. For instance, in Hržica and Kuvač Kraljević's (2022) study, instead of investigating full referential chains for each character, the analysis is restricted to pairs of referents in adjacent clauses. Six referential expressions were selected per participant, instantiating two levels of discourse status (vs maintenance) for each character. Less than six tokens were selected per participants if they did not mention all characters. These data selection procedures can only give a partial view of children's performance. For instance, could it be the case that 6-year-old children are better able to encode referent maintenance in adjacent clauses than in non-adjacent ones?

Final remarks

This collection of articles demonstrates the richness of narrative data elicited by the MAIN and their relevance for the study of typical and atypical language development in monolinguals and bilinguals. In particular, narrative data provide a privileged insight into the patterns of association and dissociation between grammar and discourse-pragmatics. The variety of methods employed across studies show this is a field in development, and I hope the questions raised in this commentary provide a constructive contribution.

Author contribution

Cecile De Cat: Conceptualization; Writing – original draft; Writing – review & editing.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Cécile De Cat  <https://orcid.org/0000-0003-0044-0527>

References

- Andreou, M., Peristeri, E., & Tsimpli, I. (2022). Reference maintenance in the narratives of Albanian-Greek and Russian-Greek children with developmental language disorder: A study on cross-linguistic effects. *First Language*, 42, 292–320.

- Ariel, M. (1994). Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics*, 30, 3–42.
- Armon-Lotem, S., de Jong, J., & Meir, N. (Eds.) (2015). *Methods for assessing multilingual children: Disentangling bilingualism from language impairment*. Multilingual Matters.
- Arnold, J. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23(4), 495–527. <https://doi.org/10.1080/01690960801920099>
- Arnold, J. (2010). How speakers refer: The role of accessibility. *Language and Linguistic Compass*, 4(4), 187–203.
- Arnold, J., Bennetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism: Effects of discourse status and processing constraints. *Cognition*, 110(2), 131–146. <https://doi.org/10.1016/j.cognition.2008.10.016>
- Arnold, J., Eisenband, J., Brown-Schmidt, S., & Trueswell, J. (2000). The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, 76(1), B13–B26. [https://doi.org/10.1016/S0010-0277\(00\)00073-1](https://doi.org/10.1016/S0010-0277(00)00073-1)
- Arnold, J., & Griffin, Z. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language*, 56(4), 521–536.
- Arnold, J., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28–55.
- Büring, D. (2003). On d-trees, beans and B-accents. *Linguistics and Philosophy*, 26(5), 511–545.
- Fichman, S., Walters, J., Melamed, R., & Altman, C. (2022). Reference to characters in narratives of Russian-Hebrew and Russian and Hebrew monolingual children with developmental language disorder and typical language development. *First Language*, 42, 263–291.
- Gagarina, N., & Bohnacker, U. (2022). A new perspective on referentiality in elicited narratives: Introduction to the special issue. *First Language*, 42, 171–190.
- Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive Status and the form of referring expressions in discourse. *Language*, 69(2), 274–307.
- Hickmann, M., & Hendriks, H. (1999). Cohesion and anaphora in children's narratives: A comparison of English, French, German, and Chinese. *Journal of Child Language*, 26, 419–452.
- Hržica, G., & Kuvač Kraljević, J. (2022). Referential choice in stories with characters of one or different genders: A study of monolingual Croatian speakers. *First Language*, 42, 216–233.
- Lindgren, J., Reichardt, V., & Bohnacker, U. (2022). Character introductions in oral narratives of Swedish-German bilingual preschoolers. *First Language*, 42, 234–262.
- Matsui, T. (1993). Bridging reference and the notions of 'topic' and 'focus'. *Lingua*, 90(1), 49–68. [https://doi.org/10.1016/0024-3841\(93\)90060-A](https://doi.org/10.1016/0024-3841(93)90060-A)
- Otwinowska, A., Opacki, M., Mieszowska, K., Białicka-Pikul, M., Wodniecka, Z., & Haman, E. (2022). Polish-English bilingual children overuse referential markers: MLU inflation in Polish-language narratives. *First Language*, 42, 191–215.
- Polinsky, M., & Scontras, G. (2020). Understanding heritage languages. *Bilingualism: Language and Cognition*, 23(1), 4–20. <https://doi.org/10.1017/S1366728919000245>
- Torregrossa, J., Bongartz, C., & Tsimpli, I. M. (2018). Bilingual reference production: A cognitive-computational account. *Linguistic Approaches to Bilingualism*, 9(4/5), 569–599. <https://doi.org/10.1075/lab.17026.tor>
- Valian, V. (2020). Variability: Definitions of language and language learning. *Bilingualism: Language and Cognition*, 23(1), 48–49. <https://doi.org/10.1017/S1366728919000609>