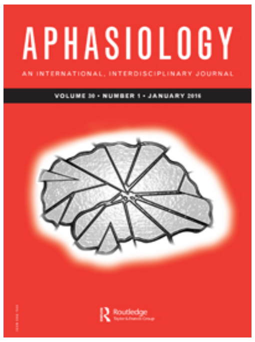

Aphasiology

ISSN: 0268-7038 (Print) 1464-5041 (Online) Journal homepage: http://www.tandfonline.com/loi/paph20

Main concepts for three different discourse tasks in a large non-clinical sample

Jessica D. Richardson & Sarah Grace Dalton

To cite this article: Jessica D. Richardson & Sarah Grace Dalton (2016) Main concepts for three different discourse tasks in a large non-clinical sample, Aphasiology, 30:1, 45-73, DOI: 10.1080/02687038.2015.1057891

To link to this article: https://doi.org/10.1080/02687038.2015.1057891

Aphasiology, 2016
Vol. 30, No. 1, 45–73, http://dx.doi.org/10.1080/02687038.2015.1057891

Main concepts for three different discourse tasks in a large non-clinical sample

Jessica D. Richardson* and Sarah Grace Dalton

Department of Communication Sciences and Disorders, University of South Carolina, Columbia, SC, USA

(*Received 26 April 2015; accepted 18 May 2015*)

Background: Semi-spontaneous speech production tasks are commonly elicited to assess discourse ability. When knowledge of a topic, story, or event is shared, it is possible to gauge the informativeness of discourse by evaluating how accurately and completely an individual produces the concepts considered to be essential to the shared topic. This analysis, main concept analysis (MCA), quantifies the degree to which speakers are able to communicate the overall gist of an event. Though MCA is an easy-to-perform, informative, and reliable measure of discourse adequacy, its widespread adoption depends on the development of standardisation and normative reference. Given the large collection of control discourse transcripts available on the AphasiaBank database, it is possible to generate main concept lists based upon a large sample of control speakers and to characterise their performance to establish preliminary normative reference.

Aims: The first aim of this study was to develop main concept checklists drawn from a control population for three semi-spontaneous discourse tasks included in the AphasiaBank protocol—a picture sequence narrative (Broken Window), storytelling (Cinderella), and a procedure (Peanut Butter and Jelly). The second aim was to report MCA results for control speakers to provide a normative reference and to stratify the normative information by age.

Methods & Procedures: Ninety-two control transcripts, stratified into four age groups (20–39 years; 40–59 years; 60–79 years; 80+ years), were downloaded from the AphasiaBank database. Relevant concepts were identified, and those spoken by at least one-third of the control sample were considered to be main concepts. A multilevel coding system was used to determine the accuracy and completeness of the main concepts produced by control speakers.

Outcomes & Results: Main concept checklists for three discourse tasks are provided. Descriptive statistics are reported and examined to assist readers with evaluation of the normative data. No differences between age groups were observed for the Broken Window narrative. For the remaining discourse tasks, the younger half of the sample generally performed differently than the older half of the sample. Additionally, the two younger age groups did not differ significantly from each other, nor did the two older groups.

Conclusions: This study provides main concept checklists drawn from a large control sample. Normative information for main concept production is provided for three discourse tasks. The sample distribution is evaluated relative to the normal probability distribution and the sample composition is described, enabling readers to determine the adequacy of normative characteristics of the sample and also the fit between their patient or client and the normative sample. This study also provides information about main concept or event production for four age groups.

Keywords: aphasia; AphasiaBank; discourse analysis; main concept; normative reference

*Corresponding author. Email: j.d.richardson@sc.edu
Present address of Jessica D. Richardson is Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, NM, USA.

Introduction

Discourse refers to a number of complex and individualised communication acts wherein individuals transmit and receive information for survival and cooperation, and use language for ritual purposes (e.g., fellowship, co-participation, kinship; Carey, 1988; Dimbleby & Burton, 1998). Aphasia is an acquired disorder characterised by impaired language production and/or comprehension that can result in life-altering changes in discourse abilities. Even persons with mild aphasia produce discourse samples that are characterised by reduced complexity, content, length, coherence, and lexical diversity (e.g., Andreetta, Cantagallo, & Marini, 2012; Capilouto, Wright, & Wagovich, 2006; Fergadiotis & Wright, 2011; Nicholas & Brookshire, 1995; Ulatowska, North, & Macaluso-Haynes, 1981). As aphasia severity increases, quality and quantity of relevant discourse decreases. Given the importance of discourse to life activities, it is not surprising that diminished communication abilities in persons with aphasia (PWAs) negatively impacts life participation and quality of life (QoL), particularly for those with chronic aphasia (approximately 10–30% of individuals who have survived a stroke or traumatic brain injury; Code & Petheram, 2011; Demır, Altinok, Aydin, & Köseoğlu, 2006; Go et al., 2014; Hilari, 2011; Kauhanen et al., 2000; Safaz, Alaca, Yasar, Tok, & Yilmaz, 2008). Notably, compared to post-stroke individuals without aphasia, PWAs demonstrate increased activity and participation limitations, higher rates of depression, and reduced QoL (Hilari, 2011; Hilari, Needle, & Harrison, 2012; Spaccavento et al., 2014).

Treatment approaches emphasising functional conversational and spoken discourse abilities are steadily evolving (see Boyle, 2011, for a review). Discourse measures documenting treatment response have varied across studies, ranging from word-level to narrative-level metrics. Measurement of discrete skills such as confrontational or generative naming, or global scales of aphasia severity, may predict performance on select discourse measures, but predictive strength is dependent upon the tasks and measures selected (Herbert, Hickin, Howard, Osborne, & Best, 2008; Mayer & Murray, 2003; Ross & Wertz, 1999; Ulatowska et al., 2003). Further, change in discourse ability may better predict socially valid change compared to standardised measures of overall language ability and functional communication (Ross & Wertz, 1999). Discourse measures are therefore strong candidates for detecting treatment response and predicting real-world communicative success. Different spoken discourse tasks have been investigated and utilised in clinical and research settings: spontaneous tasks including narration of personal events, informal conversations, or dialogue elicited via open questions (Prins & Bastiaanse, 2004); and semi-spontaneous tasks including procedural accounts of common activities, picture description, story generation using a series of pictures, storytelling, role-playing, etc. (Fergadiotis & Wright, 2011; Fergadiotis, Wright, & Capilouto, 2011; Prins & Bastiaanse, 2004). Different tasks provide different levels of support for semantic retrieval and facilitate different levels of output quality and quantity for both PWAs and control speakers (Armstrong, 2000; Capilouto, Wright, & Wagovich, 2005; Fergadiotis & Wright, 2011; Fergadiotis et al., 2011; Mayer & Murray, 2003; Ulatowska et al., 2003).

Whatever the purpose or genre, discourse production is fundamentally the communication of knowledge and beliefs to or with others (Van Dijk, 2014). It is hypothesised that humans develop an organisational cognitive schema (i.e., mental model) for knowledge representations that arise from life experience, where humans continuously undergo a series of spatio-temporally ordered events and situations with varied combinations of agents, settings, causal events, etc. (Van Dijk, 2014). These mental models are shared socially through discourse acts that involve the stringing together of propositions (e.g., concepts, sentences, statements, word-level meanings, that-clauses, fact concepts, text base, and/or conceptual chunks; Frederiksen, Bracewell, Breuleux, & Renaud, 1990; Kintsch & Van Dijk, 1978; Van Dijk, 2006, 2014). The words, phrases, and sentences used when expressing propositions form the discourse microstructure, and relevant concepts (RCs), cohesive devices, and topic markers form the discourse macrostructure (Kintsch & Van Dijk, 1978; Van Dijk, 2014). Microstructural discourse analysis includes examination of word class usage, syntactical construction, lexical diversity, or production of paraphasias, whereas macrostructural measures include coherence, cohesion, or story grammar (Armstrong, 2000). Proposition-level measures of correctness or completeness relate to both microstructure and macrostructure, since incorrect and/or incomplete concepts generally reduce overall narrative coherence (Brandão, Castelló, Van Dijk, Pimenta, & Peña-Casanova, 2009; Frederiksen et al., 1990; Kintsch & Van Dijk, 1978; Van Dijk, 2014). When the mental model for discourse is combined with a context model that determines the pragmatics of discourse, a more complete understanding of discourse is reached (Van Dijk, 2006, 2014). Because language production limitations experienced by PWAs affect their ability to share knowledge via proposition production, and generally (without co-occurring conditions) spare pragmatics, we have focused our discussion and research on the measurement and remediation of proposition-level deficits in this clinical population.

Main concept analysis (MCA) is one discourse analysis approach that focuses on the proposition-level expression of knowledge. A main concept (MC) corresponds to an utterance containing a subject, one main verb, and an object, if appropriate; it can also contain subordinate clauses, as long as it contains only one main verb (Nicholas & Brookshire, 1995). An MC list for a discourse task includes a closed list of those concepts constituting the "outline of the gist or essential information" for the task (Nicholas & Brookshire, 1995, p. 148). Each MC consists of semantic elements considered to be essential, and a multilevel coding system is applied to determine the accuracy and completeness of concept production (Boyle, 2014; Doyle, Goda, & Spencer, 1995; Hopper, Holland, & Rewega, 2002; Kong, 2009, 2011; Nicholas & Brookshire, 1993, 1995; Ross & Wertz, 1999). The first determination is whether a concept is present or absent (AB). If present, it receives one of the four following codes: accurate and complete (AC); accurate but incomplete (AI); inaccurate but complete (IC); and inaccurate and incomplete (II). Examining the proportion of AB codes allows us to determine how much of the essential information an individual attempted to produce. Utilising the additional codes allows researchers and clinicians to examine the quality of the information produced and provides more detail regarding the overall informativeness. MCA thus informs whether speakers are able to accurately and completely communicate concepts considered to be essential for expressing the overall *gist* of a discourse about which speakers share knowledge (Nicholas & Brookshire, 1993) and thus provides important information about both micro- and macrostructure (Armstrong, 2000; see also Brandão et al., 2009; Frederiksen et al., 1990; Kintsch & Van Dijk, 1978; Van Dijk, 2014).

Differences in informativeness between speaker groups can be detected using the coding system described earlier. Nicholas and Brookshire (1993) reported that control participants produced fewer MCs coded as AI or IC compared to PWAs, indicating that accuracy and completeness of attempted concepts might effectively differentiate the two groups. In a later study, control participants produced significantly more MCs that were AC, and significantly fewer MCs that were AI, IC, and AB, compared to PWAs (Nicholas & Brookshire, 1995; see also Kong, 2009). Kong (2009) reported that persons with fluent aphasia attempted to produce more MCs and were more likely to produce them accurately and completely than persons with non-fluent aphasia. Finally, MCA may be sensitive to improved language ability during spontaneous recovery (Nicholas & Brookshire, 1993) and following treatment (Hopper et al., 2002).

MCA corresponds to ecologically valid measures such as conversational ability or listener ratings of communication. Ross and Wertz (1999) asked listeners to rate change in discourse (i.e., better than, worse than, or no change) during recovery and characterised the relationships between listener ratings and several standardised impairment and discourse measures. The per cent of AC concepts correlated with listener ratings of overall communication abilities (r = .54), whereas no standardised impairment measures significantly correlated with listener ratings (Ross & Wertz, 1999). Cupit and colleagues (2010) asked listeners to rate narratives of two groups of PWAs—treated and untreated—focusing on informativeness, transmission ability, word-finding ability, and ease of retelling. The authors also reported proposition scores for each narrative using a system that assigned points to propositions based on characteristics similar to MCA codes AC (1 point) and AI (.5 points). PWAs in the treated group had higher listener ratings and higher proposition scores compared to those in the untreated group. Finally, Doyle et al. (1995) collected semi-spontaneous and spontaneous conversational discourse samples from PWAs using MCA as one of the outcome measures. Results showed a strong relationship (r = .71) between %AC MCs and an analogous conversational measure, percentage of informative minimal discourse units.

MCA is reliable across raters and testing sessions. Point-to-point inter-rater reliability for scoring of control and PWA participants' discourse is consistently above 80% (Boyle, 2014; Nicholas & Brookshire, 1993, 1995) and MCA scores from two different raters are significantly correlated (τ = .57–.96; Kong, 2009, 2011). Nicholas and Brookshire (1995) also reported point-to-point intra-rater reliability above 80% for MCA analysis of control and PWA participants' discourse. Kong (2009, 2011) reported significant intra-rater correlations (τ = .46–.99), with stronger correlations than those seen for inter-rater reliability. The test–retest reliability of MCA has been reported across both close (<2 weeks separating sessions; Boyle, 2014; Nicholas & Brookshire, 1993, 1995) and distant (approximately 1 year separating sessions; Kong, 2011) assessments. Despite the use of varied statistics, good reliability has been reported for both close and distant assessment sessions, and for the scoring codes AB, AI, and AC. Lower stability values have been reported for concepts that contain inaccurate information (i.e., IC and II).

Differences between age groups have been investigated, with mixed results, in discourse analyses similar to MCA. Age-related differences have been reported for overall efficiency of discourse (Wright, Capilouto, Srinivasan, & Fergadiotis, 2011), main event (ME) analysis (a proposition-level analysis capturing relationships between ideas; Wright, Capilouto, Wagovich, Cranfill, & Davis, 2005), and per cent correct information units (CIU) (which measures information at the word level; Capilouto et al., 2005). However, in other research no significant age-related differences were found for ME production (Capilouto et al., 2005; Wright et al., 2011). Differences in outcome measures, limits and ranges for age groups, and conflicting results make it difficult to determine what, if any, age differences exist.

In the majority of previous MC and ME research, 3–8 persons (investigators or speech-language pathologists [SLPs]) have generated a written list of MCs, MEs, or propositions for each discourse task under study following training. The investigators usually compare and combine these lists according to predetermined criteria to generate a master list of MCs, MEs, or propositions (Capilouto et al., 2005; Kong, 2009; Nicholas & Brookshire, 1993, 1995; Stark, 2010; Wright et al., 2005). Analysis then involves a determination of the number of MCs or MEs expressed during discourse relative to the number of MCs or MEs identified by investigators as being central to successfully communicating the gist using either a binary ("1" = present and accurate and "0" = absent, incomplete, or inaccurate) (Capilouto et al., 2006; Wright et al., 2011) or multilevel coding system (Kong, 2009, 2011; Nicholas & Brookshire, 1995). Few studies to date (Menn et al., 1998; Nicholas & Brookshire, 1995; Wright et al., 2011; Yorkston & Beukelman, 1980) have generated concept or proposition lists based on narratives of control speakers. Menn et al. (1998) created event checklists from small samples of both controls and PWAs (2–4 controls, 3–7 PWAs), while Nicholas and Brookshire (1993) trained 10 SLPs to write MC lists and validated the lists by examining productions of 20 control speakers. MCs written by 70% of SLPs and produced by 70% of control speakers were retained (Nicholas & Brookshire, 1993). Wright et al. (2011) sampled 20 control narratives to generate a list of events shared by 80% of speakers. Finally, Yorkston and Beukelman (1980) included every content unit (single words or groupings of words) produced by any of their 78 control speakers. Clearly, an array of methods have been employed, some of which may elicit biased target lists.

Researchers and clinicians who want to use MCA face several barriers. First, the MC lists, associated stimulus materials (e.g., pictures), and detailed assessment protocols (to facilitate standardised administration) either do not exist, are unavailable, or have limited availability to those outside academia. Second, there are shortcomings in the availability and/or quality of normative references. MC studies to date have often included small control sample sizes, restricted age groups that do not span adulthood, and little to no evaluation of distribution properties to support their use as a normative reference. Performance of controls has been documented primarily in evaluations of differences (e.g., between age groups or controls and clinical populations) and not to establish normative references or relevancy. Third, the practice of developing MC lists with input from biased samples (e.g., investigators, SLPs) or small (and presumably unbiased) control samples may undermine the relevance of the lists and the analytic approach. Given the centrality of the proposition (or MC) to discourse production and what is known about the impact of limited discourse abilities on life quality and participation, it would be well worth the expected effort to refine and make accessible an analytic approach that (a) has already proven to be informative and sensitive to group differences, (b) predicts ecologically valid measures and behaviours, and (c) can be administered reliably.

These obstacles to more widespread use of MCA could be overcome by using the AphasiaBank database (http://talkbank.org/AphasiaBank/) to develop MC lists and normative reference information. AphasiaBank is an online database of transcriptions and media files of PWAs and controls performing a variety of spontaneous and semi-spontaneous discourse tasks (Forbes, Fromm, & MacWhinney, 2012; MacWhinney, Fromm, Forbes, & Holland, 2011). AphasiaBank endorses a standardised discourse protocol (MacWhinney et al., 2011) with instructions and videos that can be used for training. Transcriptions are in CHAT (Codes for the Human Analysis of Transcripts) format, and CLAN (Computerized Language Analysis) programs can be used to perform analyses (MacWhinney, 2000). Given the richness of the AphasiaBank database, it would be possible to generate MC lists from a large sample of control speakers, ensuring a relevant foundation. Characterising the performance of controls in an MCA would also establish preliminary normative references for standardised AphasiaBank discourse tasks.

Measure development and qualitative analysis are two of the eight specific aims that motivated creation of the AphasiaBank database (MacWhinney, Fromm, Holland, Forbes, & Wright, 2010); therefore, reliance upon AphasiaBank data for this project is well aligned with the aims of the database. We sought to develop MC checklists drawn from a control population for three semi-spontaneous discourse tasks included in the AphasiaBank protocol and to report MCA results for control speakers. Specifically, we first determined MC checklists for picture sequence description, storytelling, and procedural tasks. Second, we performed MCA for control speakers and, as a secondary aim, characterised age differences in performance.

Methods

Transcripts

Control transcripts were downloaded from the AphasiaBank database, which at the time of this study included transcripts collected from 198 non-brain-injured adults between the ages of 20 and 89.5 (107 females, 91 males). Control participants in the database include typically ageing individuals with no history of neurological conditions who report English as their primary language. (All but one participant had hearing and vision adequate for testing.) Initial database screening revealed a strong age bias, with the majority of control participants age 60 and above. Ninety-two transcripts from the database were selected through stratified random sampling to guard against age-related bias in concepts (e.g., Capilouto et al., 2005; Wright et al., 2005). These 92 transcripts were contributed to AphasiaBank by the Capilouto, Kempler, Richardson and Wright laboratories and included 23 speakers not significantly different in gender and education from four age bins (20–39 years; 40–59 years; 60–79 years; 80+ years) who completed the three discourse tasks selected for this study (see Table 1).

Semi-spontaneous discourse tasks were selected for MC checklist development because MCA measures the ability to communicate essential concepts for a picture, story, or procedure about which speakers share knowledge (Nicholas & Brookshire, 1993). We selected one of each type of semi-spontaneous discourse task (picture description, storytelling, and procedural discourse) included in the AphasiaBank protocol in order to (a) include multiple and varied tasks since differences in discourse production have been reported as a function of genre and stimuli type (e.g., Capilouto et al., 2006; Fergadiotis et al., 2011); (b) include a more varied sampling of semi-spontaneous tasks that, if measured together, might better predict spontaneous speech; and (c) focus on commonly used tasks. Discourse tasks selected for this study thus included the Broken

Table 1. Demographic information for the 92 transcripts selected as the normative sample from the AphasiaBank database.

	N	Age (years)	Gender	Education (years)	Race/ethnicity
All	92	58.3 (±21.6)	55 females 37 males	15.6 (±2.5)	88 Caucasian 2 African American 2 Hispanic/Latino
20–39	23	29.6 (±5.8)	14 females 9 males	15.9 (±2.5)	20 Caucasian 2 African American 1 Hispanic/Latino
40–59	23	48.4 (±6.3)	15 females 8 males	15.7 (±2.5)	22 Caucasian 1 Hispanic/Latino
60–79	23	71.6 (±4.7)	13 females 10 males	15.7 (±2.4)	23 Caucasian
80+	23	83.9 (±2.9)	13 females 10 males	15.3 (±2.8)	23 Caucasian

Window picture sequence narrative (Menn et al., 1998), Cinderella storytelling (Grimes, 2005), and the procedure "how to make a peanut butter and jelly sandwich" (first introduced by Chandler, 1901; as referenced in Lau, 2013).

For the Broken Window narrative, participants were instructed to look at the picture sequence while telling a story with a beginning, middle, and end. For the Cinderella storytelling, participants first viewed the picture book by Grimes (2005) with the words covered. Participants looked at the book for as long as necessary, and when ready, the book was removed and the story told. For the peanut butter and jelly sandwich procedure (PB&J), participants were instructed to "tell me how you would make a peanut butter and jelly sandwich." Details regarding administration of the discourse protocol and transcription of discourse samples are included in MacWhinney et al. (2011), Forbes et al. (2012), and the AphasiaBank website (http://talkbank.org/AphasiaBank/). Using the GEM command from the CLAN tool, we isolated the selected discourse tasks from the rest of the transcript using this command (for Cinderella as an example): gem +sCinderella +d1 +fCinderella +t*PAR *.cha. The GEM command was used to create individual transcripts for each task (Broken Window, Cinderella, and PB&J).

Relevant concepts

To provide a list of MCs produced by control speakers, it was necessary to first identify the RCs produced during each discourse task. An RC was defined as an utterance related to the topic of the discourse task that contained a subject, one main verb, and an object, if appropriate. It could also contain subordinate clauses, as long as it contained only one main verb (Nicholas & Brookshire, 1993, 1995). RCs were thus candidate MCs from which the final MC list was established. The first time an RC was produced it was recorded in a running list for that task. The speaker who produced the RC received a score of "1" for production. Subsequent speakers who stated content comparable to that RC also received a score of "1", and the actual language used by the speaker to express the RC was retained and entered into the list. Although an RC is defined by its syntactic structure, the semantic content is of greatest interest and was entered into the RC list as a candidate MC. Speakers who did not state the RC received a score of "0." Speakers whose transcript was scored before the addition of a new RC to the running list retroactively received a score of "0" for new RCs. Each participant received a binary score of "1" (present) or "0" (absent) for each RC of a discourse task. We then summed across participants (column) for each RC (row), arriving at a frequency count of the number of times an RC was produced. Following training and supervised coding sessions conducted by the first author, the second author (a certified SLP) and two trained graduate speech-language pathology research assistants identified the RCs present in each transcript.

Because discourse is highly variable, RCs did not have to be produced with exactly the same wording in order to be considered the same (see also Nicholas & Brookshire, 1995). RCs across participants were judged to be comparable if they contained the same or similar subjects and/or verbs when conveying a relevant event (e.g., "The father remarried" and "The father married again" were determined to be the same RC in the Cinderella storytelling). Coders were instructed to enter concepts as new RCs if it was not clear that two concepts were the same. After completing RC coding of all transcripts for each discourse task, the authors examined RC lists and used forced agreement to determine whether any should be merged.

Main concepts

Frequency plots of the RCs for each discourse task were generated, where the *x*-axis represented the RCs and the *y*-axis represented the number of speakers (*N* = 92) who produced each RC. Plots were visually inspected for large natural gaps between RC productions. The locations of those gaps across discourse tasks were then compared to determine a shared threshold for determining which RCs would be considered "essential" and included on the MC checklists. The 33% cut-off was selected because it corresponded to natural gaps in the frequency plots of two tasks, Broken Window and PB&J. The Cinderella frequency plot was relatively free of large gaps, except two that occurred at 66% (12 RCs) and 78% (8 RCs) of speaker production, which excluded the majority of concepts generally accepted as essential to Cinderella storytelling. The 33% threshold was therefore applied to all discourse tasks so that any RC produced by 33% of the control speakers (rounded down to *N* = 30) was considered an MC. Given the discrepancy in natural gaps, we also report which MCs would survive 50% and 66% thresholds (see Appendices 1–3) to support future research efforts.

Each MC consists of several essential elements, and information about the accuracy and completeness of those essential elements is needed to determine the overall informativeness of a discourse sample. Using a multilevel coding system, we first determined whether a concept was present or absent (AB). If present, it received one of four codes based upon its degree of accuracy and completeness (Nicholas & Brookshire, 1995). In order to be coded as *accurate*, no essential elements of the MC could include incorrect information. To be *complete*, the speaker must have produced every essential element of the MC. Based on these definitions, AC concepts contained all elements of the MC with no incorrect information. AI concepts contained no incorrect information, but left out at least one essential element. IC statements contained at least one incorrect piece of information (e.g., "knight" for "prince"), but mentioned all essential elements. The coding of II was given when a statement clearly corresponded with an MC, but included at least one incorrect essential element and also failed to include at least one essential element (see Appendix 4 for statements that received each code). After MC coding, corresponding numerical scores were assigned and an overall MC score calculated for each speaker. Kong (2009) introduced the scoring formula where overall MC score = (3 × AC) + (2 × AI) + (1 × IC/II). We modified the scoring system so that the overall MC score = (3 × AC) + (2 × AI) + (2 × IC) + (1 × II). This modification separated the previously combined "inaccurate" scores (IC/II) (Kong, 2009; Nicholas & Brookshire, 1993, 1995), giving more value to IC statements than before. This was to prevent a scoring bias where semantic paraphasias, which would be coded as inaccurate, did not automatically receive fewer points than phonological paraphasias, which could be coded as accurate (along with misarticulations) "as long as they would be intelligible to a listener as the target words in the context of what the speaker is saying" (Nicholas & Brookshire, 1995, p. 154). For example, the statement "The cousins tore Cinderella's dress" (coded as IC) is at least as informative in conveying the gist of the story as "The stepsisters tore" (coded as AI) and more informative than "The cousins tore" (coded as II). Though we expected controls to primarily receive either AC or AB scores for most concepts, and we did not expect to encounter paraphasias, we included the multilevel scoring system to establish the reference for comparison of clinical populations in future investigations. This is especially important as PWAs may fall within the range of control speaker performance for AC and AB frequencies, but their production of the other codes on the accuracy-completeness continuum (AI, IC, II) is what is most distinguishable from control performance (Nicholas

& Brookshire, 1995). Following training and supervised coding sessions conducted by the first author, the second author and the same graduate research assistants scored accuracy and completeness of MCs for the discourse samples according to the guidelines established by Nicholas and Brookshire (1995).

Assessment fidelity

Discourse samples and transcripts were collected and transcribed by several different AphasiaBank contributors, potentially introducing variability to the assessment process, as different instructions, prompts, or materials could lead to discourse samples of different lengths and complexity. To measure adherence to the AphasiaBank assessment protocol, 33% of transcripts (with accompanying media) were randomly selected from the normative sample to determine how closely the assessor followed the prescribed assessment protocol. Following training by the first author, the second author and a graduate student in speech-language pathology inspected transcripts and videos for perfect adherence, omissions, or additions to the protocol. For approximately half of the reliability sample, assessor adherence to administration instructions could not be determined, as the transcripts and media had been edited to remove assessor dialogue. (However, the assessors did adhere to the protocol for providing prompts and materials while the participant completed the tasks.) For the remainder of the reliability sample, the assessors adhered to the protocol, with appropriate use of instructions, prompts, and materials.

Following training, the second author and two undergraduate students trained in transcription conducted point-by-point transcript reliability by comparing participants' productions from video or audio files to each line of the transcript. The percentage of omitted, added, and incorrect content words transcribed for each discourse task was calculated. For Broken Window, there were no omitted, added, or incorrect content words in the transcripts assessed for reliability. For Cinderella and PB&J, less than 1% of all content words were omitted, added, or incorrect. Given high assessor adherence to the AphasiaBank protocol and excellent transcription reliability, we feel confident using the transcripts from the AphasiaBank database for this study.

To determine reliability of MC coding, each rater re-scored transcripts for intra-rater reliability calculation, and the first author scored transcripts for inter-rater reliability calculation. Point-by-point (Kazdin, 1982) (or rather concept-by-concept) intra- and inter- rater reliability was conducted for 33% and 20% of transcripts, respectively. Intra-rater reliability for Broken Window, by age group from youngest to oldest, was 100%, 100%, 95%, and 97%. For Cinderella, intra-rater reliability was 91%, 91%, 91.9%, and 93.7%, and for PB&J 97.5%, 95%, 97%, and 95%. Inter-rater reliability, by age group from youngest to oldest, was as follows: Broken Window, 100%, 100%, 100%, 92.5%; Cinderella, 95.2%, 96.3%, 95.2%, and 95.6%; and PB&J, 96%, 90%, 94%, and 95%.

Data analysis

SPSS (Statistical Package for the Social Sciences) version 22.0 (Chicago, IL) was used to generate descriptive statistics and perform statistical analyses. Characteristics of the RC distribution for each discourse task were first examined to determine the suitability of the selected speaker sample for deriving MC checklists. Specifically, this examination involved assessment of descriptive statistics, assessment of skewness and kurtosis, and visual inspection of normality plots. For each discourse task, the following occurred: descriptive statistics were produced following coding and scoring of MCs; distribution characteristics were examined for acceptable normality; data were subjected to a one-way analysis of variance (ANOVA) (MC codes) or omnibus median test (overall MC score) to identify differences in MC codes and overall MC scores across the four age groups; and planned follow-up comparisons using Tukey honest significant difference (HSD) (MC codes) or pairwise median tests (overall MC score) were conducted to examine differences in the overall MC score and in each MC code across the four age groups. Non-parametric median tests with Holm–Bonferroni corrections for multiple comparisons were used to analyse the overall MC score because this variable does not meet the assumptions for ANOVA or Kruskal–Wallis tests. The MC code variables met the assumptions for use of ANOVA, and Tukey HSD tests (which correct for multiple comparisons) were utilised.

Distribution characteristics were generated so that (a) authors could examine them for acceptable normality and (b) readers would have the information needed to make decisions about the appropriateness of the norms for their use, compared to other existing or future normative data (Mitrushina, Boone, Razani, & D'Elia, 2005). What follows is a user-friendly description of evaluative procedures, and our application of these procedures is described in the results. The relationship between measures of central tendency, specifically the distance between mean and median, was examined to determine the symmetry of the distributions. When these values are equal or close together, the distribution is considered symmetric (Sullivan, 2003). Skewness values illustrate the symmetry of the distribution around the central peak, so that negative skew would indicate a greater number of larger values and positive skew would indicate a greater number of smaller values (Hopkins & Weeks, 1990; Keppel & Wickens, 2004). Kurtosis values characterise the magnitude of difference between the probability densities of the observed sample versus the normal curve, with negative kurtosis indicating that the sample distribution has fewer extreme values than a normal distribution and positive kurtosis indicating more extreme values (Hopkins & Weeks, 1990). Skewness >±2 and kurtosis >±4 (given SPSS correction for kurtosis; Keppel & Wickens, 2004; Kim, 2013) would indicate unacceptable non-normality (Fabrigar, Wegener, MacCallum, & Strahan, 1999; West, Finch, & Curran, 1995). Visual inspection of normality plots, or quantile–quantile (Q-Q) plots, allow readers to view observed quantiles of values from the control sample plotted against expected quantiles of values from a normal distribution. If our observed data points randomly cluster around the straight line, this would indicate a good match between the two distributions and confirm the suitability of the sample for development of norm-referenced measures. Statistical significance for ANOVA, omnibus median tests, and pairwise comparisons was defined as $p < .05$.

Results

Relevant concepts

For each discourse task, the following descriptive statistics for total number of RCs produced are displayed in Table 2: (a) mean, (b) standard deviation, (c) median, (d) range, (e) skewness, and (f) kurtosis. Mean and median were close in value, and skewness and kurtosis were within acceptable ranges (<±2 and <±4, respectively), indicating a sample distribution of acceptable symmetry. The smallest deviation from the normal distribution was observed for Broken Window, followed by PB&J and then Cinderella. Supplemental Figure 1 displays Q-Q plots for RCs for the three discourse tasks, with most data points clustered tightly around the straight line of the normal distribution.

Table 2. Descriptive statistics for relevant concept (RC) production on discourse tasks.

	Broken Window	Cinderella	PB&J
Mean	9.53	43.57	9.78
SD	2.18	21.83	3.79
Median	10	43.5	9
Range	4–14	5–144	3–19
Skewness	−0.009	1.218	0.428
Kurtosis	−0.339	3.885	−0.564

Main concepts

MC checklists are presented in Appendices 1–3. There were 8 concepts shared by at least 33% of our sample for Broken Window, 34 for Cinderella, and 10 for PB&J. Essential elements for each MC are listed with information about alternative word choices and sentence structures (as in Nicholas & Brookshire, 1995; Wright et al., 2005), and in some cases accompanied by additional non-essential content that was commonly (but not for 33% of speakers) produced with that MC. Also identified are the concepts included in 50% and 66% cut-off criteria, developed for future clinical utility research.

For each discourse task, the following descriptive statistics for MC codes and overall scores are presented in Tables 3–5: mean, standard deviation, median, range, skewness, and kurtosis for MC codes AC and AB; mean, standard deviation, median, and range for codes with extremely low or no occurrence (i.e., AI, IC, II); and median, range, skewness, and kurtosis for overall MC scores (mean and standard deviation excluded due to the ordinal nature of the variable). Statistics are presented for all speakers in the sample collapsed into a single group, and also divided into the four age groups. The maximum values of MC codes ranging from AC to AB are equal to the number of MCs on the checklist for the discourse tasks, and the maximum MC overall score is equal to the number of MCs on the checklist multiplied by 3, which would indicate that the speaker produced all MCs accurately and completely. Thus, MC overall maximum scores are 24 for Broken Window, 102 for Cinderella, and 30 for PB&J.

Broken Window

For each age group and for the entire sample considered together, means and medians for Broken Window were close in value for all MC codes as well as for the overall MC score. Skewness and kurtosis were within acceptable ranges for codes AC and AB, and the overall MC score. Supplemental Figure 2 displays Q-Q plots for the AC code for Broken Window, with most data points clustered tightly around the straight line of the normal distribution. The AC code was selected for display because it predicts the shape and distribution of both the AB code and the overall MC score since the other codes (AI, IC, II) rarely occurred in this sample.

Cinderella

The means and medians for Cinderella also show clustering around the straight line of the normal distribution across age groups and codes. Skewness and kurtosis were within acceptable ranges for codes AC, AB, and overall MC score for each age group and for the entire sample. Supplemental Figure 3 displays Q-Q plots for the AC code for Cinderella.

Table 3. Descriptive statistics for each main concept code for the Broken Window narration of the entire normative sample and each age group separately.

	All	20–39	40–59	60–79	80–99
Accurate–complete					
Mean	5.6	5.9	6	5.6	4.9
SD	±4.4	±1.3	±1.3	±1.6	±1.9
Median	6	6	6	6	5
Range	1–8	3–8	3–8	2–7	1–8
Skew	−0.772	−0.338	−0.499	−1.088	−0.375
Kurtosis	0.319	−0.030	−0.192	0.523	−0.461
Accurate–incomplete					
Mean	0.02	0.04	0.04	0	0
SD	±0.15	±0.21	±0.21		
Median	0	0	0	0	0
Range	0–1	0–1	0–1	0	0
Inaccurate–complete					
Mean	0.12	0.04	0.04	0.17	0.22
SD	±0.36	±0.21	±0.21	±0.5	±0.42
Median	0	0	0	0	0
Range	0–2	0–1	0–1	0–2	0–1
Inaccurate–incomplete					
Mean	0	0	0	0	0
Absent					
Mean	2.25	2.0	1.9	2.2	3.1
SD	±1.5	±1.2	±1.3	±1.4	±0.33
Median	2	2	2	2	3
Range	0–6	0–4	0–5	1–6	0–6
Skew	0.701	0.088	0.517	1.672	0.058
Kurtosis	0.106	−0.738	0.260	2.752	−0.419
Main concept score (24)					
Median	18	18	18	18	15
Range	5–24	11–24	9–24	6–21	5–24
Skew	−0.721	−0.158	−0.428	−1.204	−0.342
Kurtosis	0.170	−0.556	−0.009	0.962	−0.671

PB&J

Means and medians for all MC codes and age groups were close together for PB&J. Distribution of codes AC, AB, and the MC overall score had acceptable skewness and kurtosis. Supplemental Figure 4 displays Q-Q plots for the AC code for PB&J.

Age differences

Some age groups had significantly different MC production for two of the three discourse tasks (Tables 3–5 and Supplemental Figures 2–4), notably between the two younger age groups and the two older age groups. Compared to the older speakers (60 and older), younger speakers (59 and younger) produced more AC concepts on average, and thus had higher overall MC scores. The omnibus median tests revealed main effects for age group for the overall MC score for the Cinderella and PB&J discourse tasks (χ^2[3, N = 92]): Cinderella, $\chi^2 = 24.142$, $p < .001$; PB&J, $\chi^2 = 8.830$, $p = .032$; but not for Broken Window ($\chi^2 = 2.254$, $p = .521$). ANOVAs were run for each discourse task and MC code,

Table 4. Descriptive statistics for each main concept code for the Cinderella narration of the entire normative sample and each age group separately.

	All	20–39	40–59	60–79	80–99
Accurate–complete					
Mean	18.3	20.8	21.9	16.9	13.4
SD	±6.2	±3.6	±4.5	±5.8	±6.7
Median	20	21	22	18	13
Range	2–30	11–30	10–29	5–26	2–26
Skew	−0.699	−0.281	−1.007	−0.412	0.053
Kurtosis	0.005	2.767	1.362	−0.393	−0.573
Accurate–incomplete					
Mean	0.58	0.78	0.48	0.48	0.57
SD	±0.82	±0.95	±0.85	±0.73	±0.79
Median	0	0	0	0	0
Range	0–3	0–3	0–3	0–2	0–2
Inaccurate–complete					
Mean	0.46	0.48	0.52	0.22	0.57
SD	±0.7	±0.9	±0.67	±0.52	±0.59
Median	0	0	0	0	0
Range	0–4	0–4	0–1	0–2	0–2
Inaccurate–incomplete					
Mean	0.0543	0	0	0.04	0.17
SD	±0.23			±0.21	±0.39
Median	0	0	0	0	0
Range	0–1	0	0	0–1	0–1
Absent					
Mean	14.6	11.9	11.1	16.3	19.3
SD	±6.2	±3.84	±4.3	±5.85	±6.88
Median	13	12	11	16	19
Range	2–31	2–21	4–22	6–29	6–31
Skew	0.723	−0.038	0.84	0.473	0.049
Kurtosis	0.235	1.782	0.949	0.175	−0.582
Main concept score (102)					
Median	62	66	67	54	43
Range	7–94	37–94	34–89	15–82	7–82
Skew	−0.723	−0.064	−0.904	−0.450	−0.023
Kurtosis	0.183	0.935	1.131	−0.029	−0.531

excluding inaccurate codes (IC/II) because, for all discourse tasks, they rarely or never occurred and did not meet test assumptions. Codes AC and AB were significantly different across age groups for Cinderella and PB&J. For AC concepts, the results are, $F(3, 88)$: Broken Window, $F = 2.489$, $p = .066$; Cinderella, $F = 12.437$, $p < .001$; PB&J, $F = 4.574$, $p = .005$. For AB concepts, the results are, $F(3, 88)$: Broken Window, $F = 2.168$, $p = .097$; Cinderella, $F = 11.779$, $p < .001$; PB&J, $F = 2.988$, $p = .035$.

Results of Tukey HSD and pairwise median tests are given in Table 6. For Cinderella, both younger age groups (20–39, 40–59) performed significantly differently from both older age groups (60–79, 80+) for AB concepts and the overall MC score. In addition, both younger age groups differed from the oldest age group for code AC, and the 40–59 age group also differed from the 60–79 age group for AC concepts. For PB&J, both younger age groups differed from the 80+ age group for AC concepts, and the youngest group differed from the oldest group for AB concepts and overall MC score. Across all

Table 5. Descriptive statistics for each main concept code for the PB&J procedure of the entire normative sample and each age group separately.

	All	20–39	40–59	60–79	80–99
Accurate–complete					
Mean	5.7	6.4	6.2	5.7	4.5
SD	±2.03	±1.56	±2.17	±2.06	±1.83
Median	6	7	6	6	5
Range	1–10	4–9	2–10	1–9	1–7
Skew	−0.327	−0.488	−0.157	−0.375	−0.643
Kurtosis	−0.314	−0.975	−0.727	−0.075	−0.669
Accurate–incomplete					
Mean	0.51	0.35	0.39	0.57	0.74
SD	±0.72	±0.57	±0.58	±0.66	±0.96
Median	0	0	0	0	0
Range	0–2	0–1	0–2	0–2	0–2
Inaccurate–complete					
Mean	0.11	0.09	0.13	0.04	0.17
SD	±0.41	±0.29	±0.63	±0.21	±0.39
Median	0	0	0	0	0
Range	0–1	0–1	0–1	0–1	0–1
Inaccurate–incomplete					
Mean	0.02	0	0	0.04	0.04
SD	±0.15			±0.21	±0.21
Median	0	0	0	0	0
Range	0–1	0	0	0–1	0–1
Absent					
Mean	3.68	3.2	3.3	3.7	4.6
SD	±1.8	±1.5	±1.82	±1.8	±1.83
Median	3	3	3	3	4
Range	0–8	1–6	0–6	1–8	1–8
Skew	0.348	0.566	−0.001	0.606	0.178
Kurtosis	−0.488	−0.716	−0.917	−0.008	−0.665
Main concept score (30)					
Median	18	21	20	18	17
Range	5–30	12–27	10–30	5–27	6–23
Skew	−0.344	−0.593	−0.024	−0.484	−0.523
Kurtosis	−0.399	−0.853	−0.825	0.02	−0.972

discourse tasks and all MC codes and scores, no significant differences were observed between the two youngest groups, or between the two oldest groups.

Discussion

MCA allows for direct and comprehensive measurement of the basic building blocks used for sharing knowledge. Using this information to educate communication partners and to identify treatment targets can positively impact the ways in which persons with communication impairment interact with the world. This study is the first to generate MC checklists from a large control sample and is an important step towards developing additional clinician-friendly tools to be included in a multi-tiered discourse analysis approach. Normative information for the production of MCs is provided for select discourse tasks from the AphasiaBank protocol. The sample distribution is evaluated

Table 6. Analysis of variance (ANOVA) results for between age groups comparisons of accurate-complete and absent MC codes and overall MC scores.

One-way ANOVA $F(3, 88)$ and omnibus median test $\chi^2(3, N = 92)$			Post hoc (Tukey HSD and pairwise median tests)					
			20–39 vs. 40–59	20–39 vs. 60–79	20–39 vs. 80+	40–59 vs. 60–79	40–59 vs. 80+	60–79 vs. 80+
Cinderella	AC	12.437***			***	**	***	
	AB	11.779***		*	***	**	***	
	MC score	24.142***		**	***	***	***	
PB&J	AC	4.574**			**		*	
	AB	2.988*			*			
	MC score	8.830*			**			

Note: AC = accurate–complete; AB = absent; MC = main concept.*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

relative to the normal probability distribution and the sample composition is described, enabling readers to determine the adequacy of normative characteristics of the sample and the fit between their patient/client and the normative sample. This study provides normative information for broad age stratifications, and the mixed results (where differences in concept production across ages were observed for only two tasks) provide further motivation for investigating age-related differences in discourse.

Checklist development

There are proposition checklists currently in existence for Broken Window (Menn et al., 1998) and Cinderella (Stark, 2010). Menn et al. (1998) presented 11 events for Broken Window, developed from a small mixed sample of control and aphasic speakers. Control judges generated 41 propositions considered to be crucial for the telling of the Cinderella story (Stark, 2010). We established MC lists by identifying every RC (or candidate MC) produced by a large sample of control speakers. We applied a 33% cut-off threshold, determined through visual inspection of the frequency of RC production across the speaker sample, to arrive at our final MC lists. As expected, more RCs were produced for Cinderella than the other discourse tasks; correspondingly, the Cinderella storytelling included a greater number of MCs than the other tasks.

A cursory comparison of the lists developed in this study and existing proposition lists highlights the importance of sampling discourse from a large group of control speakers rather than small or biased samples. For example, some propositions listed by Stark (2010), particularly those elements of the story where Cinderella is not present, were never produced by our control sample (e.g., “the guests eat,” “the prince greets all the guests”) or were produced by only a single speaker (e.g., “the guests dance”). Conversely, several MCs on our checklist did not appear on Stark’s list (e.g., “the family received the invitation,” “the women are excited about the ball,” “the stepsisters tear Cinderella’s dress,” and “the prince falls in love with Cinderella.”). Using lists created by smaller and/or atypical samples of speakers or judges might lead to inaccurate estimation of the normality of discourse abilities.

MC scoring

For this project, we utilised the reliable MCA approach to measure speakers’ communication adequacy during semi-spontaneous discourse production (Boyle, 2014; Kong, 2009; Nicholas

& Brookshire, 1993, 1995). We used our enhanced checklists and adhered to Nicholas and Brookshire's (1995) detailed appendices for coding concepts as accurate/inaccurate and complete/incomplete, introducing also a modification to Kong's (2009) scoring approach to arrive at an overall MC score. We presented information about control speaker performance for each MC code and the overall MC score. Despite expecting most control speakers to either produce the concept correctly or not at all, we elected to utilise the multilevel coding system, as opposed to a binary system, in order to generate a normative reference for all codes. Readers interested in using these lists and normative information with a binary coding system would simply rely upon values reported for code AC. Intra- and inter-rater reliability using the multilevel coding system was consistent with previous reports using control subjects (94–97%, Nicholas & Brookshire, 1993; 94%, 1995).

Age-related similarities and differences in MC production

Some age-related differences in MC production were observed for two discourse tasks, with the younger half of the sample generally differing from the older half of the sample. No significant age differences were found for Broken Window. The two younger groups did not differ significantly from each other, nor did the two older groups for any of the discourse tasks. For comparisons that revealed significant age differences, the older group had higher rates of AB codes, fewer AC, and lower MC scores compared to the younger group. These results are consistent with previous research reporting higher per cent CIU production (Capilouto et al., 2005) and ME production (Wright et al., 2005, 2011; but not Capilouto et al., 2005) in younger, as compared to older, adults and support the need for further development of age-stratified norm-references for MCA.

Though stroke and resultant aphasia is generally believed to be something that occurs in older populations, stroke in younger populations has steadily increased in the past two decades (Hankey, 2013; Kissela et al., 2012; Singhal et al., 2013). We know very little about the impact of stroke in this younger population, and cannot currently predict what proportion of these individuals will have resultant aphasia. Normative references for language abilities, including discourse, are scarce to non-existent for young strokes. This information deficit can begin to be remedied by reliance upon the large and ever-expanding AphasiaBank database (or similar databases) to develop norms for numerous discourse measures and tasks.

Normative sample

Our preliminary normative information could potentially be used for both descriptive and diagnostic purposes. Descriptive norms allow evaluators to determine how a patient or client is performing when compared to the normative sample (e.g., within or outside range; below, at, or above average) and rely on measures of central tendency from samples with acceptable normal distribution properties (Busch, Chelune, & Suchy, 2005). For codes AC, AB, and for the overall MC score, the range of control speaker performance was wide and variable and the distributions had acceptable symmetry. Subsequently, the normative information is most appropriate for descriptive purposes and potentially for documenting longitudinal change in normative position along the sample's distribution (e.g., movement from below average to average), consistent with previous research where MCA revealed improvement during spontaneous recovery periods and also in response to treatment (Hopper et al., 2002; Nicholas & Brookshire, 1993).

Diagnostic norms focus on the areas that are the least variable in their presentation in control individuals (e.g., behaviours that never or rarely occur in the reference group;

Busch et al., 2005). Consistent with previous MC research conducted by Kong (2009) and Nicholas and Brookshire (1993, 1995), speaker performance was largely invariable for codes AI, IC, and II, indicating that these types of concept productions never or rarely occur in the control sample. This normative information could therefore be used diagnostically since the presence of these codes would generally indicate abnormal or pathological behaviour (see also Brandão et al., 2009).

Limitations and future directions

Readers are encouraged to examine the normative information provided from several different angles before deciding to rely upon it for descriptive or diagnostic purposes for a specific individual. Methods for evaluating normative reports generally include asking the following questions (Mitrushina et al., 2005): (a) were administration procedures adequately described?; (b) what normative data are reported, and do they at least include mean and standard deviation?; (c) was sample size reported and is it adequate?; (d) was sample composition described?; and (e) what stratifications are there for age, gender, education, and/or other important demographic variables? The first question regarding administration procedures is critical, as reporting of normative data first requires that the method for collecting and scoring the behavioural data be standardised. The procedures for collecting the discourse samples have been described fully, as referenced in the "Introduction." Regarding the type of normative data reported, the minimum requirement was met (means and standard deviations), and we provided additional information (median, range, skewness, kurtosis, normality plot) to allow readers to assess symmetry and distribution of the normative sample.

The overall sample size included transcripts from 92 control speakers, but each age group included only 23 speakers. This is an improvement compared to previous reports but falls below preferred sample size, which would be >50 speakers per stratification (Mitrushina et al., 2005). With respect to sample composition, the following speaker characteristics were reported: age, gender, years of education, race/ethnicity, and English as native language. The normative sample was slightly gender-biased (more females than males). The average years of education for our sample is in line with recent education census data, where approximately 13% of adults in the United States do not complete high school, 31% complete high school, 26% complete some college or associate's degree, 19% complete bachelor's degree, and 11% complete master's degree or above (Snyder & Dillow, 2011). The sample is not racially/ethnically diverse and may not be an appropriate comparison group for all races and ethnicities. However, thematic content, and thus perhaps concepts, may be stable across African American and Caucasian samples at least (Olness, Ulatowska, Wertz, Thompson, & Auther, 2002). With regard to stratifications to increase the specificity of comparison groups and resultant interpretations, data were grouped by 20-year-wide age intervals, and significant differences in age groups were revealed for some age groups, for some discourse tasks. No other stratifications were applied. Taken together with previous sample composition shortcomings, recommendations for improving normative reference group stratifications for MCA would include increasing the sample size overall, increasing recruitment of more diverse speakers and more male speakers, ensuring similar distribution and variance of each age group, and producing norm-references for narrower age intervals further subdivided by gender and education. As researchers continue to participate in and contribute to shared databases such as AphasiaBank, these goals may actually be achievable.

A recent development by the AphasiaBank group is a transcription-based discourse analysis tool called EVAL (Forbes et al., 2012) with the following functionality: generation of comprehensive microstructure profile; comparison of discourse profiles collected at multiple time points; and comparison of an individual to a group (control or PWA) matched by age and/or gender. MCA using the MC checklists developed here could be used in conjunction with the EVAL program to fully characterise micro-structural and concept-level performance of PWAs, to compare PWA performance to normative reference groups within the AphasiaBank database, and to potentially chart change throughout recovery and rehabilitation. As emphasised by Wright et al. (2005), and reiterated by Kong (2009), MCA does not capture whether speakers are able to express the relationships between the essential concepts or whether the gist was communicated in a cohesive manner; still, the completeness of each individual concept and the completeness of the MC checklist strongly influences discourse coherence. Although MCA has already been shown to be a good predictor of listener perceptions of socially relevant change (Ross & Wertz, 1999), the addition of a macrostructural measure would expand the MCA approach to have multiple dimensions and allow for assessment of the overall organisation of the story or procedure. This addition, plus perhaps an efficiency measure (e.g., MC/min; Kong, 2009), would be especially important if extending these checklists and norms to other populations with discourse impairment, such as individuals with traumatic brain injury, who may be able to produce key concepts, but not in a cohesive manner (Coelho, Liles, & Duffy, 1991; Hartley & Jensen, 1991; Mentis & Prutting, 1987).

When developing measures and determining which collection of measures are optimal for predicting real-world discourse and conversational abilities, barriers to discourse assessment and clinician needs must be considered if implementation of such measures in the near future is desired. MCA is ideal for use as a non-transcription-based tool and has so far proven to be clinician-friendly. Raters and clinicians have commented that the approach is easy, useful, efficient, and ideal for completing with orthographic transcription or by audio-video recording (Armstrong, Brady, Mackenzie, & Norrie, 2007; Kong, 2009).

Conclusions

MC checklists for three standardised discourse tasks were developed from a large control sample. Normative information for MC codes and overall MC score is presented for the overall sample as well as for four age groups. Those interested in utilising these checklists should rely upon Appendices 1–3 to identify essential elements, as well as Appendix A in Nicholas and Brookshire (1995) to determine the accuracy and completeness of productions. The normative sample composition and distribution should be thoroughly evaluated in relation to the patient or client so that interpretations can be made and/or qualified appropriately. Given differences in age groups, reliance upon the age-stratified normative values is recommended.

Development of MC checklists for other semi-spontaneous AphasiaBank discourse tasks (Umbrella and Cat Rescue) is underway. Refinement of the checklists and MCA approach will continue, as will investigations of multidimensional discourse assessment using these checklists in combination with measures of efficiency, cohesion, and coherence. It is hoped that norm-references for all discourse measures and AphasiaBank discourse tasks will eventually be located in a centralised and accessible repository to facilitate implementation of discourse assessment, and likely discourse treatment, in both research laboratories and clinical settings.

Acknowledgements

We would like to thank the Chapman Foundation for financial support, AphasiaBank developers and contributors, and graduate assistants Sarah Ledbetter and Nicole Williams for their contribution to this project. We would also like to thank the editor and reviewers for their valuable comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Chapman Foundation.

Supplemental data

Supplemental data for this article can be accessed at http://dx.doi.org/10.1080/02687038.2015.1057891

References

Andreetta, S., Cantagallo, A., & Marini, A. (2012). Narrative discourse in anomic aphasia. *Neuropsychologia*, *50*, 1787–1793. doi:10.1016/j.neuropsychologia.2012.04.003

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, *14*, 875–892. doi:10.1080/02687030050127685

Armstrong, L., Brady, M. C., Mackenzie, C., & Norrie, J. (2007). Transcription–less analysis of aphasic discourse: A clinician's dream or a possibility? *Aphasiology*, *21*, 355–374. doi:10.1080/02687030600911310

Boyle, M. (2011). Discourse treatment for word retrieval impairment in aphasia: The story so far. *Aphasiology*, *25*, 1308–1326. doi:10.1080/02687038.2011.596185

Boyle, M. (2014). Test-retest stability of word retrieval in aphasic discourse. *Journal of Speech, Language, and Hearing Research*, *57*, 966–978. doi:10.1044/2014_JSLHR-L-13-0171

Brandão, L., Castelló, F. G., Van Dijk, T. A., Pimenta, M. A., & Peña-Casanova, J. (2009). Cognition and discourse production in Alzheimer's disease: Using informative prompts. *Psychology & Neuroscience*, *2*, 147–155. doi:10.3922/j.psns.2009.2.006

Busch, R. M., Chelune, G. J., & Suchy, Y. (2005). Using norms in neuropsychological assessment of the elderly. In D. K. Attix & K. A. Velsh-Bohmer (Eds.), *Geriatric Neuropsychology: Assessment and Intervention* (pp. 133–157). New York, NY: Guilford Press.

Capilouto, G., Wright, H. H., & Wagovich, S. A. (2005). CIU and main event analyses of the structured discourse of older and younger adults. *Journal of Communication Disorders*, *38*, 431–444. doi:10.1016/j.jcomdis.2005.03.005

Capilouto, G. J., Wright, H. H., & Wagovich, S. A. (2006). Reliability of main event measurement in the discourse of individuals with aphasia. *Aphasiology*, *20*, 205–216. doi:10.1080/02687030500473122

Carey, S. (1988). Conceptual differences between children and adults. *Mind & Language*, *3*, 167–181. doi:10.1111/mila.1988.3.issue-3

Chandler, J. D. (1901). *The Boston cooking school magazine of culinary science and domestic economics*. Boston, MS: Boston Cooking School.

Code, C., & Petheram, B. (2011). Delivering for aphasia. *International Journal of Speech-Language Pathology*, *13*, 3–10. doi:10.3109/17549507.2010.520090

Coelho, C. A., Liles, B. Z., & Duffy, R. J. (1991). Discourse analyses with closed head injured adults: Evidence for differing patterns of deficits. *Archives of Physical Medicine and Rehabilitation*, *72*, 465–468.

Cupit, J., Rochon, E., Leonard, C., & Laird, L. (2010). Social validation as a measure of improvement after aphasia treatment: Its usefulness and influencing factors. *Aphasiology*, *24*, 1486–1500. doi:10.1080/02687031003615235

Demır, S. Ö., Altinok, N., Aydin, G., & Köseoğlu, F. (2006). Functional and cognitive progress in aphasic patients with traumatic brain injury during post-acute phase. *Brain Injury*, *20*, 1383–1390. doi:10.1080/02699050601081844

Dimbleby, R., & Burton, G. (1998). *More than words: An introduction to communication*. New York, NY: Routledge.

Doyle, P. J., Goda, A. J., & Spencer, K. A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech Language Pathology*, *4*, 130–134. doi:10.1044/1058-0360.0404.130

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299. doi:10.1037/1082-989X.4.3.272

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, *25*, 1414–1430. doi:10.1080/02687038.2011.603898

Fergadiotis, G., Wright, H. H., & Capilouto, G. (2011). Productive vocabulary across discourse types. *Aphasiology*, *25*, 1261–1278. doi:10.1080/02687038.2011.606974

Forbes, M. M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: A resource for clinicians. *Seminars in Speech and Language*, *30*, 217–212. doi:10.1055/s-0032-1320041

Frederiksen, C. H., Bracewell, R. J., Breuleux, A., & Renaud, A. (1990). The cognitive representation and processing of discourse: Function and dysfunction. In Y. Joannette (Ed.), *Discourse ability and brain damage* (pp. 69–110). New York, NY: Springer-Verlag.

Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Blaha, M. J., … Turner, M. B. (2014). Heart disease and stroke statistics – 2014 update: A report from the American Heart Association. *Circulation*, *129*, e28–e292. doi:10.1161/01.cir.0000441139.02102.80

Grimes, N. (2005). *Walt Disney's Cinderella*. New York, NY: Random House.

Hankey, G. J. (2013). Stroke in young adults: Implications of the long-term prognosis. *Journal of the American Medical Association*, *309*, 1171–1172. doi:10.1001/jama.2013.2319

Hartley, L. L., & Jensen, P. J. (1991). Narrative and procedural discourse after closed head injury. *Brain Injury*, *5*, 267–285. doi:10.3109/02699059109008097

Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology*, *22*, 184–203. doi:10.1080/02687030701262613

Hilari, K. (2011). The impact of stroke: Are people with aphasia different to those without? *Disability and Rehabilitation*, *33*, 211–218. doi:10.3109/09638288.2010.508829

Hilari, K., Needle, J. J., & Harrison, K. L. (2012). What are the important factors in health-related quality of life for people with aphasia? A systematic review. *Archives of Physical Medicine and Rehabilitation*, *93*, SS86–SS95. doi:10.1016/j.apmr.2011.05.028

Hopkins, K. D., & Weeks, D. L. (1990). Tests for normality and measures of skewness and kurtosis: Their place in research reporting. *Educational and Psychological Measurement*, *50*, 717–729. doi:10.1177/0013164490504001

Hopper, T., Holland, A., & Rewega, M. (2002). Conversational coaching: Treatment outcomes and future directions. *Aphasiology*, *16*, 745–761. doi:10.1080/02687030244000059

Kauhanen, M.-L., Korpelainen, J. T., Hiltunen, P., Maatta, R., Mononen, H., Brusin, E., … Myllyla, V. V. (2000). Aphasia, depression, and non-verbal cognitive impairment in ischaemic stroke. *Cerebrovascular Diseases*, *10*, 455–461. doi:10.1159/000016107

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York, NY: Oxford University Press.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). New York, NY: Pearson.

Kim, H.-Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution using skewness and kurtosis. *Restorative Dentistry and Endodontics*, *38*, 52–54. doi:10.5395/rde.2013.38.1.52

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394. doi:10.1037/0033-295X.85.5.363

Kissela, B. M., Khoury, J. C., Alwell, K., Moomaw, C. J., Woo, D., Adeoye, O., … Kleindorfer, D. O. (2012). Age at stroke: Temporal trends in stroke incidence in a large, biracial population. *Neurology*, *79*, 1781–1787. doi:10.1212/WNL.0b013e318270401d

Kong, A. (2009). The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: A preliminary report. *Journal of Communication Disorders*, *42*, 442–464. doi:10.1016/j.jcomdis.2009.06.002

Kong, A. (2011). The main concept analysis in Cantonese aphasic oral discourse: External validation and monitoring chronic aphasia. *Journal of Speech, Language, and Hearing Research*, *54*, 148–159. doi:10.1044/1092-4388(2010/09-0240)

Lau, M. (2013, June). Who made that? *New York Times Magazine*.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, *25*, 1286–1307. doi:10.1080/02687038.2011.589893

MacWhinney, B., Fromm, D., Holland, A., Forbes, M., & Wright, H. (2010). Automated analysis of the Cinderella story. *Aphasiology*, *24*, 856–868. doi:10.1080/02687030903452632

Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology*, *17*, 481–497. doi:10.1080/02687030344000148

Menn, L., Reilly, K. F., Hayashi, M., Kamio, A., Fujita, I., & Sasanuma, S. (1998). The interaction of preserved pragmatics and impaired syntax in Japanese and english aphasic speech. *Brain and Language*, *61*, 183–225. doi:10.1006/brln.1997.1838

Mentis, M., & Prutting, C. A. (1987). Cohesion in the discourse of normal and head-injured adults. *Journal of Speech, Language, and Hearing Research*, *30*, 88–98. doi:10.1044/jshr.3001.88

Mitrushina, M., Boone, K. B., Razani, J., & D'Elia, L. F. (Eds). (2005). *Handbook of normative data for neuropsychological assessment*. Oxford: Oxford University Press.

Nicholas, L. E., & Brookshire, R. H. (1993). A system for scoring main concepts in the discourse of non-brain-damaged and aphasic speakers. *Clinical Aphasiology Conference*, *21*, 87–99.

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech and Hearing Research*, *38*, 145–156. doi:10.1044/jshr.3801.145

Olness, G. S., Ulatowska, H. K., Wertz, R. T., Thompson, J. L., & Auther, L. L. (2002). Discourse elicitation with pictorial stimuli in African Americans and Caucasians with and without aphasia. *Aphasiology*, *16*, 623–633. doi:10.1080/02687030244000095

Prins, R., & Bastiaanse, R. (2004). Review: Analysing the spontaneous speech of aphasic speakers. *Aphasiology*, *18*, 1075–1091. doi:10.1080/02687030444000534

Ross, K. B., & Wertz, R. T. (1999). Comparison of impairment and disability measures for assessing severity of, and improvement in, aphasia. *Aphasiology*, *13*, 113–124. doi:10.1080/026870399402235

Safaz, I., Alaca, R., Yasar, E., Tok, F., & Yilmaz, B. (2008). Medical complications, physical function and communication skills in patients with traumatic brain injury: A single centre 5-year experience. *Brain Injury*, *22*, 733–739. doi:10.1080/02699050802304714

Singhal, A. B., Biller, J., Elkind, M. S., Fullerton, H. J., Jauch, E. C., Kittner, S. J., ... Levine, S. R. (2013). Recognition and management of stroke in young adults and adolescents. *Neurology*, *81*, 1089–1097. doi:10.1212/WNL.0b013e3182a4a451

Snyder, T. D., & Dillow, S. A. (2011). *Digest of Education Statistics, 2010* (NCES 2011-015). Washington, DC: National Center for Education Statistics.

Spaccavento, S., Craca, A., Del Prete, M., Falcone, R., Colucci, A., Di Palma, A., & Loverre, A. (2014). Quality of life measurement and outcome in aphasia. *Neuropsychiatric Disease and Treatment*, *10*, 27–37. doi:10.2147/NDT.S52357

Stark, J. A. (2010). Content analysis of the fairy tale Cinderella–A longitudinal single-case study of narrative production: "From rags to riches". *Aphasiology*, *24*, 709–724. doi:10.1080/02687030903524729

Sullivan, M. (2003). *Statistics: Informed decisions using data* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language*, *13*, 345–371. doi:10.1016/0093-934X(81)90100-0

Ulatowska, H. K., Olness, G. S., Wertz, R. T., Samson, A. M., Keebler, M. W., & Goins, K. E. (2003). Relationship between discourse and Western Aphasia Battery performance in African Americans with aphasia. *Aphasiology*, *17*, 511–521. doi:10.1080/0268703034400102

Van Dijk, T. A. (2006). Discourse, context and cognition. *Discourse Studies*, *8*, 159–177. doi:10.1177/1461445606059565

Van Dijk, T. A. (2014). *Discourse and knowledge: A sociocognitive approach*. Cambridge: Cambridge University Press.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56–75). Newbury Park, CA: Sage.

Wright, H. H., Capilouto, G. J., Srinivasan, C., & Fergadiotis, G. (2011). Story processing ability in cognitively healthy younger and older adults. *Journal of Speech Language and Hearing Research*, *54*, 900–917. doi:10.1044/1092-4388(2010/09-0253)

Wright, H. H., Capilouto, G. J., Wagovich, S. A., Cranfill, T. B., & Davis, J. E. (2005). Development and reliability of a quantitative measure of adults' narratives. *Aphasiology*, *19*, 263–273. doi:10.1080/02687030444000732

Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders*, *45*, 27–36. doi:10.1044/jshd.4501.27

Appendix 1. Main concepts for Broken Window

Essential information is italicised and bolded. Each essential segment is numbered (superscript) with alternative productions (if any were produced) listed by number below. These alternative productions are not intended to be an exhaustive list, but represent some of the more common productions of the normative sample and are included to aid in scoring. Additional, but non-essential, information often spoken to complete the main concept is in normal font.

†

1) ***[1]A/The boy [2]was [3]outside***.
   1. *"He" since referent is unambiguous; some give the boy a name*
   2. *Is, decided to go*
   3. *In his front yard, on the lawn, out of the house, etc.*

   Note: Sometimes, this concept was combined with number 2 in a statement like "The boy was playing soccer outside" or "The boy was kicking the ball in the yard". These statements would receive full credit for both concepts 1 and 2.

††

2) ***[1]A/The boy [2]was playing [3]soccer.***
   1. *See 1.1*
   2. *Played, is kicking, kicks, is practicing, etc.*
   3. *With the soccer ball, with the ball, with the football* (*if not from U.S.)*

   Note: "He has a ball" or "He has a soccer ball" did not count towards this concept because it does not imply any kind of action with the soccer ball, and boy–action–ball was the concept that met criterion.

††

3) ***[1]The ball [2]breaks [3]the*** man's/neighbor's ***window.***
   1. *Soccer ball, football**
   2. *Goes through, went through, crashes through/into, flew through, sails through/into, shattered, is kicked through*
   3. *glass*

†

4) ***[1]The man [2]is sitting*** in a chair and/or inside the house.
   1. *His dad, his father, the father, the neighbor, the guy; some give the man a name*
   2. *Lounging, resting, relaxing, inside*

   Note: Most common were "The man is sitting," "The man is inside," "The man is sitting inside."

   Note: "The man is watching TV" or something similar did not count for this concept; that was a separate relevant concept that did not meet criterion. However, if an individual said, "The man is sitting watching TV," then they would receive credit for this concept since they included "sitting."

5) ***[1]The man [2]was startled.***
   1. *See 4.1*
   2. *Surprised, amazed, afraid, astonished, freaked out, stunned, shocked, angry, upset, not happy, mad*

   Note: Occasionally, this concept was combined with number 3, in a statement such as "The ball crashed through the window and startled the man."

†

6) ***[1]The ball [2]broke [3]a lamp.***
   1. *See 3.1*
   2. *Knocks down/over, smashes into, breaks, hit*
   3. *No alternative for lamp was produced*

††
7) [1]***The man*** [2]***picked up*** [3]***the ball.***
1. *See 4.1*
2. *Grabs, gets, holds/is holding, catches, captures, has*
3. *See 3.1*
Note: Occasionally, "The man stands up with the ball" and "The man jumps up with the ball" was used to express this concept, expressing that the man had performed some action to hold on to the ball.

††
8) [1]***The man*** [2]***looked*** [3]***out of the window.***
1. *See 4.1*
2. *Looks, is looking*
3. *Outside, out, out of the glass*
Note: "The man goes to the window," "The man went to the window," or "The man goes outside", etc., did not count towards this concept. These were separate relevant concepts that did not meet criterion.

† Indicates concepts produced by 50% of the normative sample.
†† Indicates concepts produced by 66% of the normative sample.

Appendix 2. Main concepts for Cinderella

Essential information is italicised and bolded. Each essential segment is numbered (superscript), with alternative productions (if any were produced) listed by number below. These alternative productions are not intended to be an exhaustive list, but represent some of the more common productions of the normative sample and are included to aid in scoring. Additional, but non-essential, information often spoken to complete the main concept is in normal font.

†
1) [1]***Dad*** [2]***remarried*** [3]***a woman*** with two daughters.
1. *Daddy/Father*
2. *Got married to, got remarried, married again*
3. *A lady*

2) [1]***Cinderella*** [2]***lives with*** [3]***stepmother/stepsisters.***
1. *She**
2. *Is left with, moves in with, grows up with, has*
3. *Stepfamily, new family, the women, they*
i. *If they do not mention the word "step", there must be a clear indication that the stepmother and stepsisters (the lady and her two daughters, the mean woman and her beautiful daughters) are a unit separate from Cinderella.*
Note 1: After Cinderella has been introduced into the story, "she" is an acceptable alternative as long as there is a clear pronoun referent.
Note 2: After the stepmother and stepsisters have been introduced into the story, "they" is an acceptable alternative as long as there is a clear pronoun referent.

†
3) [1]***Stepmother/stepsisters*** [2]***were mean*** [3]***to Cinderella.***
1. *See 2.3*
2. *Were cruel, were wicked, treated Cinderella poorly, were awful, hated*

4) [1]***Cinderella*** [2]***was*** [3]***a servant*** to the stepmother and stepsisters.
1. *She*
2. *Was forced to be, had to be*
3. *Maid, slave, domestic*
Note: If they say the sentence in another way that expresses servitude, for example "had to wait on," they must include stepmother and/or stepsisters, because the verb requires an object. This would be the only time they are essential for this concept.

†
5) [1]***Cinderella*** [2]***has to do*** [3]***the housework.***
1. *She*
2. *Is forced to do, must do, has to take care of*
3. *Chores, cleaning, taking care of the house, everything*

6) The king thinks ***[1]the prince [2]should get married.***

1. *He**
2. *Needs to get married/find a wife, must get married, has to get married*

Note: After the prince has been introduced into the story, "he" is an acceptable alternative as long as there is a clear pronoun referent.

†

7) King announces ***[1]there is going to be [2]a ball*** in honor of son who needs to find a wife.

1. *Will be, is to be, is*
2. *Dance, big party, celebration, gala*

Note: Occasionally this concept was combined with number 8 in a statement like, "They got an invitation to the ball the king was hosting for his son." This should receive full credit for concepts 7 and 8.

8) ***[1]They [2]got [3]an invitation*** *to the ball*.

1. *The women, the stepmothers and/or stepsisters and/or Cinderella, everyone in the household, the household*
2. *Received, was delivered (if word order altered so that the invitation is delivered to the women)*
3. *No alternatives were produced for "invitation"*

Note 1: *Not essential if clear from context or previously stated; otherwise, see note for number 7.
Note 2: Alternatively, the speaker could say "They were invited to the ball" or something similar.

9) ***[1]They [2]are excited*** *about the ball*.

1. *See 8.1*
2. *Are happy, are pleased*

Note 1: *See number 8.
Note 2: If they say something like "They are looking forward to," they must include "the ball" because an object is required.

†

10) ***[1]Cinderella*** is told by the stepmother she ***[2]cannot go*** *to the ball* unless/because (insert reason).

1. *She*
2. *Could not go, has to stay home, is not allowed to go*

Note 1: *See number 8.
Note 2: An alternative is "If Cinderella could get all of her chores done, she could go to the ball."

11) ***[1]The stepsisters [2]tore [3]Cinderella's dress.***

1. *They*
2. *Ruined, destroyed, ripped up, shredded*
3. *Her dress*

12) ***[1]Stepmother/stepsisters [2]went*** *to the ball*.

1. *Everyone but Cinderella*
   i. *If "They went to the ball" is the sentence, the "they" must clearly exclude Cinderella in the context*
2. *Go, left, departed*

Note: *See number 8.

13) ***[1]Cinderella [2]was [3]upset.***

1. *She*
2. *Is*
3. *Crying, sad, disappointed*

††

14) ***[1]A fairy godmother [2]appeared*** to Cinderella.

1. *No alternative for "fairy godmother" was produced*
2. *Shows up, appears, surprises, comes*
   i. *Some may say "Cinderella sees" or "meets" or "finds", in which case Cinderella then becomes an essential element*

Note: Another popular way of expressing this is "Along came a fairy godmother" (which is basically "appeared a fairy godmother").

†

15) ***[1]The fairy godmother [2]makes [3]{item(s)} turn into {items}.***

1. *See 14.1*
2. *Turns, creates, changes, any other verb indicating transformation/creation*
   i. *Must be a verb that indicates some kind of transformation or creation*

3. *Pumpkin and mice OR carriage/coach and horses*
   - i. *When producing this concept, only one pair needs to be mentioned, however, it must be correctly paired to receive full credit*
     - a. *Pumpkin → carriage/coach (and horses)*
     - b. *Mice → horses (and carriage)*
     - c. *If they initially mention both pumpkin and mice, they do not necessarily have to mention both after the transformation occurs in order to receive full credit, and only one needs to be accurate*
     - d. *Do not take points off for mentioning other transformations, such as dog → coachman as these are not incorrect, they simply did not reach significance.*

††

16) ***[1]The fairy godmother [2]makes [3]Cinderella [4]into a beautiful princess.***
1. *See 14.1*
2. *Turns, creates, changes, gives*
3. *The regular girl, her regular clothes*
4. *Dress/shoes into gown/slippers, beautiful*

††

17) ***[1]Cinderella [2]went [3]to the ball*** in the coach.
1. *She*
2. *Goes, arrives, reaches*
3. *See 7.2*

††

18) She knew ***[1]she [2]had to be [3]home by midnight*** because everything will turn back at midnight.
1. *Cinderella*
2. *Must be, needs to be, must return*
3. *Leave by midnight*

Note: An alternative could be "The fairy godmother told her that if she wasn't home by midnight, XXX would happen" or something similar.

††

19) ***[1]The prince and Cinderella [2]danced*** around the room/all night/with no one else.
1. *They*
2. *Were dancing, kept dancing*

†

20) ***[1]Prince [2]falls in love [3]with Cinderella.***
1. *He*
2. *Is enamored with, is delighted with, is awestruck by, likes, is hooked on*
3. *Her*

Note: If someone says "Prince/They fall in love at first sight" that individual can receive credit only if Cinderella has been mentioned before or it is clearly indicated who "they" are.

†

21) Cinderella realized ***[1]it [2]is [3]midnight.***
1. *Clock, something indicates that it is*
2. *Is, gets to be, rings, strikes*
3. *Twelve o'clock, twelve midnight, almost midnight*

††

22) ***[1]She [2]ran [3]down the stairs***.
1. *See 18.1*
2. *Was running, flew, rushed, sprinted, left, was leaving*
3. *Out of the ball/castle, away from the ball/castle/prince, out, away*

††

23) As she was running down the stairs ***[1]she [2]lost one of the [3]glass slippers.***
1. *See 18.1*
2. *Leaves, steps out of*
3. *Shoes, glass shoe*

24) ***[1]Prince [2]finds [3]Cinderella's shoe.***
1. *Any other royal figure, king, servant, duke, prime minister, chamberlain*
2. *Had, got, retrieved, was brought*
3. *See 23.3*

Note: An alternative way to say this is "The servant brings the slipper to the prince."

25) ***[1]Everything [2]turns back [3]to its original form.***
1. *Pumpkin, mice, and/or clothes/dress*
2. *Goes back, returns, disappears*
i. *If the speaker uses "disappears" they do not have to specify what disappears, for example, "Everything disappears."*
3. *To normal (can specify what it turns back into)*
Note: The addition of "again" at the end of the sentence paired with a verb that does not indicate change is acceptable because it implies a return to the original state (e.g., "she got home and the dress was old again").

26) ***[1]She [2]returned [3]home*** in time.
1. *See 18.1*
2. *Gets, makes it, goes*
3. *to the house*

††
27) ***[1]The prince [2]searched*** door to door ***[3]for Cinderella.***
1. *See 24.1, the servant*
2. *Was trying to find, looked for*
3. *For the person who would fit into the glass slipper, for the girl from the ball*
Note: Alternatively, this could be stated as "The prince/his servant was trying the slipper on all the girls."

28) ***[1]Prince [2]comes [3]to Cinderella's house.***
1. *See 27.1*
2. *Arrives at, went, found, shows up at*
3. *Her*

29) ***[1]The stepsisters [2]try on [3]the glass slipper.***
1. *The stepsisters and stepmother, the other girls (if the reference to the stepsisters is clear), etc.*
i. *For this concept, speakers may include the stepmother. However, the stepmother alone is not sufficient.*
2. *Attempt to put on, cram, try to fit into, try their foot in*
3. *See 23.3*

†
30) ***[1]The slipper [2]didn't fit [3]the stepsisters.***
1. *See 23.3*
2. *Would not go on the feet of, did not work for, couldn't fit on*
3. *See 29.1*

31) ***[1]He [2]put [3]the slipper on*** Cinderella's foot.
1. *See 27.1*
2. *Tried, slid, slipped, placed*
3. *See 23.3*
Note 1: Though most did not specifically mention Cinderella in these concepts (31, 32), it was clear that at this point in the story they were referring to Cinderella.
Note 2: An alternate way to say this is "Cinderella tried on the slipper."

††
32) ***[1]The slipper [2]fits*** Cinderella perfectly.
1. *See 23.3*
2. *Belonged to*
Note: An alternate way to say this is "The shoe slid easily onto her foot" or "The slipper was Cinderella's."

33) ***[1]Cinderella and the prince [2]were married.***
1. *See 19.1*
2. *Got married, were wed, had a wedding, had a marriage celebration*
Note: An alternate way to say this is, "The prince took Cinderella as his bride."

††
34) ***[1]Cinderella and the prince [2]lived happily ever after.***
1. *See 19.1*
i. *The speaker must include or refer to both Cinderella and the prince in order to receive an accurate and complete score*
2. *Lived forever, lived a long time, were happy for life.*

i. The speaker must indicate an extended length of time in order to receive an accurate and complete score. For example, "ever after," "forever," "a long time," "life"
Note: Variations can include "They lived happily every after," "They were together forever," and "They had a wonderful life."

† Indicates concepts produced by 50% of the normative sample.
†† Indicates concepts produced by 66% of the normative sample.

Appendix 3. Main concepts for PB&J

Essential information is italicised and bolded. Each essential segment is numbered (superscript) with alternative productions (if any were produced) listed by number below. These alternative productions are not intended to be an exhaustive list, but represent some of the more common productions of the normative sample and are included to aid in scoring. Additional, but non-essential, information often spoken to complete the main concept is in normal font.

1) ***[1]Get [2]bread [3]out*** of the pantry/cupboard/refrigerator/freezer/etc.
   1. *Take out, remove, grab, find, pull out*
   2. *Loaf, bread loaf, bread bag*
   3. *From (which if used, must be followed by a location from the speaker)*

††
2) ***[1]Get [2]two slices [3]of bread.***
   1. *See 1.1*
   2. *A couple of slices, two pieces*
   i. Most speakers indicated more than one piece of bread. However, if the speaker uses one piece of bread throughout the story, and for #8 uses "fold" or "close", or indicates in some way that they made a half sandwich, then one slice/piece of bread is allowed. Speakers must be consistent throughout the telling for this to be counted as correct.
   3. *If the speaker received full credit for the first concept, they do not necessarily have to repeat "of bread" to be counted correct/complete for this concept. For example a speaker could say "You take the bread out of the pantry and get two slices."*

†
3) ***[1]Get [2]the peanut butter.***
   1. *See 1.1*
   2. *A jar of peanut butter*
   Note: A concept like "take off the lid on the peanut butter" or "open the peanut butter" cannot be used for this concept. This was a relevant concept that did not reach significance.
4) ***[1]Get [2]the jelly.***
   1. *See 1.1*
   2. *A jar of jelly, jam, preserves, honey*
   Note: See note for number 3.
5) ***[1]Get [2]a knife.***
   1. *See 1.1*
   2. *Spatula*
6) ***[1]Put [2]the bread [3]on the plate.***
   1. *Place, set, lay*
   2. *The slices, the pieces, it, them*
   i. *It must be clear that the individual is referring only to the bread, not the jelly, peanut butter, or knife*
   3. *Counter, breadboard, cutting board, napkin, down (on a surface)*

††
7) ***[1]Put [2]peanut butter [3]on bread.***
   1. *Place, smear, spread, slap, slather, spoon out, cover*
   2. *No alternative for "peanut butter was produced*
   3. *On top of jelly, on the other piece of bread, on one slice, on one side, on one half*

††
8) ***[1]Put [2]jelly [3]on bread.***
   1. *See 7.1*
   2. *See 4.2*
   3. *See 7.3*

††
9) ***[1]Put [2]the two pieces [3]together***.
1. *Place, smash, slap, smack, stick*
2. *The bread, the two slices of bread, the two sides, the peanut butter and jelly, the two, the two halves, them*
3. *If the speaker does not say "together" they must give some indication that the two pieces become one (i.e., "Put one piece on top of the other", "Combine the pieces of bread", "Put the second piece of bread on top.")*

i. The verbs "fold" and "close" cannot be used for this concept, unless the speaker tells the entire story with one piece of bread as if making half of a sandwich, see 2.2.i.

10) ***[1]Cut [2]the sandwich [3]in pieces.***
1. *Slice*
2. *The bread, it*
3. *In half, in quarters, in two, diagonally, across, on the bias, down the middle, with an x, however you like*

Note: For concepts 1–5, "put" is not an acceptable verb. For each of those concepts, there was a similar relevant concept (i.e., "put the bread on the counter"); however, none of these relevant concepts reached criterion. In these cases, the speaker would receive a score of absent, and any information associated with the verb "put" should be treated as extra information that is not scored.

† Indicates concepts produced by 50% of the normative sample.
†† Indicates concepts produced by 66% of the normative sample.

Appendix 4. Examples of statements that received each MC code for the three discourse tasks

Broken Window

[1]A/The boy [2]was playing [3]soccer.

AC **"boy was playing outside with a ball."**

AI **"boy ball"**
- *No verb is produced.*

IC **"they kick this around"**
- *Clear pronoun referent for "this" from previous statement.*
- *Incorrect pronoun "they."*

II **"and baseball or something"**
- *Incorrect noun use.*
- *No verb is produced.*

Cinderella

[1]She [2]ran [3]down the stairs.

AC **"she was running down the steps"**
- *Clear pronoun referent from previous statement.*

AI **"and she had to run"**
- *Clear pronoun referent from previous statement.*
- *Omitted essential element "down the stairs."*

IC **"so he gets out"**
- *Incorrect pronoun "he."*

II **"she run"**
- *No clear pronoun referent for "she" from previous statement.*
- *Omitted essential element "down the stairs."*

Peanut Butter and Jelly Sandwich

[1]Get [2]the peanut butter.

AC **"oh, well first you get the peanut butter out"**

AI **"peanut butter"**
- *Clearly indicating with gestures that this is the beginning of the procedure.*
- *No verb is produced.*

IC **"and you get out the butter"**
- *Incorrect noun use.*

II **"and then the peanuts and the jellies"**
- *No verb is produced.*
- *Incorrect noun use.*