HHS Public Access
Author manuscript
J Clin Exp Neuropsychol. Author manuscript; available in PMC 2019 November 01.

Published in final edited form as:
J Clin Exp Neuropsychol. 2018 November ; 40(9): 917–939. doi:10.1080/13803395.2018.1446513.

Connected Speech and Language in Mild Cognitive Impairment and Alzheimer's Disease: A Review of Picture Description Tasks

Kimberly D. Mueller, Ph.D.[a,b,*], **Bruce Hermann, Ph.D.**[c], **Jonilda Mecollari**[b], and **Lyn S. Turkstra, Ph.D.**[a,d]

[a]Department of Communication Sciences and Disorders, University of Wisconsin-Madison, USA

[b]Wisconsin Alzheimer's Institute, University of Wisconsin-Madison, USA

[c]Department of Neurology, University of Wisconsin-Madison, USA;

[d]School of Rehabilitation Science, McMaster University, Canada

Abstract

The neuropsychological profile of people with Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) dementia includes a history of decline in memory and other cognitive domains, including language. While language impairments have been well described in AD, language features of MCI are less well understood. A potentially sensitive measure of language in MCI is analysis of connected speech. Connected speech analysis is the study of an individual's spoken discourse, usually elicited by a target stimulus, the results of which can facilitate understanding of how language deficits typical of MCI and AD manifest in everyday communication. Among discourse genres, picture description is a constrained task that relies less on episodic memory and more on semantic knowledge and retrieval, within the cognitive demands of a communication context. Here we review picture description task data from adults with MCI and AD, identify speech and language parameters assessed, discuss the potential value of this task for MCI diagnosis, and provide recommendations for future research.

Keywords

Mild Cognitive Impairment; Alzheimer's Disease; connected speech; language; discourse; picture description; spontaneous speech; dementia

BACKGROUND

Alzheimer's Disease (AD) is a neurodegenerative disease that is most often diagnosed based on clinical symptoms, most notably gradual worsening of problems with learning and memory and other cognitive abilities that affect activities of daily living (McKhann et al., 2011; Rentz et al., 2013; Sperling et al., 2011; Sperling et al., 2010). Accumulating evidence (Albert et al., 2011; Sperling et al., 2011) suggests that the neuropathology of AD

*Address correspondence to Kimberly D. Mueller, Wisconsin Alzheimer's Institute, University of Wisconsin-Madison, 610 Walnut Street, 9th Floor, Madison, WI 53726 kdmueller@wisc.edu). *addresses of co-authors:* Bruce Hermann, Ph.D: hermann@neurology.wisc.edu, Jonilda Mecollari: mecollari@wisc.edu, Lyn S. Turkstra: turkstrl@mcmaster.ca.

begins decades before the onset of clinical symptoms. Determining the most sensitive assessment tools for detecting cognitive change is crucial not only for early identification, but also for disease monitoring in clinical trials. Considerable research has focused on assessment of memory and executive functions (Buckner, 2004; Small, Perera, DeLaPaz, Mayeux, & Stern, 1999), which are the typical hallmarks of both AD and Mild Cognitive Impairment (MCI) (Albert et al., 2011), but language assessment also may be sensitive to early cognitive change (Taler & Phillips, 2008). Analysis of connected language may be particularly revealing. Connected language analysis is the study of self-generated discourse, and has been an intriguing area of AD research for several reasons: 1) connected language involves ongoing interactions among diverse cognitive processes including semantic storage and retrieval, executive functions, and working memory, which contrasts with isolated tasks such as picture-naming; 2) connected language more closely approximates language production in everyday contexts than do standardized tests; and 3) connected language can be a quick means of assessment with relatively low burden for the participant. Most connected language research has used subjects with mild to moderate AD, but some retrospective analyses of connected language (Ahmed, Haigh, de Jager, & Garrard, 2013b; Berisha, Wang, LaCross, & Liss, 2015; Garrard, Maloney, Hodges, & Patterson, 2005) have revealed changes during the MCI phase or earlier. Subtle, preclinical changes in language have been particularly problematic to document, because memory is often the primary symptom of AD, and although language abilities and memory are necessarily intertwined, there is no standardized approach to differentiating language from memory impairments. Further, it has been difficult to differentiate between the language changes in MCI and those associated with normal aging. Thus, quantifying memory difficulties and language problems simultaneously may increase diagnostic sensitivity, as well as provide important prognostic indicators. For example, one study (De Jager, Blackwell, Budge, & Sahakian, 2005; De Jager & Budge, 2005) found that a subgroup of participants who evidenced both associative learning and naming deficits showed steeper cognitive decline over four years than those with deficits in only one of these two domains. As communication problems have been shown to be directly correlated with challenging behaviors in AD (Gitlin, Winter, Dennis, Hodgson, & Hauck, 2010; Rao, Anderson, Inui, & Frankel, 2007) as well as with increases in caregiver burden (Savundranayagam, Hummert, & Montgomery, 2005), it also is critical to identify communication difficulties early in the disease so non-pharmacological communication interventions can be implemented for both the patient and caregiver. In this paper, we provide a comprehensive review of studies of connected language in AD, to determine its diagnostic utility at the MCI or even pre-MCI phases of disease.

Language deficits in AD dementia have been well documented (Price et al., 1993; Taler & Phillips, 2008) and the language profile of adults with AD typically is characterized by “empty speech” (Nicholas, Obler, Albert, & Helm-Estabrooks, 1985), referring to word retrieval deficits that result in the use of circumlocutions, nonspecific language, and an overabundance of words conveying limited ideas. In the moderate to severe stages of disease, communication skills degrade further with deficits in both production (Kemper, Thompson, & Marquis, 2001) and comprehension of language (Bickel, Pantel, Eysenbach, & Schröder, 2000; Grossman et al., 1996; MacDonald, Almor, Henderson, Kempler, & Andersen, 2001; Martin & Fedio, 1983), reflected in communication breakdowns in everyday interactions (Savundranayagam & Orange, 2014) and increased frustration that may result in challenging behaviors (Woodward, 2013). Often the end stage of AD is characterized by a complete lack of verbal communication, and the person with AD becomes socially disengaged (Blair, Marczinski, Davis-Faroque, & Kertesz, 2007; Chung & Cummings, 2000; Frisoni, Fratiglioni, Fastbom, Viitanen, & Winblad, 1999; Hart et al., 2003; Ripich & Terrell, 1988).

Connected language analysis has long been used as a means of evaluating expressive language performance, with its roots in child language development and developmental psychology (Bloom et al., 1975; Brown, 1973; Evans et al., 1992; Botting, 2002). Connected language more closely approximates a functional communication skill than performance on isolated, single-word language tasks, and it has therefore received increased attention over the last two decadesas a means toward understanding and documenting longitudinal change in language in persons with cognitive decline, particularly AD dementia (Kemper et al., $ $1990). Because expressive language in AD dementia is characterized as “empty” or devoid of content words, measures capturing semantic content of language samples have been especially useful. For example, “idea density” or “propositional idea density (P-Density)” is a measure of semantic content words relative to the total number of words in a sentence (Kintsch & Keenan, $$1973) that has been widely examined in studies of dementia and aging. The seminal study of retrospective language analysis in AD was the Nun Study (Snowdon et al., 1996), in which researchers examined early life writing samples of cloistered nuns, and were able to predict cognitive decline based on the P-density of these early life writing samples. The authors theorized that low idea density reflected suboptimal “neurocognitive development” and thus an increased risk for cognitive decline or AD neuropathology. Kemper et al. (2001) further examined the Nuns’ writing samples and found that rates of decline in both idea density and grammatical complexity were similar for those nuns who developed dementia and those who did not, but that those who developed dementia had lower baseline levels on both measures, supporting the theory of developmental disadvantage.

.Other measures used in connected language studies capturing semantic content include the number of nouns relative to the number of pronouns plus nouns (i.e., pronouns without referents), as a means of quantifying “nonspecific language” (Almor et al., $$1999). Connected language analysis has also yielded measures of grammatical complexity such as percent verbs or verb indices (Kim & Thompson (2004)), and measures of coherence and informativeness (Chapman et al., $$1998; Laine et al., 1998; Ripich & Terrell, 1988).

There are many ways to elicit connected language samples, several of which have been explored in the dementia literature. Researchers have made use of existing language samples from individuals with known AD dementia that are a matter of public record, such as the unscripted speeches of former US President Ronald Reagan (Berisha et al., 2015). In this study, Berisha and colleagues compared longitudinal unscripted speeches (responses to questions by reporters in press conferences) to those of a “control,” President George W. Bush who had no known dementia, and found that President Reagan displayed a lower proportion of unique words and a higher proportion of filled and unfilled pauses over time. Similarly, other retrospective studies examined language in the novels of the British writer

Iris Murdoch, and found that decreases in language abilities were apparent years before her formal diagnosis of AD (Garrard et al., 2005). These approaches of using available language samples, while informative, are problematic in that they lack standardization and the ability to generalize findings to other populations of individuals with AD dementia.

Language samples have also been elicited in more structured ways, such as through open-ended questions or semi-structured interviews. For example, researchers have asked participants to describe “the happiest moment of their lives” (Armaki et al., $$2015), or to answer general questions along the lines of career, family, life and hobbies (Sajjadi et al., 2012; Graham, Patterson, & Hodges, $$2004; Orange & Peacock, $$1998; Ripich & Terrell, 1988).

Other studies of connected speech and language have used language samples obtained from more constrained tasks such as picture description. While open-ended elicitation methods may provide a larger quantity of output, they can be highly variable within and across individuals and contexts, and thus cannot be easily standardized for between- and within-group comparisons. By contrast, picture description tasks can be easily evaluated with standardized measures, and if the picture is visible throughout the task they rely less on episodic memory. Picture stimuli most frequently cited in the literature are Norman Rockwell prints such as “Easter Morning” (Tomoeda, Bayles, Trosset, Azuma, & McGeagh, 1996), and the “Cookie Theft” picture from the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1983). The “Cookie Theft” picture is the most widely used overall, and was designed to include aspects of person, time, place, and actions, and depicts key vocabulary which is acquired early in life (Giles, Patterson, & Hodges, 1996)

As the language deficits in Alzheimer’s Disease dementia are typically semantically-based, measures of informational content of language are particularly important. Picture description readily lends itself to these measures and has the added benefit of minimizing demands on episodic and autobiographical memory. While there is a large literature on picture description in AD dementia, such literature on picture description or any other aspect of language in MCI is comparatively small. However, smaller retrospective studies do suggest that declines in language may evolve in prodromal phases (Berisha et al., 2015; Garrard et al., 2005). Picture description tasks could be informative at the MCI and pre-MCI phase, not only to help inform diagnosis, but also to reveal the earliest point at which cognitive-communication intervention could be most useful. In this review, we set out to examine picture description tasks and explore the ways in which the literature reports on their utility in both MCI and AD dementia. Our aim was to answer the following research questions:

1. What measures of speech and language are assessed in picture description tasks?
2. What psychometric properties of connected speech analysis have been reported?
3. What are the limitations of the reported studies, and needs for future research?
4. What is the evidence to support using connected language analysis as a means to detect and describe how cognitive problems may manifest in everyday speech and language at the MCI stage?

METHODS

Inclusion Criteria for the Review

Study Design Criteria—We included only English-language experimentally-based peer reviewed journal articles that focused on diagnostic testing with both quantitative and qualitative data results. All studies were required to have included a picture description task such as "Cookie Theft" from the Boston Diagnostic Aphasia Examination (BDAE, Goodglass & Kaplan, 1983), the "picnic scene" from the Western Aphasia Battery (WAB; Kertesz, 1982), or Norman Rockwell prints.

Participant Criteria—Additional criteria included that study participants were diagnosed with dementia due to probable Alzheimer's disease, autopsy-confirmed Alzheimer's disease, or individuals with biomarker evidence of Alzheimer's disease neuropathology, or persons with familial AD (mutations of Presenilin I or II genes) with or without clinical symptoms. Also included were adults diagnosed with Mild Cognitive Impairment, based on Peterson et al. criteria (1999) or Albert et al. criteria (2011), or with evidence of early MCI (defined by demonstrated longitudinal decline on cognitive testing. Studies that included other types of dementia (e.g., Vascular dementia) or other neurodegenerative disorder (e.g., Parkinson's) were included provided that there was also a group with dementia due to possible or probable AD,MCI, or a biomarker-based, preclinical AD group. We included participants who spoke languages other than English, and studies were not excluded based on age or sex characteristics of participants.

Exclusion Criteria

Studies that met one or more of the following seven criteria were excluded: 1) studies published in a language other than English for which translations were not available; 2) studies not published as a peer-reviewed journal article (i.e. poster abstracts, conventions, etc.); 3) studies that did not include quantitative, linguistic or communication analyses; 4) studies that did not use picture descriptions as an assessment tool, but rather for other purposes such as intervention targets or as a means of monitoring a particular therapeutic intervention; 5) studies in which the picture description task was limited to *written* descriptions rather than verbal; 6) studies where picture description consisted of sequential action descriptions only (e.g. listing events in a sequence of pictures, picture naming, or story re-telling); 7) studies in which picture description was part of a larger language-test battery, but where scores for picture description alone were not available.

Method for Searching and Identifying Studies

We conducted an electronic literature search for articles reported up to December 2016 using the PubMed, PsychINFO and CINAHL databases. A librarian verified indices and search terms used to identify records, and suggested additional terms that were used to search WEB OF SCIENCE and SCOPUS. We randomized and divided relevant search terms into eleven different groups, each containing at least seven words or more of the relevant search terms that were entered into the database. For example, "Dementia OR Alzheimer's disease AND Connected Speech OR Spontaneous Language AND Cookie Theft OR Picture description OR Norman Rockwell Prints" was one group of search terms that fit our criteria. For randomization, we put each of the terms listed into a random name generator and examined results until it was evident to the reviewers that a cohesive group of search terms was established. Each group of terms was then entered into a database, resulting articles were collected, and duplicates were removed. We tracked the number of articles retrieved during each screening process using the PRISMA (Moher, Liberati, Tetzlaff, Altman, & and the, 2009) Flow Diagram

The review process was as follows: two authors (KDM, JM) first independently screened titles for potentially eligible studies following the inclusion criteria. Next, abstracts and/or full text versions of those articles were reviewed, and articles that did not meet the inclusion criteria were removed. The final set of articles was then scanned for additional references that were not returned in the electronic search. Any disagreements between the two reviewers were resolved by discussion.

RESULTS

Results of the search process are presented in Figure 1. The search initially returned 871 results. After exclusion criteria were applied, the final selection was 34 papers. The most common exclusions included studies that were unrelated to language analysis, or to AD dementia or MCI (n=478), Some results were excluded due to the fact that they were qualitative discourse analysis studies that examined underlying themes or ideas (e.g., perceptions of stigma in AD diagnosis), rather than linguistic analyses... We excluded studies that did not include a single picture description task (n=236), or articles that examined written picture descriptions only. Remaining articles were scanned for additional references and 2 more were added, bringing the total to 36.

Of the 36 studies included in the review, 25 studies used the “Cookie Theft” picture from the BDAE (Fig. 1), 4 of which used the Cookie Theft combined with other pictures. One study used the picture from the Dutch version of the Comprehensive Aphasia Test (Swinburn, Porter, & Howard, 2004), 5 studies used Norman Rockwell prints, and one study used the Picnic Scene from the Western Aphasia Battery (Kertesz, 1982; Fig. 2). 31 studies examined patients with AD dementia, and 5 studies included an MCI or pre-MCI group. Across all 36 studies, there was a total of 1,127 patients with AD dementia and 274 with MCI or pre-MCI. The average number of participants with AD or MCI per study was 35, with age ranges of 65.1 – 82.5 years for AD, and 43.2 – 73.5 years for MCI or pre-MCI. Participant characteristics and major study findings are summarized in Table 1.

Psychometric Properties of Connected Speech Analysis in Picture Description Tasks

The following sections outline the measures of reliability and validity that were reported in the selected 36 studies, including test-retest reliability, rater reliability, parallel form reliability, normative measures, construct validity, and discriminant validity.

Test-Retest Reliability—As part of a normative study of simple and complex picture description tasks, Forbes-McKay & Venneri (2005) examined test-retest reliability in a randomly selected group of 40 healthy adults, ages 23–84 (mean age=60) who were retested after a one-week delay. There were low test-retest correlations for measures that distinguished AD groups from controls, including information content (.56), pictorial themes (.50), word finding delays (.50), and semantic paraphasias (−.10). This can be explained by practice effects (particularly following the short one-week interval), and ceiling effects in this healthy and relatively young cohort. We were unable to identify any other study reporting test-retest reliability for measures specifically designed for the MCI/AD population.

Inter-Rater Reliability—The majority of studies examined in this review reported on some form of inter-rater reliability, with coefficients ranging between .78 and .99 for coding of language variables (most notably content units), although some studies did not report actual correlation coefficients. Very few studies reported on inter-rater reliability of the transcription process itself, which may also add unnecessary variability to results (Garrard, Haigh, & de Jager, 2011; Macwhinney, Fromm, Forbes, & Holland, 2011).

Parallel Test Form Reliability—Forbes McKay & Venneri (2005) collected normative data for two simple picture stimuli (Cookie Theft and Tripping Woman) and two complex pictures (Chaos Traffic Scene and Bus Stop Scene)(Forbes-McKay & Venneri, 2005). Responses to the two simple pictures showed high correlations for grammatical form, information content, and error monitoring, but low correlations for pictorial themes and word-finding delays, most likely due to ceiling effects of these measures in a healthy, young group. Results were similar for the complex pictures – i.e., high correlations on most measures. Another study compared descriptions of two pictures differing in content level (high vs. low) in an AD group and a normal control group. Interestingly, there was a significant group by content interaction, as the AD group produced more content units in the low content picture than in the high content picture relative to controls (Ehrlich, Obler, & Clark, 1997). March et al. (2006) compared the Cookie Theft to other elicitation tasks (cartoon, describing a map) and found that the Cookie Theft was more sensitive to reduced noun use than the other more complex methods (March, Wales, & Pattison, 2006).

Normative scores for discourse measures.—Forbes-McKay & Venneri (2005) provided a normative study with cut-off scores for normal individuals up to age 90 years. Specifically, the cut off scores were for the measures included in the BDAE discourse coding scheme (Goodglass & Kaplan, 1983), as well as some additional measures specifically for AD (response to word finding delays, information content, speech monitoring, and information content specific to the picture). They performed multiple regressions adjusting for age, sex and education, and obtained cut-off scores based on tolerance limits, and found that with these cut-off scores, individuals with “minimal” AD were able to be correctly identified with 70% accuracy.

Construct Validity—Only one study reported on correlations among connected language measures and standardized language measures: Kave & Goral (2016) compared language performance across picture naming, semantic verbal fluency, and Cookie Theft picture description tasks in adults with AD (n=20) versus controls (n=20). Some measures of connected speech correlated with the picture-naming scores (e.g., percent content words, nouns, and pronouns; mean frequency of words; mean word length), but only one dimension of connected speech (percent nouns) correlated with semantic fluency. The authors suggested that the lack of significant correlation between measures of connected speech and semantic fluency tasks could be due to two factors: 1) the picture description task, like the picture naming task, had visual support which the fluency task did not; 2) verbal fluency tasks are highly dependent on executive functions, whereas picture description may be more reliant upon semantic stores. The authors support their hypothesis by noting that the control group showed a lack of correlation between the naming scores, verbal fluency scores, and the connected speech variables, implying that each task is dependent upon different retrieval demands.

Discriminant Validity—Several studies compared picture description task data among clinical groups, such as vascular or mixed dementia vs. pure AD, early- vs. late-stage AD, or AD vs. fronto-temporal variants. These studies aimed to determine if picture description tasks were sensitive to between-groups differences. The following is a description of the ways in which picture description tasks were used to discriminate among groups:

Stages of Cognitive Decline.

MCI.: There is mixed evidence that discourse measures discriminate among stages of disease. Bschor et al. (2001) found that standardized tasks such as Boston Naming Test and category/letter fluency tasks were more sensitive to MCI than the picture description task measures. In the study by Forbes et al. (2002), adults with "minimal" AD dementia (n=10; MMSE=24–30; akin to "MCI" definition (Petersen et al., 1999)) had significantly more word-finding delays and produced more semantic paraphasias than controls, and those measures also differentiated patients in the mild AD stage (n=10; MMSE = 19–23) from controls. Ahmed compared retrospective data from adults at different stages of disease who had autopsy-confirmed AD to data from controls, and found that measures of semantic and lexical content and syntactic complexity captured the progression of impairment through stages of disease from MCI to moderate AD (Ahmed et al., 2013a).

AD Dementia: Carlomango et al. (2005) calculated total words, information units, information units per minute, words per minute and percent correct information units in discourse samples, all of which distinguished between controls and adults with AD, but not between adults with mild versus moderate AD.

Primary Progressive Aphasia versus AD.: Primary progressive aphasia (PPA) is a neurodegenerative disorder in which speech and language are the primary clinical symptoms. Three subtypes of PPA described in the literature include the semantic variant (svPPA), the nonfluent agrammatic variant (naPPA), and the logopenic variant (lvPPA) (characterized by anomia, and difficulty with sentence repetition). Particularly with recent technological advances regarding biomarker detection, the underlying pathology of lvPPA may be characterized as an AD variant more than a frotontemporal lobar degeneration (Knopman & Nestor, 2017).

Nonetheless, examination of spontaneous speech for PPA variants Primary Progressive Aphasia has become part of routine diagnostic examination (Gorno-Tempini et al., 2011).

Sajjadi et al. (2012) compared data among Semantic Dementia (SD), AD dementia, and normal control groups, and found that the AD dementia and SD groups were similar in speaking rate, proportions of closed-class words, and number of redundant words and phrases, and both produced fewer pictorial themes than controls. The SD group differed from the AD group on a separate spontaneous speech task (a semi-structured interview), in which the SD group produced significantly more semantic and morpho-syntactic errors than the AD and control groups. Ahmed et al. (2012) asked if a group of autopsy-confirmed AD patients showed similar language patterns to those seen in the logopenic variant of Primary Progressive Aphasia (lvPPA), and found that only one patient with AD pathology out of 18 showed a similar pattern to lvPPA. The authors concluded that lvPPA represents an atypical clinical presentation of AD rather than a common clinical feature of typical AD.

Alzheimer's Disease versus Parkinson's Disease.: One study compared speech monitoring between persons with AD and PD and normal controls (McNamara et al., 1992). Both the AD and PD groups corrected significantly fewer speech errors than controls. While AD patients tended to use single-word repairs, PD patients were more like to use both single-word and sentence-level reformulations.

Alzheimer's Disease versus Vascular Dementia.: Nicholas et al. (1985) used the Cookie Theft task to describe empty speech in four groups: AD, Wernike's aphasia due to stroke, post-stroke anomic aphasia and normal controls. The AD group consisted of 19 people with AD dementia ranging from mild to moderately severe. The group with AD produced significantly fewer content elements than control subjects, and adults with Wernike's aphasia due to stroke had the least informative samples of all 4 groups. The distinguishing factor between Wernike's aphasia post-stroke and AD dementia was that the Wernike's group produced more neologisms than the AD group, and the authors concluded that neologistic speech can serve as a diagnostic marker. The AD group was heterogeneous in severity of dementia, which potentially confounded interpretation of results, as lexical deficits increase in severity as the disease progresses. Vuorinen et al. (2000) and Carlomango et al. (2005) examined semantic measures in Vascular Dementia/stroke-related aphasia and AD dementia and found that the two groups performed similarly and worse than control subjects, and that semantic measures did not distinguish between the two dementia groups.

Alzheimer's Disease versus Depression.: Although adults with Alzheimer's disease dementia often present with comorbid depressive symptoms (Rapp et al., 2008), and depressive symptoms have been associated with an increased risk of developing dementia (Modrego et al., 2004), some elderly patients with depression may be misdiagnosed with dementia. This may be due to the similarities in clinical presentation of depression and dementia, such as deficits in memory, attention, and processing speed; social withdrawal; anxiety; and apathy (Dobie, 2002). Murray (Murray, 2010) used a picture description task and found that adults with depression performed similarly to controls on all measures of discourse, while adults with AD expressed fewer content information units than both groups. The authors conclude that narrative analysis may help with differential diagnosis of depression and dementia.

Language Measures Examined in Picture Description Tasks

The studies we reviewed examined a variety of measures across several domains of expressive language. The following is a summary of the measures grouped by semantic content, syntax and pragmatic language.

Semantic Content

Picture-related Content Thematic Elements.: Several studies examined the amount of thematic elements expressed that were directly relevant to the picture stimulus. The studies used a variety of phrases to denote these "thematic elements," including "pictorial themes," "relevant observations," and "semantic units." The following is a description of the findings regarding picture thematic elements: Nicholas et al (1985) identified 8 thematic elements of the Cookie Theft picture and used the number of elements expressed as the outcome measure in four groups: AD, Wernike's aphasia resulting from stroke, post-stroke anomic aphasia and normal controls. The AD group consisted of 19 people with AD ranging from mild to moderately severe (mean age=67). Patients with AD expressed significantly fewer content elements than the controls; differences between the AD group and the groups with stroke-related aphasia were not significant.

Hier et al. (Hier, Hagenlocker, & Shindler, 1985) assessed content using a similar list of 8 actions and objects (e.g., cookie, mother, washing, stool tipping) in the Cookie Theft picture, which they referred to as "relevant observations", and divided their AD group into "early" and "late" stage. The late-stage AD group produced significantly fewer relevant observations than the early-stage group, and the AD group combined produced fewer relevant observations than controls. These findings were replicated by Vuorinen, Laine, & Rinne (2000).. Cuetos (Cuetos, Arango-Lasprilla, Uribe, Valencia, & Lopera, 2007) used a similar set of 8 themes ("semantic units"), to analyze Cookie Theft picture descriptions in a group of 19 carriers of the E280A mutation in the Presenilin1 gene, and found that, prior to the onset of clinical symptoms, carriers produced fewer semantic units than non-carriers. Conversely, Smith, Chenery, & Murdoch (1989) applied Hier's pictorial "theme" construct with the Picnic Scene from the Western Aphasia Battery with a control group and participants with moderate to moderately severe AD. The authors found that the two groups produced an equal number of semantic elements, but the group with AD required more time and more syllables to communicate these elements.

Forbes-McKay and Venneri (2005) performed a normative study of discourse variables from Cookie Theft descriptions (as well as "complex" pictures with more elements) using 240 healthy individuals, ages 20 to 90 years, with varying years of education. The authors developed cut scores for a modified "Rating Scale Profile of Speech Characteristics" from the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983). They then used these cut scores to determine the sensitivity of the measures in a group of 30 adults with probable AD (mean age 75) across "minimal," mild and moderate severity (based on Mini Mental State Examination (MMSE) (Cockrell & Folstein, 1987)). The scores that displayed the best discriminant power between controls and AD were *information content* (proportion of phrases containing indefinite terms, inappropriate phrases and redundant words), *7 pictorial themes*, word finding delays, and the response to word finding delays (70% correct classification rate). Sajjadi et al. (Sajjadi, Patterson, Tomek, & Nestor, 2012) examined 10 pictorial themes in pictures descriptions from the Comprehensive Aphasia Test (Swinburn et al., 2004), and found that the group of 20 adults with mild AD produced significantly fewer themes than controls. Bschor (Bschor, Kuhl, & Reischies, 2001) examined Cookie Theft picture descriptions at 4 stages of AD, defined by the Global Deterioration Scale. The authors identified 12 elements of the Cookie Theft picture, including persons, objects, localizations, actions and features, and found that while each AD group differed significantly from the others and from controls, the measures did not distinguish between MCI and normal controls. The authors concluded that there was no advantage to using the Cookie Theft picture task over standardized language tests such as verbal fluency or the Boston Naming Test in distinguishing among severities of AD.

Beginning with Croisile et al. (1996) several studies (Ahmed, de Jager, Haigh, &Garrard, 2012a; Ahmed et al., 2013a; Ahmed et al., 2013b; Carlomagno, Santoro, Menditti, Pandolfi, & Marini, 2005; Fraser et al., 2014; Kavé & Levy, 2003; Lira, Minett, Bertolucci, & Ortiz, 2014; Shimada et al., 1998) used measures combining thematic elements and unspecified information units, resulting in a list of 23 possible information units of the Cookie Theft picture. The authors claimed that analysis of information units provided a more liberal (yet still constrained) tally of relevant content produced by the speaker, including objects, actions/facts, places and subjects, and that by having this wider range of criteria, more subtle deficits could be observed. Studies using the 23-information-unit system found differences between adults with AD and controls, and some studies found differences among stages of AD (Ahmed et al., 2013b; Forbes, Venneri, & Shanks, 2001).

General Information Units.: Other studies (Arkin & Mahendra, 2001; Giles et al., 1996; C. K. Tomoeda & K. A. Bayles, 1993; Tomoeda et al., 1996) have used a more general concept of content, defining "information units" as "the smallest non-redundant meaningful fact or inference," whether or not the information conveyed was specific to the picture stimulus. Giles (Giles et al., 1996), for example, studied 48 adults with "minimal," mild, or moderate AD and found that adults with AD produced fewer overall information units than controls.

Conciseness of Information.: Conciseness has been operationalized as the number of words the speaker uses to express ideas. The theory is that people with AD will need more words to convey ideas because of word-finding-related behaviors such as circumlocutions and repetitions. Conciseness is calculated by dividing the number of ideas expressed by the total number of words. a measure commonly referred to as "idea density," and has also been described as a measure of richness of language content (Smolik et al., 2016)Sixteen studies reviewed here examined some index of content words to total words (S. Ahmed, C. A. de Jager, A.-M. F. Haigh, & P. Garrard, 2012; S. Ahmed et al., 2012a; Ahmed et al., 2013a; Ahmed et al., 2013b; Bayles et al., 1999; Croisile et al., 1996; Duong, Giroux, Tardif, & Ska, 2005; Forbes-McKay, Shanks, & Venneri, 2013; Forbes-McKay & Venneri, 2005; Groves-Wright, Neils-Strunjas, Burnett, & O'Neill, 2004; Hier et al., 1985; McNamara, Obler, Au, Durso, & Albert, 1992; Murray, 2010; Sajjadi et al., 2012; Shimada et al., 1998; C. K. Tomoeda & K. A. Bayles, 1993; Tomoeda et al., 1996) Authorsdenoted these indices by a variety of names, including "lexical index," "information content," "idea density," and "information unit conciseness index." For example, Snowdon et al. (Snowdon et al., 1996) examined written discourse from the Nun Study and found that low idea density in early life was associated with reduced cognitive performance later in life. Riley et al. (Riley, Snowdon, Desrosiers, & Markesbery, 2005) extended these findings by concluding that early life idea density was associated with lower brain weight, higher degree of cerebral atrophy, and increased neurofibrillary pathology in later life. Ahmed et al. (2013a) examined idea density in 18 patients with autopsy confirmed AD at the mild stage (MMSE mean = 21) and found that they produced fewer total semantic units than controls, but differences in idea density between the two groups did not reach significance. The study of "empty speech" by Nicholas et al. (1985) examined conciseness with measures thought to contribute to the "non-specificity" of discourse in AD, such as *empty phrases* (common idioms contributing no relevant information), *deictic terms* (e.g., "this," "that" without referents), indefinite terms (e.g., "thing" or "stuff"), pronouns without proper noun antecedents, and repetitions. AD patients produced more of these behaviors than did controls.

Efficiency.: Efficiency is the rate at which meaningful information is conveyed over time, calculated by dividing the total number of information units by the duration in seconds of the speech sample. Smith et al (1989) found that 18 adults with moderately severe AD produced 21 fewer content units per minute than controls, due to increased circumlocutions and repetitions. Murray (2010) used an analogous measure referred to as "performance deviations per minute," in which fillers, irrelevant words, revisions or false starts, vague or nonspecific vocabulary and inaccurate output (e.g., paraphasias) were divided by the total number of minutes in the speech sample; this measure was lower for adults with AD (n=17) than for those with depression (n=18) or normal aging (n=14). The authors suggested that discourse information measures may help disentangle the similarities in symptoms of early AD versus depression in older adults.

Lexical Richness/Diversity.: Eleven studies (Ahmed et al., 2012a; Ahmed et al., 2013b; Croisile et al., 1996; Cuetos et al., 2007; Fraser et al., 2014; Hier et al., 1985; Kavé & Levy, 2003; Lira et al., 2014; Murray, 2010; Sajjadi et al., 2012; Sitek et al., 2015) measured some aspect of lexical richness or lexical diversity, by inspecting word types and totals. Measures included total number of words, total unique words, and the ratio of unique words to total words, referred to as "type-token ratio." More recently, studies have used the Moving Average Type Token Ratio (MATTR) (Covington & McFall, 2010) to identify lexical diversity, because unlike type-token ratio, it is independent of text length. Only two studies (Fraser, Meltzer, & Rudzicz, 2015; Rentoumi, Raoufian, Ahmed, de Jager, & Garrard, 2014) used MATTR and other computational measures such as Brunet's index and Honore's statistic. Fraser (Kathleen C Fraser et al., 2015), using cookie theft samples from a corpus of 167 adults with AD, found that Honore's statistic was lower for adults with AD than controls, and Rentoumi's (Rentoumi et al., 2014) machine classifier results showed lower Honore's statistic for 18 patients with mixed pathology versus those 18 participants with AD pathology alone. Other studies investigated lexical content through ratios of open class/closed class words (Ahmed et al., 2012; Ahmed et al., 2013a; Ahmed et al., 2013b), measuring different parts of speech such as verbs and pronouns (Ahmed et al., 2013b; Fraser et al., 2014); proportions of pronouns, nouns, adjectives and verbs; ratio of nouns to verbs, pronouns to nouns plus pronouns, and verbs to nouns plus verbs. For example, in Ahmed et al.'s (2013b) study of 9 adults with autopsy-confirmed AD, differences in the proportion of pronouns were significantly different between MCI and moderate AD groups, and Fraser et al. (2015) found that proportion of pronouns and proportion of nouns and verbs, classified language samples of participants with AD versus healthy controls.

Quantity – Total Number of Words.: Several studies (Croisile et al., 1996; Giles et al., 1996; Cheryl K Tomoeda & Kathryn A Bayles, 1993) reported that adults with moderate AD produce fewer words than controls on picture-description tasks, whereas others (Bschor et al., 2001; Smith et al., 1989; Tomoeda et al., 1996) found no differences in total words among groups of controls and patients with MCI, mild AD, and moderately severe AD. Murray (Murray, 2010) investigated normal controls, AD patients, and older adults with depression, and found no group differences in total words; and Nicholas (Nicholas et al., 1985) found no differences in total words among groups of patients with anomic aphasia post-stroke, stroke-related Wernike's aphasia, AD, or normal controls. Conversely, De Lira (Lira et al., 2014) found that controls produced more total words than patients with AD, but there was no difference in quantity between mild and moderate AD groups.

Syntax and Morphology (Language Form)—"Syntax" refers to the rules that govern how words can be combined to form sentences, while "morphology" is the system that governs the structure of words and the construction of word forms. Multiple studies included at least one measure of syntactic complexity. Common constructs included *words per clause* (Kave & Levy, 2003); *grammatical form* (measures of an "appropriate use" of syntactic conjunctions, tenses, conditionals, subordinate clauses and passive constructions) (Forbes-McKay & Venneri, 2005; Lai, Pai, & Lin, 2009); measures of "phrase length" (also called *t-unit length*) (Fraser, Meltzer, & Rudzicz, 2015; Sajjadi et al., 2012), and *proportion of words in sentences* (S. Ahmed, C. A. de Jager, A. M. F. Haigh, & P. Garrard, 2012b)); number and type of clauses in a sentence (e.g., dependent vs. subordinate clauses); clause types that differ in complexity (e.g., *passive vs. infinitive vs. coordinate clauses* (Hier et al., 1985; Kavé & Levy, 2003)*)*; the proportion of verbs to nouns plus verbs (based on the premise that syntactic simplification is indexed by number of verbs used in an utterance) (Kavé & Levy, 2003); *pronoun/referent ratio* (Chapman, Ulatowska, King, Johnson, & McIntire, 1995); and errors in sentences, referred to either as *grammatical errors, syntactic errors* (Kavé & Levy, 2003), or *verb-agreement errors (*Sajjadi et al., 2012). One study used a syntactic complexity index, consisting of complex clauses divided by total number of clauses (Duong et al., 2005). Several studies examined *mean length of utterance (MLU),* defined as the number of morphemes (smallest unit of meaning) divided by the total number of utterances (S. Ahmed et al., 2012a; Ahmed et al., 2013b; Hier et al., 1985; Murray, 2010; Ripich, Fritsch, Ziol, & Durand, 2000). The general hypothesis motivating these studies is that either working memory limitations or semantic memory limitations in AD affect one's ability to use complex grammatical constructions.

Mean Length of Utterance (MLU): Murray (2010) found that MLUwas not a distinguishing factor among healthy adults, adults with depression and adults with AD dementia. Ripich et al. (2000) found a decrease in MLU in adults with severe AD over time, but not for those with mild or moderate AD.

Proportion of verbs to nouns plus verbs.: Kave & Levy (2003) used a verb index to capture synactic complexity, and found that adults with AD dementia expressed the same amount of verbs to nouns plus verbs as the adult controls.

Syntactic Complexity – Composite measures of MLU, syntactic errors, verbs.: Ahmed et al. (2012a; 2013b) found differences in syntactic complexity between adults with MCI and controls, and between MCI and moderate AD stages. The differences in syntactic complexity were not significant when individual measures were tested, but were apparent using a composite score consisting of MLU, words in sentences, syntactic errors, nouns with determiners, and verbs with inflections.

Pragmatic Language—The following section outlines measures that fall under the pragmatic language domain, which refers to the social rules of language for the purpose of communication, including: 1) using language to achieve goals (Ciccia & Turkstra, 2002); 2) using information from the context to achieve these goals; and 3) using the interaction between people to initiate, maintain, and terminate conversations (Ciccia, 2011).

Questions, turn-taking, unsure statements, egocentric comments.: One study (Ripich et al., 2000) examined several pragmatic abilities among patients at different stages of AD (early (n=10; MMSE=23); middle (n=10; MMSE=16.2) late (n=10; MMSE=10.3 The severe AD group asked more questions over time than the other groups. The authors argued that question-asking was a compensatory strategy, and as a result, adults in late-stage AD may have had insight into their communication problems. The authors acknowledged, however, that the study had a number of possible confounds. First, a caregiver was asked to administer the picture description task in order to mirror a more typical communicative interaction. While this method achieved the ecological goal of the study, caregiver interactions were varied and uncontrolled. Some caregivers asked questions, and provided prompts and encouragement while others did not. Second, due to the constrained nature of the picture description task, it is unlikely that it captured pragmatic skills that are typical of conversations in everyday life.

Coherence.: Coherence refers to the appropriate maintenance of topic in discourse (Halliday & Hasan, 2014), measured by how closely an utterance is thematically related to the immediately preceding utterance (local coherence) and by how closely an utterance relates to the general topic at hand (global coherence) (Laine, Laakso, Vuorinen, & Rinne, 1998). One study examined coherence: Chapman et al. (1995) used picture descriptions of Norman Rockwell prints within a frame analysis, with “frames” being defined as internalized knowledge structures. The authors identified aspects of content, including typical frames of interpretation, atypical, incorrect, or no frames, propositions supporting frames, and propositions disrupting frames as measures of coherence. They examined these variables in three groups: early stage AD (n=12), “old-elderly” (n=12) and normal controls (n=12). The old-elderly and normal controls produced significantly more typical frames, and more frame-supporting information than the AD group. The authors attributed AD patients’ difficulties to memory deficits, attentional deficits, visual perceptual problems, disruption of internalized frame representation, or failure to access frame knowledge.

Perseveration.: One study (Bayles, Tomoeda, Kaszniak, Stern, & Eagans, 1985) examined verbal perseveration in the description of Norman Rockwell prints, dividing total number of words within perseverations by total number of words in the speech sample. The authors also calculated rate of perseveration on two other language tasks: confrontation naming and generative naming. Across all tasks, the AD group produced significantly more perseverations than controls; however, on the picture description task alone, there were no significant differences between the two groups. The authors postulated that this was because picture description was an easier task because there was the visual stimulus, similar to the argument made by Bschor et al. (2001).

Speech Intonation/Prosody.: Three studies (Forbes-McKay et al., 2013; Forbes-McKay & Venneri, 2005; Forbes, Venneri, & Shanks, 2002) examined "melodic line", a subjective measure of speech prosody defined as "the appropriate use of intonational contour, including alterations in pitch, volume and duration" (Goodglass & Kaplan, 1983). For instance, Forbes-McKay et al. compared melodic line using a "simple" picture (Cookie Theft or "The Tripping Woman," (Semenza & Cipolotti, 1989) versus a "complex" picture ("Bus Stop" or "Chaos", unpublished, designed by the researchers). The number of pictorial themes differentiated the simple tasks from the complex. Results showed no group differences on simple picture tasks (Cookie Theft and Tripping Woman) but there were differences in melodic line for the complex picture tasks. Fraser et al. (2015) examined several acoustic features of speech in patients with AD using machine learning methods that captured both rate and phonation patterns, and found acoustic abnormalities to be a significant factor. Konig et al. (2015) also used an automated feature analysis examining vocal and temporal aspects of speech among controls, and patients with MCI or AD, and reported a classification accuracy of 81%.

Discourse Fluency—"Verbal fluency" is a term used in neuropsychological contexts generally referring to timed, word-generation tasks, while in Speech-Language Pathology contexts, "fluency disorders" are defined as interruptions in the flow of speaking characterized by atypical rate, rhythm and repetitions in sounds, syllables, words and phrases. "Fluency," in the literature of discourse of adults with AD, typically refers to the smoothness or flow of spoken language. Abnormalities of fluency in this population are typically characterized by filled and unfilled pauses, word repetitions, circumlocutions, and revisions. In contrast with fluency disorders (i.e., stuttering), abnormalities in the fluency of adults with MCI or AD are rarely characterized at the phonological level.

Ten studies reported on some aspect of discourse fluency. The study of "empty speech" by Nicholas et al. (1985) was one of the first to examine aspects of fluency in the connected speech of persons with AD. The authors found that adults with AD (n=19) had significantly more repetitions than controls (n=30) in Cookie Theft descriptions. Similarly, Tomoeda et al. (1996) found more aborted phrases, revisions, and ideational repetitions in the AD group than the controls. Several other studies showed a greater number of repetitions (Fraser et al., 2014; Hier et al., 1985) and revisions (Hier et al., 1985; Sajjadi et al., 2012) in AD groups versus controls. These findings contrast with studies examining fluency earlier on the disease continuum, which did not find differences between adults with vs. without AD. (Ahmed et al., 2012a; Ahmed et al., 2013b). From these findings, it appears that dysfluent behavior is a reflection of moderate to severe stages of disease when semantic processing is overtly compromised.

Speech Rate—Speech rate typically is calculated as syllables per minute. Of the six studies that have evaluated speech rate (Ahmed et al., 2012a; Ahmed et al., 2013a; Ahmed et al., 2013b; Giles et al., 1996; Murray, 2010; Smith et al., 1989), none found significant differences among groups.

Speech Monitoring—Speech monitoring is related not only to word retrieval deficits and to pragmatic language skill, but also to "anosognosia," the awareness of one's own deficits. McNamara et al. (1992) investigated word error monitoring skills by comparing uncorrected and repaired errors in adults with AD (n=15) or Parkinson Disease (PD) (n=22) vs. normally aging adults (n=141). AD and PD groups were equally impaired in error monitoring as compared to the controls. Severity of naming deficits correlated negatively to the amount of uncorrected errors ($r=-.37$). The authors suggested that this impairment was related to executive function difficulties in the clinical groups. The authors did not report correlations between error monitoring and executive function test scores, however, which could have strengthened that hypothesis.

Another measure of speech monitoring used in the picture description literature is "response to word finding delays," defined as "whether patients appear unaware of their problem, produce an approximation of the target word, or actively search and produce the target word" (Forbes-McKay & Venneri, 2005). Response to word finding delays differed significantly between minimal AD and normal controls (Forbes-McKay & Venneri, 2005; Forbes et al., 2002). The measure was based on Goodglass & Kaplan's (Goodglass & Kaplan, 1983) scale for rating discourse on the Boston Diagnostic Aphasia Examination, and is comprised of clinical judgment of behavior that is rated on a Likert-type scale ranging from 1 to 7 (7 being no abnormality).

Summary of Language Measures Results.—In general, semantic content of picture descriptions is the most frequently cited measure of language degradation, followed by some measure of idea density. A major limitation of the studies reviewed here was lack of age-matching of healthy controls to subjects with AD, as several studies included AD groups that were older than the controls. Variables such as voice onset time, syntactic complexity, and fluency can be affected by aging, so age difference between groups is a potential confounding factor. In some studies, AD groups were not adequately described in terms of severity, which may lead to an oversimplification of results. In addition, studies used multiple names for the same construct (e.g., 'content units', 'thematic units', 'semantic units'), which adds to difficulty in aggregating and interpreting results across studies. Additionally, measures such as "melodic line", "unspecified content units", and "response to word finding" are subjective, and thus would be difficult to reproduce in other studies. Finally, some studies reported either limited information on inter- and intra-reliability of transcriptions/scoring or no information at all.

Sex Differences in Picture Description Abilities—Bayles et al. (1999) examined sex differences in idea density (information units/total words) in a group of men and women with probable AD with similar dementia severity. Within both cross-sectional and longitudinal designs, the authors found no significant differences between men and women on any measure.

Longitudinal Analyses using Picture Description—Four studies examined picture description results longitudinally. Forbes-McKay et al.(2013) found that phonological paraphasias (speech sound errors) were significantly increased at a 12-month follow up, but that other deficits noted at baseline (word finding delays, semantic paraphasias, and fewer repaired errors) had not changed significantly at 12 months. Ahmed et al. (2013b) found increasing deficits in semantic and lexical content and syntactic complexity across three stages of disease in subjects with pathologically confirmed AD. Bayles et al. (1999) found no significant differences in idea density between men and women at a two-year follow-up.

Brain Imaging and Picture Description Tasks—Shimada et al. (1998) studied regional cerebral blood flow (rCBF) during the Picture Description Task (Ulatowska, 1985), and found significant correlations between number of information units and blood flow changes in the occipital lobe and left thalamus, suggesting “thalamo-cortical functional circuits.” Due to the small sample size and the single-time point measurement, these results need replication.

DISCUSSION

Without question, Alzheimer’s Disease degrades the language system, and more broadly, the *communication* system, at some point on the continuum of cognitive decline. Table 2 provides a summary of the key findings of this review, and Table 3 provides an aggregated list of future research directions resulting from this review. These results show that a task as simple and non-invasive as describing a picture can yield rich information about language content, syntax, and pragmatics at several points in the course of cognitive decline. Results are more robust at later stages of AD, but are more fragile and inconclusive at the MCI stage. Certainly the retrospective studies of earlier life writings indicate that a process of change may be occurring pre-clinically, or that they had reduced cerebral reserve of earlier origin; however, writing and speaking engage different cognitive processes. Perhaps the earliest evidence of change in connected language comes from the prospective study by Cuetos et al. (2007) that examined preclinical disease in carriers of the E280A mutation of the PSN-1 gene, and these results were promising: a reduction in information units at a young age and at a truly preclinical state with no evidence of clinical impairment. This finding was replicated in a pre-clinical population showing sub-clinical declines in memory by Mueller et al. (2016) although the sample was an at-risk cohort; whether or not these individuals represent a true preclinical AD group is unknown. Longitudinal designs that study connected language in preclinical or at-risk groups prospectively are needed to determine when connected language can be a sensitive and informative measure of early cognitive decline.

The relatively few studies of the spoken language of adults with MCI showed promise that detectable differences in connected speech are evident early on the AD continuum, but clearly more research is needed.

The psychometric properties of connected language analysis from picture description tasks have been minimally explored. Forbes-McKay et al. (2005) gathered normative data, cut-off scores using a sample of 240 adults between 18 and 90 years of age. The examined test-retest reliability in a subset of 40 individuals following a one-week delay. Test-retest stability measures of picture descriptions at longer intervals (e.g., one-to two-year intervals more typical of longitudinal studies of aging) are not represented in the current literature. In order to make more informed assumptions about changes in normal aging and in persons with cognitive decline, more studies examining the test-retest reliability stability of these measures are needed. Measures of convergent validity (the extent to which measures of discourse are related to the theoretical constructs they are purported to measure) and ecological validity (the extent to which picture descriptions are related to everyday communicative functions) are also needed in order to make inferences about performance.

The study by Kave & Goral (2016) illustrates the probable underlying differences between speech and language in discourse versus in isolation: while picture description measures accurately distinguished controls from AD patients, the authors found weak or minimal correlations of several connected speech measures with focal tasks such as picture naming and verbal fluency. This indicates that connected speech may be able to tap into additional processes not accessed by way of standardized measures, and may serve as an important marker for early diagnosis and for clinically meaningful change.

Review of discourse studies revealed that the notion that syntax is preserved at least until the moderate stages of AD is an oversimplification. While many of the rules of grammar are maintained until late in the disease, likely because they are governed by procedural memory systems, there is accruing evidence that syntax becomes simplified even in early stages of disease (Ahmed et al., 2013b; Ripich et al., 2000). Automated methods, larger sample sizes, and longitudinal analyses are necessary to understand when and to what degree syntax is affected.

Many of the studies here presented with common limitations. First, the sample sizes were often small (mean *n* across the 38 studies = 35) and lacking in diverse ethnic and cultural backgrounds, or did not report on racial background at all. The majority of studies did not report on hearing status, which may have confounded results (particularly with respect to concurrent neuropsychological tests). Some studies did not clearly define stages of disease. For example, Chapman et al. (1995) classified the AD group as “mild AD,” however the group’s mean Logical Memory score was 3.7 (SD=2.2), which is significantly lower than the scores reported in the literature for mild AD (Petersen et al., 1999). Finally, many studies did not match controls according to age, gender and education, which would have strengthened results.

The general conclusions of this review highlight the importance of longitudinal analysis of connected speech and language, to better and more accurately describe the progression of language changes beginning in MCI or before. Only five of the studies reviewed were longitudinal designs (Ahmed et al., 2013b; Bayles et al., 1999; Bayles, Tomoeda, & Trosset, 1993; Forbes-McKay & Venneri, 2005; Ripich et al., 2000), and each showed that some aspect of connected language showed change at follow-up. These studies' sample sizes ranged from 9 to 31 (mean=16), and the samples' mean age ranged from 71 to 80 (mean=74.9). Larger-scale studies, beginning at younger ages and with more points of follow-up may not only better characterize language change, but also the test-retest stability of these measures in people who are not declining in language and memory.

Only one study examined sex differences in picture description and AD (Bayles et al., 1999) and found no differences; however, other findings suggest that AD pathology is differentially manifested clinically between sexes (Barnes et al., 2005). For instance, findings suggest that the association between AD pathology and clinical AD is substantially stronger in women than in men (Barnes et al., 2005), while sex differences favor women in verbal learning and memory (Lewin, Wolgers, & Herlitz, 2001). However, potential confounds exist in studying sex-based differences in cognitive outcomes for MCI and AD, such as: differences in hormonal history; history of depression, anxiety or substance abuse; and differences in education, lifestyle and occupational history. Therefore, an area of need for future research is to examine sex differences in connected speech using longitudinal analyses with larger sample sizes, detailed personal, health and lifestyle histories, and across various stages of AD. Using multi-modal imaging biomarkers as a concurrent means of assessing AD pathology while studying sex-based differences in connected speech over time may reveal further information about AD risk.

Much of the research reported here required arduous, labor-intensive hand transcription, counting, and coding. Other promising research methods are automatic feature classification systems, machine learning techniques, and natural language processing, which have now been applied to the study of speech and language in AD (Fraser et al., 2015; König et al., 2015; Rentoumi et al., 2014). Automated methods also may assist with analysis of "melodic line" or other aspects of speech prosody that to date have been analyzed using subjective judgments. Despite good training methods and inter-rater reliability, subjective judgements are prone to human error, and machine classification removes some of the human error in obtaining these measures. Technology also allows faster and easier analysis of rate, prosody, pitch, and loudness, and may detect patterns not perceptible to humans. For example, acoustic analyses accurately identified MCI, mild AD and moderate AD with up to 89% accuracy, using vocal features that were not detectable to the human ear (König et al., 2015). If such automated methods prove reliable and valid, the effects could have far-reaching clinical implications, including a low-cost, low-burden longitudinal screening measure that may trigger referrals for more in-depth neuropsychological evaluation.

One conclusion of this review is that the studies reviewed here showed varied if not limited cohesiveness in the terms used to describe identical language constructs. A factor contributing to this heterogeneity of terms may be the varied disciplines approaching the problem (e.g., linguistics, neuropsychology, speech-language pathology). A more cohesive approach by standardizing terms across disciplines may help to aggregate future results, and to better understand the systematic effects of studying connected language for the diagnosis and monitoring of language in AD.

The narrow focus of our review, in terms of both tasks and populations, can be considered a limitation. In particular, much information can be garnered from the Primary Progressive Aphasia research about progressive and specific language decline. Several of the PPA studies included picture description tasks, and by excluding those studies without an AD component, we may have missed valuable comparisons among progressive disorders. Because we excluded other connected language elicitation methods, such as personal event narratives and conversations, we were unable to assess strengths and weaknesses of different methods that attempt to achieve the same goal. Future reviews should include task comparisons, to identify the most sensitive, specific, effective, and efficient tasks for eliciting and analyzing connected speech and language.

We included studies that were conducted in languages other than English. For example, Kave et al., 2016 used participants who spoke Hebrew, and found that participants with AD dementia produced a lower percentage of content words relative to all words, and a higher percentage of pronouns to all words. Although these findings are consistent with studies conducted in English reviewed, the linguistic nuances across languages limit generalization of findings from one language to another, and should therefore be interpreted with caution.

In addition to the research limitations discussed above, picture description tasks have inherent limitations. First, while these tasks elicit spontaneous, unrehearsed speech and language, they by no means reflect conversational speech, where the breakdowns in patient-caregiver communication actually occur. Picture description is essentially a monologue: the needs of the listener are limited as opposed to the demands in less concrete, natural conversations. Conversational tasks, in which there is a back-and-forth exchange between two or more people, also require a good deal of executive function, metacognitive and pragmatic skill, and language comprehension, and thus may be more useful in understanding communication challenges in activities of daily living. Conversational speech and language, and its relation to picture description, in an area of research that needs further investigation in MCI and AD.

Second, one of the strengths of picture description tasks can also be considered a limitation; that is, while the task allows for standardization, and thus lends itself to longitudinal assessment and comparison across studies, the picture used should be one that is culturally and socially representative of patients' experiences to best capture that individual's semantic knowledge. Pictures such as the "Cookie Theft" can become dated and irrelevant to future cohorts, and may not be an adequate stimulus across cultures. Future research using picture description tasks should aim to address cultural and generational relevance as the population demographics change, by developing and investigating new and diverse stimuli.

Despite these limitations, there is a significant body of evidence to suggest that analyzing connected speech and language elicited by picture description tasks is a useful and informative tool for detecting and describing language changes across stages of cognitive decline. This method can have far-reaching clinical implications. For example, current advances in biomarkers of Alzheimer's pathology, such as neuroimaging, are expensive and frequently inaccessible to clinical practice; whereas connected language analysis is noninvasive, inexpensive, and poses minimal participant burden. If continued research suggests that spontaneous language analysis is sensitive to subtle declines, the tool could be used to identify individuals who are most likely to benefit from pharmacological clinical trials. Furthermore, there are clinical implications for using connected speech analysis in intervention: cognitive-communication interventions and other non-pharmacological interventions, much like pharmacological interventions, will likely be more effective earlier in the disease course before neuropathological changes become diffuse (Jack et al., 2011; Oren, Willerton, & Small, 2014). At the same time, connected speech can serve as a functional marker for disease progression and response to pharmacological interventions, as it may be a more objective measure than self- or caregiver-report of functional activities of daily living. A 2015 report on "Innovative Diagnostic Tools for Early Detection of Alzheimer's Disease" (Laske et al., 2015) listed analysis of spontaneous speech and voice (rate, voice onset time, variations in pitch and amplitude) and spontaneous language (content, syntax, pragmatics) as one of the innovative tools that deserves further research.

CONCLUSION

Our review shows that picture description tasks, as an elicitation method for obtaining connected language samples, are useful in detecting differences in semantic processing, syntactic complexity, pragmatic language use, and speech and voice parameters between persons with Alzheimer's disease and those who are aging typically, and across stages of disease. While evidence for the tool's sensitivity at the MCI phase or before is limited, there is still indication that subtle changes in speech and language can be detected, particularly longitudinally. Clinically, providing spontaneous speech in this context may present a quick and efficient means of gaining information about language, while posing less burden to the adults with MCI and AD. Finally, technological advances in machine learning, automatic speech recognition, and natural language processing offer exciting opportunities for expansion in knowledge about subtle changes, while affording more efficient and expedient means to analyze speech and language.

REFERENCES

Ahmed S, de Jager CA, Haigh AM, & Garrard P (2012a). Logopenic aphasia in Alzheimer's disease: clinical variant or clinical feature? J Neurol Neurosurg Psychiatry, 83(11), 1056–1062. 10.1136/jnnp-2012-302798 [PubMed: 22842206]

Ahmed S, de Jager CA, Haigh AM, & Garrard P (2013a). Semantic processing in connected speech at a uniformly early stage of autopsy-confirmed Alzheimer's disease. Neuropsychology, 27(1), 79–85. 10.1037/a0031288 [PubMed: 23356598]

Ahmed S, de Jager CA, Haigh AMF, & Garrard P (2012b). Logopenic aphasia in Alzheimer's disease: clinical variant or clinical feature? Journal of Neurology Neurosurgery and Psychiatry, 83(11), 1056–1062. 10.1136/jnnp-2012-302798

Ahmed S, Haigh AM, de Jager CA, & Garrard P (2013b). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. Brain, 136(Pt 12), 3727–3737. 10.1093/brain/awt269 [PubMed: 24142144]

Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, … Phelps CH (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the

National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement, 7(3), 270–279. 10.1016/j.jalz.2011.03.008 [PubMed: 21514249]

Arkin S, & Mahendra N (2001). Discourse analysis of Alzheimer's patients before and after intervention: Methodology and outcomes. Aphasiology, 15(6), 533–569.

ArmakiBarnes LL, Wilson RS, Bienias JL, Schneider JA, Evans DA, & Bennett DA (2005). Sex differences in the clinical manifestations of alzheimer disease pathology. Archives of General Psychiatry, 62(6), 685–691. 10.1001/archpsyc.62.6.685 [PubMed: 15939846]

Bayles KA, Azuma T, Cruz RF, Tomoeda CK, Wood JA, & Montgomery EB, Jr (1999). Gender differences in language of Alzheimer disease patients revisited. Alzheimer Disease and Associated Disorders, 13(3), 138–146. [PubMed: 10485572]

Bayles KA, & Tomoeda CK (1983). Confrontation naming impairment in dementia. Brain Lang, 19(1), 98–114. [PubMed: 6222782]

Bayles KA, Tomoeda CK, Kaszniak AW, Stern LZ, & Eagans KK (1985). Verbal perseveration of dementia patients. Brain Lang, 25(1), 102–116. [PubMed: 3161580]

Bayles KA, Tomoeda CK, & Trosset MW (1993). Alzheimer's disease: Effects on language. Developmental Neuropsychology, 9(2), 131–160.

Berisha V, Wang S, LaCross A, & Liss J (2015). Tracking discourse complexity preceding Alzheimer's disease diagnosis: a case study comparing the press conferences of Presidents Ronald Reagan and George Herbert Walker Bush. J Alzheimers Dis, 45(3), 959–963. 10.3233/JAD-142763 [PubMed: 25633673]

Bickel C, Pantel J, Eysenbach K, & Schröder J (2000). Syntactic comprehension deficits in Alzheimer's disease. Brain Lang, 71(3), 432–448. [PubMed: 10716871]

Blair M, Marczinski CA, Davis-Faroque N, & Kertesz A (2007). A longitudinal study of language decline in Alzheimer's disease and frontotemporal dementia. Journal of the International Neuropsychological Society, 13(02), 237–245. [PubMed: 17286881]

Bloom L, Lightbown P, Hood L, Bowerman M, Maratsos M, & Maratsos MP (1975). Structure and variation in child language. Monographs of the society for Research in Child Development, 1–97. [PubMed: 1102959]

Botting N (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. Child language teaching and therapy, 18(1), 1–21.

Brown R (1973). A first language: The early stages Harvard U. Press.

Bschor T, Kuhl KP, & Reischies FM (2001). Spontaneous speech of patients with dementia of the Alzheimer type and mild cognitive impairment. Int Psychogeriatr, 13(3), 289–298. [PubMed: 11768376]

Buckner RL (2004). Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. Neuron, 44(1), 195–208. [PubMed: 15450170]

Carlomagno S, Santoro A, Menditti A, Pandolfi M, & Marini A (2005). Referential communication in Alzheimer's type dementia. Cortex, 41(4), 520–534. [PubMed: 16042028]

Chapman SB, Ulatowska HK, King K, Johnson JK, & McIntire DD (1995). Discourse in early Alzheimer's disease versus normal advanced aging. American Journal of Speech-Language Pathology, 4(4), 124–129.

Chung JA, & Cummings JL (2000). Neurobehavioral and neuropsychiatric symptoms in Alzheimer's disease. Neurologic clinics, 18(4), 829–846. [PubMed: 11072263]

Ciccia A (2011). Pragmatic Communication. In Kreutzer JS, DeLuca J, & Caplan B (Eds.), Encyclopedia of Clinical Neuropsychology (pp. 1994–1995). New York, NY: Springer New York.

Ciccia AH, & Turkstra LS (2002). Cohesion, communication burden, and response adequacy in adolescent conversations. Advances in Speech Language Pathology, 4(1), 1–8.

Clark LJ, Gatz M, Zheng L, Chen YL, McCleary C, & Mack WJ (2009). Longitudinal verbal fluency in normal aging, preclinical, and prevalent Alzheimer's disease. Am J Alzheimers Dis Other Demen, 24(6), 461–468. 10.1177/1533317509345154 [PubMed: 19759254]

Cockrell J, & Folstein MF (1987). Mini-Mental State Examination (MMSE). Psychopharmacology bulletin, 24(4), 689–692.

Covington MA, & McFall JD (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). Journal of Quantitative Linguistics, 17(2), 94–100.

Croisile B, Ska B, Brabant MJ, Duchene A, Lepage Y, Aimard G, & Trillet M (1996). Comparative study of oral and written picture description in patients with Alzheimer's disease. Brain Lang, 53(1), 1–19. 10.1006/brln.1996.0033 [PubMed: 8722896]

Cuetos F, Arango-Lasprilla JC, Uribe C, Valencia C, & Lopera F (2007). Linguistic changes in verbal expression: a preclinical marker of Alzheimer's disease. J Int Neuropsychol Soc, 13(3), 433–439. 10.1017/S1355617707070609 [PubMed: 17445292]

De Jager C, Blackwell AD, Budge MM, & Sahakian BJ (2005). Predicting cognitive decline in healthy older adults. The American Journal of Geriatric Psychiatry, 13(8), 735–740. [PubMed: 16085791]

De Jager CA, & Budge MM (2005). Stability and predictability of the classification of mild cognitive impairment as assessed by episodic memory test performance over time. Neurocase, 11(1), 72–79. [PubMed: 15804927]

De Renzi A, & Vignolo LA (1962). Token test: A sensitive test to detect receptive disturbances in aphasics. Brain: a journal of neurology

Dobie D (2002). Depression, dementia, and pseudodementia. Paper presented at the Seminars in clinical neuropsychiatry.

Duong A, Giroux F, Tardif A, & Ska B (2005). The heterogeneity of picture-supported narratives in Alzheimer's disease. Brain Lang, 93(2), 173–184. 10.1016/j.bandl.2004.10.007 [PubMed: 15781305]

Ehrlich JS, Obler LK, & Clark L (1997). Ideational and semantic contributions to narrative production in adults with dementia of the Alzheimer's type. J Commun Disord, 30(2), 79–98; quiz 98–79. [PubMed: 9100125]

Evans JL, & Craig HK (1992). Language Sample Collection and AnalysisInterview Compared to Freeplay Assessment Contexts. Journal of Speech, Language, and Hearing Research, 35(2), 343–353.

Forbes-McKay K, Shanks MF, & Venneri A (2013). Profiling spontaneous speech decline in Alzheimer's disease: a longitudinal study. Acta neuropsychiatrica, 25(06), 320–327. [PubMed: 25287871]

Forbes-McKay KE, & Venneri A (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. Neurol Sci, 26(4), 243–254. 10.1007/s10072-005-0467-9 [PubMed: 16193251]

Forbes KE, Venneri A, & Shanks MF (2001). Distinct patterns of spontaneous speech deterioration: an early predictor of Alzheimer's disease. Brain Cogn, 48(2–3), 356–361.

Forbes KE, Venneri A, & Shanks MF (2002). Distinct patterns of spontaneous speech deterioration: an early predictor of Alzheimer's disease. Brain Cogn, 48(2–3), 356–361. [PubMed: 12030467]

Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE, & Rochon E (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex, 55, 43–60. 10.1016/j.cortex.2012.12.006 [PubMed: 23332818]

Fraser KC, Meltzer JA, & Rudzicz F (2015). Linguistic features identify Alzheimer's disease in narrative speech. Journal of Alzheimer's Disease, 49(2), 407–422.

Fraser KC, Meltzer JA, & Rudzicz F (2015). Linguistic Features Identify Alzheimer's Disease in Narrative Speech. Journal of Alzheimer's Disease, 49(2), 407–422.

Frisoni GB, Fratiglioni L, Fastbom J, Viitanen M, & Winblad B (1999). Mortality in nondemented subjects with cognitive impairment: the influence of health-related factors. American journal of epidemiology, 150(10), 1031–1044. [PubMed: 10568618]

Garrard P, Haigh A-M, & de Jager C (2011). Techniques for transcribers: assessing and improving consistency in transcripts of spoken language. Literary and Linguistic Computing, fqr018.

Garrard P, Maloney LM, Hodges JR, & Patterson K (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. Brain, 128(Pt 2), 250–260. 10.1093/brain/awh341 [PubMed: 15574466]

Giles E, Patterson K, & Hodges JR (1996). Performance on the Boston Cookie Theft picture description task in patients with early dementia of the Alzheimer's type: Missing information. Aphasiology, 10(4), 395–408.

Gitlin LN, Winter L, Dennis MP, Hodgson N, & Hauck WW (2010). Targeting and managing behavioral symptoms in individuals with dementia: A randomized trial of a nonpharmacological intervention. Journal of the American Geriatrics Society, 58(8), 1465–1474. [PubMed: 20662955]

Goodglass H, & Kaplan E (1983). Boston diagnostic aphasia examination booklet: Lea & Febiger.

Gorno-Tempini M, Hillis A, Weintraub S, Kertesz A, Mendez M, Cappa S. e., … Boeve B (2011). Classification of primary progressive aphasia and its variants. Neurology, 76(11), 1006–1014. [PubMed: 21325651]

Grossman M, D'Esposito M, Hughes E, Onishi K, Biassou N, White-Devine T, & Robinson K (1996). Language comprehension profiles in Alzheimer's disease, multi-infarct dementia, and frontotemporal degeneration. Neurology, 47(1), 183–189. [PubMed: 8710075]

Groves-Wright K, Neils-Strunjas J, Burnett R, & O'Neill MJ (2004). A comparison of verbal and written language in Alzheimer's disease. J Commun Disord, 37(2), 109–130. 10.1016/j.jcomdis.2003.08.004 [PubMed: 15013729]

Halliday MAK, & Hasan R (2014). Cohesion in english: Routledge.

Hart D, Craig D, Compton S, Critchlow S, Kerrigan B, McIlroy S, & Passmore A (2003). A retrospective study of the behavioural and psychological symptoms of mid and late phase Alzheimer's disease. Int J Geriatr Psychiatry, 18(11), 1037–1042. [PubMed: 14618556]

Hier DB, Hagenlocker K, & Shindler AG (1985). Language disintegration in dementia: effects of etiology and severity. Brain Lang, 25(1), 117–133. [PubMed: 2411334]

Jack CR, Jr., Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, … Phelps CH (2011). Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement, 7(3), 257–262. 10.1016/j.jalz.2011.03.004 [PubMed: 21514247]

Kave G, & Goral M (2016). Word retrieval in picture descriptions produced by individuals with Alzheimer's disease. Journal of Clinical and Experimental Neuropsychology, 38(9), 958–966. 10.1080/13803395.2016.1179266 [PubMed: 27171756]

Kavé G, & Goral M (2016). Word retrieval in picture descriptions produced by individuals with Alzheimer's disease. Journal of Clinical and Experimental Neuropsychology, 38(9), 958–966. [PubMed: 27171756]

Kave G, & Levy Y (2003). Morphology in picture descriptions provided by persons with Alzheimer's disease. J Speech Lang Hear Res, 46(2), 341–352. [PubMed: 14700376]

Kavé G, & Levy Y (2003). Morphology in picture descriptions provided by persons with Alzheimer's disease. Journal of Speech, Language, and Hearing Research, 46(2), 341–352.

Kemper S, Thompson M, & Marquis J (2001). Longitudinal change in language production: effects of aging and dementia on grammatical complexity and propositional content. Psychology and aging, 16(4), 600. [PubMed: 11766915]

Kertesz A (1982). Western aphasia battery test manual: Psychological Corp.

König A, Satt A, Sorin A, Hoory R, Toledo-Ronen O, Derreumaux A, … Robert PH (2015). Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 1(1), 112–124.

Knopman DS, & Nestor PJ (2017). Beyond clinical syndromes in primary progressive aphasia Seeking etiologic diagnoses

Lai YH, Pai HH, & Lin YT (2009). To be semantically-impaired or to be syntactically-impaired: Linguistic patterns in Chinese-speaking persons with or without dementia. Journal of Neurolinguistics, 22(5), 465–475. 10.1016/j.jneuroling.2009.03.004

Laine M, Laakso M, Vuorinen E, & Rinne J (1998). Coherence and informativeness of discourse in two dementia types. Journal of Neurolinguistics, 11(1), 79–87.

Laske C, Sohrabi HR, Frost SM, López-de-Ipiña K, Garrard P, Buscema M, … O'Bryant SE (2015). Innovative diagnostic tools for early detection of Alzheimer's disease. Alzheimer's & Dementia, 11(5), 561–578. 10.1016/j.jalz.2014.06.004

Lewin C, Wolgers G, & Herlitz A (2001). Sex differences favoring women in verbal but not in visuospatial episodic memory. Neuropsychology, 15(2), 165. [PubMed: 11324860]

Lira J. O. d., Minett TSC, Bertolucci PHF, & Ortiz KZ (2014). Analysis of word number and content in discourse of patients with mild to moderate Alzheimer's disease. Dementia & Neuropsychologia, 8(3), 260–265. [PubMed: 29213912]

MacDonald MC, Almor A, Henderson VW, Kempler D, & Andersen ES (2001). Assessing working memory and language comprehension in Alzheimer's disease. Brain Lang, 78(1), 17–42. [PubMed: 11412013]

Macwhinney B, Fromm D, Forbes M, & Holland A (2011). AphasiaBank: Methods for Studying Discourse. Aphasiology, 25(11), 1286–1307. 10.1080/02687038.2011.589893 [PubMed: 22923879]

March EG, Wales R, & Pattison P (2006). The uses of nouns and deixis in discourse production in Alzheimer's disease. Journal of Neurolinguistics, 19(4).

Martin A, & Fedio P (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. Brain Lang, 19(1), 124–141. [PubMed: 6860932]

McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Jr., Kawas CH, … Phelps CH (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement, 7(3), 263–269. 10.1016/j.jalz.2011.03.005 [PubMed: 21514250]

McNamara P, Obler LK, Au R, Durso R, & Albert ML (1992). Speech monitoring skills in Alzheimer's disease, Parkinson's disease, and normal aging. Brain Lang, 42(1), 38–51. [PubMed: 1547468]

Modrego PJ, & Ferrández J (2004). Depression in patients with mild cognitive impairment increases the risk of developing dementia of Alzheimer type: a prospective cohort study. Archives of neurology, 61(8), 1290–1293. [PubMed: 15313849]

Moher D, Liberati A, Tetzlaff J, Altman DG, & and the, P. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. Annals of Internal Medicine, 151(4), 264–269. 10.7326/0003-4819-151-4-200908180-00135 [PubMed: 19622511]

Mueller KD, Koscik RL, LaRue A, Clark LR, Hermann B, Johnson SC, & Sager MA (2015). Verbal Fluency and Early Memory Decline: Results from the Wisconsin Registry for Alzheimer's Prevention. Arch Clin Neuropsychol, 30(5), 448–457. 10.1093/arclin/acv030 [PubMed: 26025231]

Mueller KD, Koscik RL, Turkstra LS, Riedeman SK, Larue A, Clark LR, … Johnson SC (2016). Connected Language in Late Middle-Aged Adults at Risk for Alzheimer's Disease. Journal of Alzheimers Disease, 54(4), 1539–1550. 10.3233/jad-160252

Murray LL (2010). Distinguishing clinical depression from early Alzheimer's disease in elderly people: Can narrative analysis help? Aphasiology, 24(6–8), 928–939.

Nicholas M, Obler LK, Albert ML, & Helm-Estabrooks N (1985). Empty speech in Alzheimer's disease and fluent aphasia. J Speech Hear Res, 28(3), 405–410. [PubMed: 4046581]

Nutter-Upham KE, Saykin AJ, Rabin LA, Roth RM, Wishart HA, Pare N, & Flashman LA (2008). Verbal fluency performance in amnestic MCI and older adults with cognitive complaints. Arch Clin Neuropsychol, 23(3), 229–241. 10.1016/j.acn.2008.01.005 [PubMed: 18339515]

Oren S, Willerton C, & Small J (2014). Effects of Spaced Retrieval Training on Semantic Memory in Alzheimer's Disease: A Systematic Review. Journal of Speech, Language, and Hearing Research, 57(1), 247–270.

Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, & Kokmen E (1999). Mild cognitive impairment: clinical characterization and outcome. Archives of Neurology, 56(3), 303–308. [PubMed: 10190820]

Price BH, Gurvit H, Weintraub S, Geula C, Leimkuhler E, & Mesulam M (1993). Neuropsychological patterns and language deficits in 20 consecutive cases of autopsy-confirmed Alzheimer's disease. Archives of Neurology, 50(9), 931–937. [PubMed: 8363447]

Rao JK, Anderson LA, Inui TS, & Frankel RM (2007). Communication interventions make a difference in conversations between physicians and patients: a systematic review of the evidence. Medical care, 45(4), 340–349. [PubMed: 17496718]

Rapp MA, Schnaider-Beeri M, Purohit DP, Perl DP, Haroutunian V, & Sano M (2008). Increased neurofibrillary tangles in patients with Alzheimer disease with comorbid depression. The American Journal of Geriatric Psychiatry, 16(2), 168–174. [PubMed: 18239198]

Rentoumi V, Raoufian L, Ahmed S, de Jager CA, & Garrard P (2014). Features and machine learning classification of connected speech samples from patients with autopsy proven Alzheimer's disease with and without additional vascular pathology. J Alzheimers Dis, 42 Suppl 3, S3–17. 10.3233/JAD-140555 [PubMed: 25061045]

Rentz DM, Rodriguez MAP, Amariglio R, Stern Y, Sperling R, & Ferris S (2013). Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review. Alzheimer's research & therapy, 5(6), 58.

Riley KP, Snowdon DA, Desrosiers MF, & Markesbery WR (2005). Early life linguistic ability, late life cognitive function, and neuropathology: findings from the Nun Study. Neurobiology of aging, 26(3), 341–347. [PubMed: 15639312]

Ripich DN, Fritsch T, Ziol E, & Durand E (2000). Compensatory strategies in picture description across severity levels in Alzheimer's disease: A longitudinal study. Am J Alzheimers Dis Other Demen, 15(4), 217–228.

Ripich DN, & Terrell BY (1988). Patterns of discourse cohesion and coherence in Alzheimer's disease. J Speech Hear Disord, 53(1), 8–15. [PubMed: 3339870]

Sajjadi SA, Patterson K, Tomek M, & Nestor PJ (2012). Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. Aphasiology, 26(6), 847–866. 10.1080/02687038.2012.654933

Savundranayagam MY, Hummert ML, & Montgomery RJ (2005). Investigating the effects of communication problems on caregiver burden. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60(1), S48–S55.

Savundranayagam MY, & Orange J (2014). Matched and mismatched appraisals of the effectiveness of communication strategies by family caregivers of persons with Alzheimer's disease. International Journal of Language & Communication Disorders, 49(1), 49–59. [PubMed: 24372885]

Semenza C, & Cipolotti L (1989). Neuropsicologia con carta e matita Padova: Cleup Editrice Padova.

Shimada M, Meguro K, Yamazaki H, Horikawa A, Hayasaka C, Yamaguchi S, … Yamadori A (1998). Impaired verbal description ability assessed by the Picture Description Task in Alzheimer's disease. Archives of Gerontology and Geriatrics, 27(1), 57–65. [PubMed: 18653151]

Sitek EJ, Kluj-Kozłowska K, Barczak A, Kozłowski M, Wieczorek D, Przewłócka A, … Sławek J (2015). Overlapping and distinguishing features of descriptive speech in Richardson variant of progressive supranuclear palsy and non-fluent progressive aphasia. Postępy Psychiatrii i Neurologii, 24(2), 62–67.

Small SA, Perera GM, DeLaPaz R, Mayeux R, & Stern Y (1999). Differential regional dysfunction of the hippocampal formation among elderly with memory decline and Alzheimer's disease. Annals of neurology, 45(4), 466–472. [PubMed: 10211471]

Smith SR, Chenery HJ, & Murdoch BE (1989). Semantic abilities in dementia of the Alzheimer type. II. Grammatical semantics. Brain Lang, 36(4), 533–542. [PubMed: 2720369]

Smolik F, Stephankova H, Vyhnalek M, Nikolai T, Horakova K, & Mategjka S (2016). Propositional density in spoken and written language of Czech-speaking patients with Mild Cognitive Impairment. Journal of Speech, Language, and Hearing Research, 59, 1461–1470.

Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH, Wekstein DR, & Markesbery WR (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. JAMA, 275(7), 528–532. [PubMed: 8606473]

Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, … Montine TJ (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia, 7(3), 280–292.

Sperling RA, Dickerson BC, Pihlajamaki M, Vannini P, LaViolette PS, Vitolo OV, … & Johnson KA (2010). Functional alterations in memory networks in early Alzheimer's disease. Neuromolecular medicine, 12(1), 27–43. [PubMed: 20069392]

Swinburn K, Porter G, & Howard D (2004). CAT: comprehensive aphasia test: Psychology Press.

Taler V, & Phillips NA (2008). Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. J Clin Exp Neuropsychol, 30(5), 501–556. 10.1080/13803390701550128 [PubMed: 18569251]

Testa JA, Ivnik RJ, Boeve B, Petersen RC, Pankratz VS, Knopman D, … Smith GE (2004). Confrontation naming does not add incremental diagnostic utility in MCI and Alzheimer's disease. J Int Neuropsychol Soc, 10(4), 504–512. 10.1017/S1355617704104177 [PubMed: 15327729]

Tomoeda CK, & Bayles KA (1993). Longitudinal effects of Alzheimer disease on discourse production. Alzheimer Dis Assoc Disord

Tomoeda CK, & Bayles KA (1993). Longitudinal effects of Alzheimer disease on discourse production. Alzheimer Dis Assoc Disord, 7(4), 223–236. [PubMed: 8305190]

Tomoeda CK, Bayles KA, Trosset MW, Azuma T, & McGeagh A (1996). Cross-sectional analysis of Alzheimer disease effects on oral discourse in a picture description task. Alzheimer Dis Assoc Disord, 10(4), 204–215. [PubMed: 8939280]

Ulatowska HK (1985). The aging brain: Communication in the elderly: College-Hill.

Vuorinen E, Laine M, & Rinne J (2000). Common pattern of language impairment in vascular dementia and in Alzheimer disease. Alzheimer Dis Assoc Disord, 14(2), 81–86. [PubMed: 10850746]

Woodward M (2013). Aspects of communication in Alzheimer's disease: clinical features and treatment options. International Psychogeriatrics, 25(06), 877–885. [PubMed: 23522497]

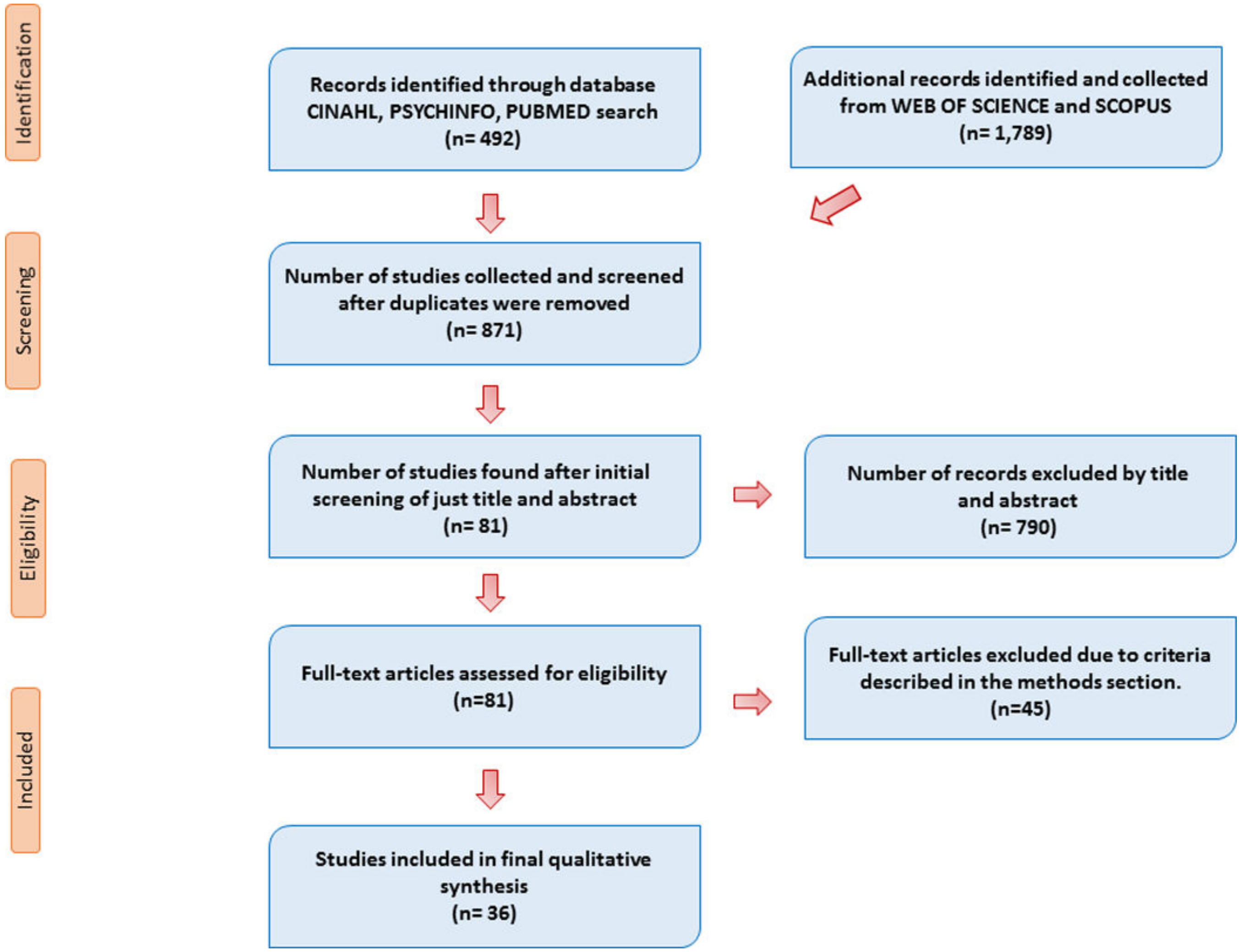

Figure 1.
The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) search and selection process.

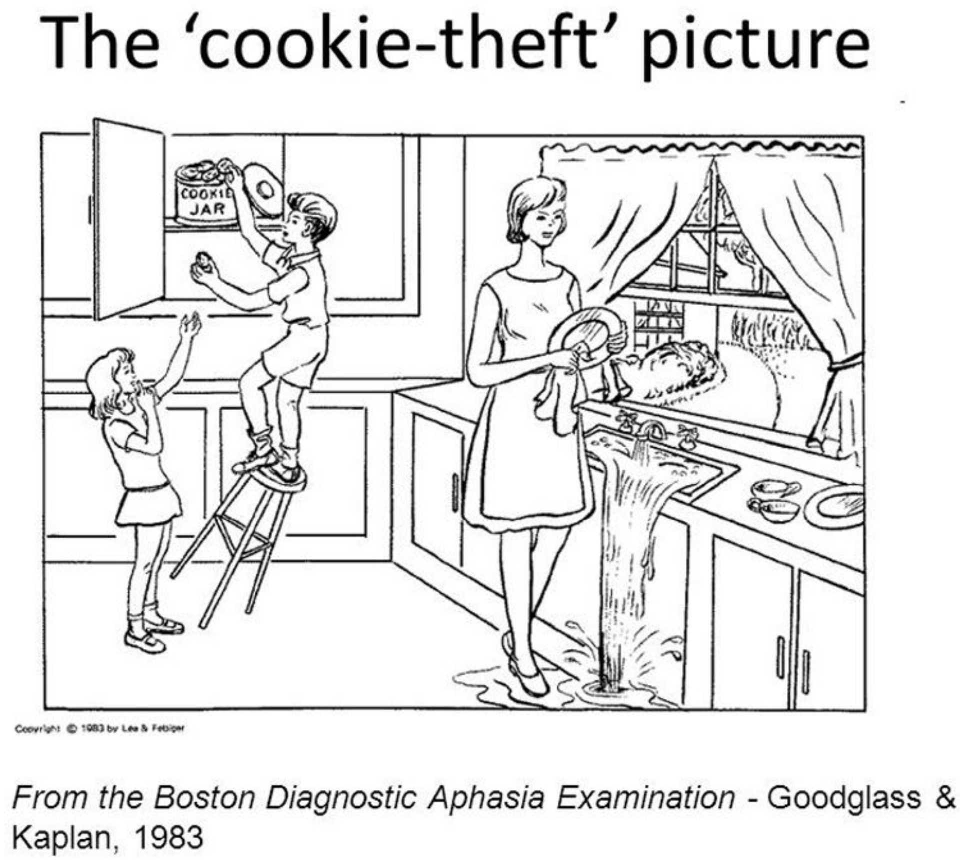

Figure 2.
The "Cookie Theft Picture" from the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983). Reprinted with permission by PRO-ED Inc.

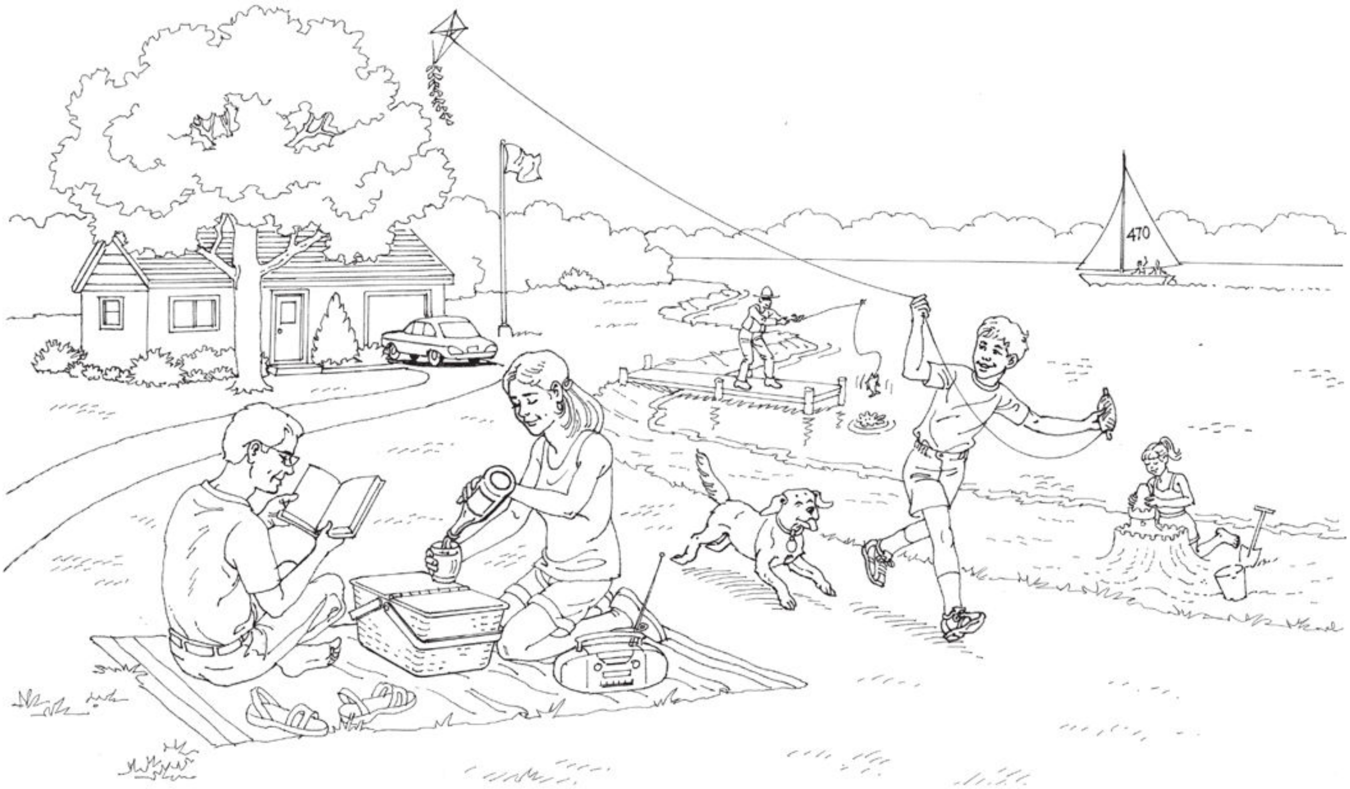

Figure 3.

The “Picnic Scene” from the Western Aphasia Battery, Revised (WAB-R). Copyright c 2006 NCS Pearson, Inc. Used with permission. All rights reserved.

Table 1.

A description of studies selected for final review.

Authors	*Stimulus Picture Used*	*Groups + sample sizes + mean age*	*MMSE Score*	*Diagnostic Criteria*	*Significant Language Variables*
Ahmed et al. (2012)	Cookie Theft (CT)	NC(n=18;age=79.1) AD(n=18; age=74.0)	NC=28.9 AD=21.7	NINCDS-ADRDA; Autopsy- confirmed	No group differences in features associated with logopenic PPA
Ahmed et al. (2013)	CT	NC(n=18;age=79.1) AD(n=18;age=74.0)	NC=28.9 AD=21.7	NINCDS-ADRDA; Autopsy- confirmed	AD <NC on **semantic units, efficiency; decreased verbs** predicted group membership.
Ahmed et al. (2013)	CT	NC(n=9;age=76.0) MCI through moderate AD(n=9;age=71.2)	NC=29.2 MCI=24.2 mildAD=22.2 modAD=12.9	NINCDS-ADRDA; Petersen criteria; Autopsy- confirmed	Longitudinal analysis examining change across MIC, Mild and Moderate AD. No significant differences between each clinical stage; unadjusted p values were <.05 for **proportion of pronouns, semantic units, references to objects, idea density, efficiency**.
Bayles et al. (1993)	NRP	AD(n=94; age=74.2) NC(n=53; age=NR) Longitudinal AD(n=31)	AD:NR NC=29.1	NINCDS-ADRDA	Longitudinal analysis of AD GDS staging and language change; changes in picture descriptions noted during first 12-month period only for moderate AD group (GDS Stage 5)
Bayles et al. (1999)	Norman Rockwell Prints (NRP)	AD Male(n=30, age=77.7) AD Female(n=33, age=80.1)	ADm=15.2 ADf=15.9	NINCDS-ADRDA	No significant differences between male and female **Information Units/total words**, cross-sectional and longitudinal
Bayles et al. (2004)	NRP	NC(n=40; 69.4) AD(n=30; 82.23)	NC=29.1 AD=15.0	NINCDS-ADRDA	No significant differences in rates of perseveration
Bschor et al. (2001)	CT	NC(n=40; 57.2) MCI(n=34; 64.0) mildAD(n=21; 75.1) modAD(n=20; 77.0)	NC=28.3 MCI=25.1 mildAD=20.6 modAD=13.3	ICD-10 criteria; Global Deterioration Scale (GDS)	Significant overall differences between diagnostic groups on **objects, locations, actions** (p<.01). NC and MCI < mildAD and modAD.
Carlomango et al. (2005)	CT	modAD(n=12; 66) mildAD(n=9; 63) NC(n=18; 61.5) Aphasic(n=11; 56.7)	modAD=16.3 mildAD=22.2 NC=NR Aphasic=NR	McKhan et al. $$1984	**Information content** and **words/minute** differentiated NC from other groups (p<.03), but did not differentiate among severity of AD or Aphasics
Chapman et al. (1995)	NRP	NC(n=12; 65.7) OE(n=12; 81.9) AD(n=12; 67.5)	NC=29.0 OE=26.1 AD=22.4	NINCDS-ADRA	AD <NC and OE on discourse **coherence** (p=.03)
Croisile et al. (1996)	CT	NC(n=24; 68.2) AD(n=22; 70.7)	NC= not reported AD=18.9	NINCDS-ADRA	AD < NC on all **word categories**; AD < NC **information units**; AD > NC on **word finding difficulties** (p<.01)
Cuetos et al. (2007)	CT	PSN1-(n=19; 45.3) PSN1+(n=21; 43.2)	PSN1-=28.8 PSN1+=27.8	Pre-clinical; genetic testing	PSN1+ < PSN1- on **semantic units** and **objective situations** (p=.001). No sig. differences on measures of fluency or syntactic complexity.

Authors	*Stimulus Picture Used*	*Groups + sample sizes + mean age*	*MMSE Score*	*Diagnostic Criteria*	*Significant Language Variables*
de Lira et al. (2014)	CT	NC(n=20; 71.1) mildAD(n=15; 68.3) modAD(n=11; 75.7)	NC=27.6 mildAD=23.7 modAD=15.7	NINCDS-ADRA; Brucki criteria	mildAD/modAD < NC on **information units** and **number of words** (p<.0001). NC>mildAD>modAD on **information units** (p<.0001)
Duong et al. (2005)	Bank Robbery	NC(n=53; 73.8) AD(n=46; 74.3)	NR	NINCDS-ADRA; GDS; Reisberg scale	Cluster analysis of discourse patterns accurately classified 41% of AD patients; authors conclude **discourse is too heterogeneous** to be an accurate classifier
Ehrlich et al. (1997)	CT and Experimental Pictures (simple vs complex)	NC(n=16; 73.2) AD(n=16; 74.3)	NC=28.7 AD=19.4	NINCDS-ADRA	AD group produced more **content units** in the simplified pictures than the complex (p<.034) (group X content interaction).
Forbes et al. (2002)	CT (simple); Tripping Woman (simple); Traffic Chaos(complex); Bus Stop(complex)	NC(n=22; 78.2) minAD(n=11; 73) mildAD(n=11; 78)	NC=NR minAD=26.6 mildAD=21.5	NINCDS-ADRA	minAD>controls **word-finding delays**, **semantic paraphasias** (p<.01). minAD < controls on **repairing errors** and **pictorial themes.** Complex pictures were more sensitive than simple pictures.
Forbes-McKay et al. (2005)	CT (simple); Tripping Woman (simple); Traffic Chaos(complex); Bus Stop(complex)	Normative Sample(n=240;18-90) minAD(n=10; 74) mildAD(n=10;78.7) modAD(n=10;74.4)	NormS=NR; minAD=24–30 mildAD=19–23 modAD=12–18	NINCDS-ADRA	Age, sex and education had significant effects on measures of melodic line, response to word finding, information content, and pictorial themes (p<.05) in the normative study. Cut-off scores for **information content, word-finding, semantic paraphasias, error monitoring and pictorial themes** correctly classified 80% of AD patients and 99% of controls.
Forbes-McKay et al. (2013)	CT (simple); Tripping Woman (simple); Traffic Chaos(complex); Bus Stop(complex)	NC(n=30; 78.2) ADtime1(n=31;76) ADtime2 (n=15;76.5)	NC>=27 ADtime1=22.2 ADtime2 = 21.1	NINCDS-ADRA	Baseline cross-sectional: AD < NC in **grammatical form, response to word finding, error monitoring, information content** (p < .001). Longitudinal: at 12 months, AD produced significantly more **phonological paraphasias** (p<.05).
Fraser et al. (2015)	CT	NC(n=233; 65.2) AD(n=240; 71.8)	NC=29.1 AD=18.5	Record review, physical assessment, neuropsychological assessment; criteria not identified	4 latent variables that classified AD speech identified from machine learning/factor analysis: **syntax/fluency**, **semantics, acoustic differences, information content**.
Giles et al. (1996)	CT	NC(n=18; 68.4) minAD(n=16; 72.8) mildAD(n=15; 66.6) modAD(n=17; 63.3)	NC=29.1 minAD=25.8 mildAD=21 modAD=10.3	NINCDS-ADRDA; CDR; MMSE;	NC > minimal=mild>moderate on **information units** (p<.0001); NC=min=mild>moderate (p<.01)
Groves-Wright et al. (2004)	CT	NC(n=14; >45) mildAD(n=14; >45) modAD(n=14; >45)	NC= >24 mildAD=NR modAD=NR	NINCDS	ModAD <mildAD=NC (p<.01) on total sum of discourse measures.
Hier et al. (1985)	CT	NC(n=15; >59) AD(n=26; NR) vasD(n=13; NR)	NR	CT scans, clinical exclusions	AD/vasD < NC on **total words, unique words, MLU, subordinate clauses, relevant observations** (p<.05)

Authors	*Stimulus Picture Used*	*Groups + sample sizes + mean age*	*MMSE Score*	*Diagnostic Criteria*	*Significant Language Variables*
Kavé et al. (2003)	CT	NC(n=48; 74.5) AD(n=14; 76.6)	NC=29.02 AD=21.8	NINCDS-ADRDA	AD < NC in **information units, circumlocutionary comments, pronouns/nouns+pronouns, words per clause** (p<.05). AD >NC in **semantic errors** (p<.05)
Kavé et al. (2016)	CT	NC(n=20; 76.2) AD(n=20; 76.1)	NC=29 AD=22	NINCDS-ADRDA	AD<NC **%content words** and **%nouns**; AD>NC **%pronouns** (p<.05)
Koing et al (2015)	NR	NC(n=15; 72) MCI(n=23; 73) AD(n=26; 80)	NC=29 MCI=26 AD=19)	NINCDS-ADRDA	Vocal feature analysis yield high classification accuracy ≥81%
Lai et al. (2009)	CT	NC(n=32; 69) AD(n=30; 72)	CDR NC=0 CDR AD=1.3	CDR	AD<NC on information units and independent clauses; AD>NC on circumlocutions and semantic errors; no differences in syntax
McNamara et al. (1992)	CT	NC(n=141; groups from 30s to 70s) AD(n=15; 65.1) PD(n=22; 61.3)	NR	NINCDS-ADRDA	AD > PD > NC on **undetected errors** (p<.05). AD < PD < NC on **repairs/revisions** (p=.0001)
Mueller et al (2016)	CT	NC(n=39; 63.1) pMCI(n=39; 63.1)	NR	NR – preclinical MCI	Psychometric MCI = pMCI < AD on **semantic units** and **unique words** (p<.05)
Murray (2010)	NRP	NC (n=14; 73.5) DEP (n=18; 73.8) AD(n=17; 75.9)	NR; Dementia Rating Score: NC: 139.43 DEP: 135.7 AD: 115.8	Depression diagnosed by psychiatrist DSM-IV criteria; NINCDS-ADRDA	AD < DEP=NC for % **correct information units**; AD > DEP=NC for **performance deviations per minute and proportion of uninformative utterances** (p<.01).
Nicholas et al. (1985)	CT	NC (n=30; 62.9) AD(n=19;66.5) WernikeA(n=16; 61) AnomicA(n=8; 57.6)	NR	DSM-III criteria for primary progressive dementia	NC > AD and WA (p<.001) on content elements. AD > NC on **deictic terms, semantic paraphasias, pronouns without antecedents, repetitions** (p<.05). NS differences between AD and AN; AD < WA (p <.05) on neologisms and literal and unrelated verbal paraphasias.
Rentoumi et al. (2014)	CT	mixAD (n=18; 75.1) pureAD (n=18; 74.3)	mixAD=18.6 pureAD=21.1	Pathological diagnosis confirmed post-mortem	With Correlation-based Feature Selection (CFS) and machine learning, CFS accurately classified mixed dementia from pure AD with 75% accuracy, based on syntactic complexity (reduced in the mixed AD) and lexical variation.
Ripich et al. (2000)	CT	earlyAD (n=10;70.9) midAD (n=10; 78) lateAD (n=10; 77.6)	earlyAD=23 midAD=16.2 lateAD=1-.3	NINCDS-ADRDA	6-month f/u longitudinal analysis: **less information units, less concise, smaller MLU (p < .05)**. Only late AD showed decreased information units over 12 months.
Sajjadi et al. (2012)	Picture from Dutch version of CAT (not described in paper)	NC (n=30; 67.5) AD (n=20; 68) SD (n=16; 67	NC=29.2 AD=22.5 SD=23	NINCDS-ADRDA; criteria according to Hodges & Patterson	AD=SD<NC in **speaking rate** (p<.001). AD >SD=NC in **phonological errors, hesitation markers, verb agreement**

Authors	*Stimulus Picture Used*	*Groups + sample sizes + mean age*	*MMSE Score*	*Diagnostic Criteria*	*Significant Language Variables*
					errors. AD & SD produced higher **proportions of closed-class words, more redundant words and phrases, and less percentage of pictorial themes.** Compared tasks of interview format and picture description; picture description exposed semantic differences, while interview exposed syntactic differences.
Shimada et al. (1998)	"The Picture Description Task"*	NC=17; 73 AD=23; 74	NC=29 AD=17.5	NINCDS-ADRDA	AD < NC in **information units** and **information units/total words** (p<.01); correlations between information units and rCBF** in occipital lobes and thalamus (p<.01)
Smith et al. (1989)	Pictures from the WAB	NC(n=18; 80) AD(n=18; 82.5)	NR 5.6	DSM and clinical exclusionary criteria	AD < NC **content units per minute, elements per clause**. AD > NC **duration of sample, number of total syllables**.
Tomoeda et al. (1996)	NRP	NC(n=52; 72.2) MCI(n=5; 73.3) modAD(n=32; 74.9) modsvAD(n=31; 78.2)	CN=29.2 MCI=21.4 modAD=17.8 modsvAD= 10.3	NINCDS-ADRDA	NC>MCI>modAD>modsvAD on **information units and conciseness** (information units/total words; p <.0001)
Vuorinin et al. (2000)	CT	VasD (n=10; 68.4) AD (n=13; 66.6)	VasD=20.7 AD=18.2	NINCDS-ADRDA	AD=vasD < NC on **pictorial themes** (p = .001)

Abbreviations: AD=Alzheimer's Disease; preAD=preclinical Alzheimer's Disease; MCI=Mild Cognitive Impairment; NC=normal controls; OE=Older Elderly; CT=Cookie Theft; NR=Not Reported; NRP=Norman Rockwell prints; CDR=Clinical Dementia Rating; NINCDS-ADRDA: National Institute for Neurological and Communication Disorders and Stroke-Alzheimer's Disease Related Disorders Association[88, 89]; WAB=Western Aphasia Battery[48]; BDAE=Boston Diagnostic Aphasia Examination[45]; PSN1-/+=E280A mutation in Presenilin 1 gene noncarrier/carrier; vasD=Vascular Dementia; PD=Parkinson Disease; DEP=Depressed; WA=Wernike's Aphasia; AN=Anomic Aphasia; SD=Semantic Dementia; PSP-RS=Richardson's Syndrome of Progressive Supranuclear Palsy; nfvPPA=nonfluent variant of Primary Progressive Aphasia; lvPPA=logopenic variant Primary Progressive Aphasia; svPPA=semantic variant Primary Progressive Aphasia

* The Picture Description Task (Sasanuma et al., $$1985)

** rCBF=regional cerebral blood flow using SPECT.

Table 2.

Key findings from literature review

Measures
• Measures of semantic content were widely used across studies, most often in some form of pre-defined content "units." There was evidence of reduced content across multiple stages (pre-symptomatic, MCI, mild to severe AD).
• Accruing evidence of simplification of syntax even in MCI, which refutes the common notion that syntax is preserved in early stages
• Picture description measures accurately distinguish controls from adults with AD dementia, but show weak correlations with naming tasks and verbal fluency tasks, suggesting connected language may tap into additional cognitive processes
Limitations
• Studies had small sample sizes (mean n = 35), lacking in ethnic and cultural diversity
• Hearing status not reported in majority of studies
• Stages of disease were not always clearly defined
• Many studies of adults with AD dementia, few with MCI
• Few longitudinal studies
• Minimal cohesiveness in terms used to identify language constructs, thus limiting the ability to aggregate results across studies

Table 3.

Recommendations for future research based on results of the literature review.

•	More studies of connected language from picture descriptions in patients with MCI or asymptomatic preclinical AD
•	More longitudinal studies of connected language in aging and across the continuum of cognitive decline associated with AD, using larger and more ethnically and culturally diverse samples
•	Studies of connected language that examine participants with and without biomarker evidence of AD pathology
•	Studies that utilize automated methods of both analysis (e.g., machine learning, natural language processing) and transcription (i.e., automatic speech recognition) to minimize subjective scoring and bias
•	Studies that examine sex-based differences in connected language across the aging spectrum
•	Development of new picture stimuli that have cross-cultural relevance and efficacy in eliciting a comprehensive language sample