Factor Structure and Clinical Applicability of New Semantic Tasks in Alzheimer's Disease and Aphasia

Luotonen Ida[1], Karrasch Mira[2], Korpilahti Pirjo[1] & Renvall Kati[1, 3]

[1]Department of Psychology and Speech-Language Pathology, University of Turku, Turku, Finland

[2]Department of Psychology, Åbo Akademi University, Turku, Finland

[3]Department of Cognitive Science, Macquarie University, Sydney, Australia

Correspondence: ida.luotonen@utu.fi, Department of Psychology and Speech-Language Pathology, University of Turku, Assistentinkatu 7, Turku

This is an original manuscript of an article published by Taylor & Francis ***in Applied Neuropsychology: Adult*** *on 16 Oct, 2021, available online:*
https://doi.org/10.1080/23279095.2021.1986511

Factor Structure and Clinical Applicability of New Semantic Tasks in Alzheimer's Disease and Aphasia

Semantic tasks are frequently used when examining language functions in patients with acquired disorders such as Alzheimer's disease (AD) and aphasia. Little is known about the possible covariation between different types of tasks or their factor structure in healthy adults. Additionally, few studies have examined semantic task performances in different patient groups. The aims of this data-driven study were to examine the factor structure in a wide range of semantic tasks in healthy older adults, the possible differences in factor variables between healthy controls, patients with AD and patients with stroke aphasia, as well as the clinical applicability of tasks in differentiating the two patient groups from controls. Participants included 59 healthy older adults, 13 patients with AD and 14 patients with aphasia. The results indicated a four-factor solution for the semantic task variables: 1) the Semantic association factor, 2) the Time factor, 3) the Verbal factor and 4) the Synonym factor. The Verbal factor was the only distinguishing factor between the two patient groups. Three factors reliably discriminated between the controls and the AD patients, and the Verbal factor reliably discriminated between the controls and the aphasia patients. In addition, a few single task variables showed outstanding discrimination for both patient groups. This study supports the notions of semantic tasks tapping into more than one cognitive subcomponent and a more general semantic impairment in AD than in aphasia. In clinical assessment, choosing appropriate semantic tasks is crucial in order to reliably detect the characteristics of the impairment.

Keywords: semantic memory, neuropsychological assessment, Alzheimer's disease, aphasia

Introduction

Semantic memory is a part of long-term memory including knowledge of concepts and objects (Hodges, Salmon & Butters, 1992; Tulving, 1972; Snowden, 2015). Retrieving semantic information from long-term memory is crucial for several cognitive processes such as expressing both verbal and non-verbal knowledge as well as understanding different types of stimuli (e.g. words, pictures and sounds). Thus, impairments in semantic memory may lead to difficulties in everyday life. They can be caused by a range of neuropsychological disorders such as Alzheimer’s Disease (AD; e.g. Verma & Howard, 2012), stroke aphasia (e.g. Thompson, Robson, Lambon Ralph & Jefferies, 2015) and semantic dementia (SD; e.g. Corbett, Jefferies, Ehsan & Lambon Ralph, 2009).

A variety of tests and tasks has been developed to assess semantic functions and impairments, many of them have been used in both clinical and research settings. In research settings, these tasks have been utilized to study the nature and neural organization of semantic impairments (for a review see Lambon Ralph, Jefferies, Patterson & Rogers, 2017). However, we are unaware of any studies exploring the factor structure of or covariance between different semantic tasks in a healthy adult population. Only a few studies have compared patient groups directly using these tasks and discussed the detection of semantic impairment in diverse clinical populations (Chapman, Hasan, Schulz & Martin, 2020; Corbett et al., 2009; Jefferies & Lambon Ralph, 2006). Therefore, there is insufficient knowledge about both the factor structure of semantic tasks and how sensitive and specific different types of tasks are in identifying semantic impairments in different clinical populations.

Prior research has thoroughly investigated different types of individual tasks that have been developed separately, based on different theoretical frameworks and using different patient samples. Generally, the controls in these studies (i.e. healthy adults) perform similarly in different semantic tasks (Catricalà et al., 2013); this is, at least partly, because of the ceiling effects (Moreno-Martínez & Rodríguez-Rojo, 2015; Ohman, Sheppard, Monetta & Taler, 2020). In patient populations, studies have aimed for a description of the skill structure underlying semantic impairment. For example, Tchakoute, Sainani and Henderson (2017) found evidence for a semantic-lexical retrieval factor and a lexical search factor in AD. Even so, the factor structure of semantic tasks in a healthy adult population has not been applied to patient studies, and clinical discrimination between controls and patients has not been established at a factor level. In addition, interpreting and comparing the results is uncertain because of the use of different sets of semantic tasks that are often limited to specific task types.

In patient populations, only a few studies have explored the correlation of different semantic tasks. In semantic type aphasia, correlations between semantic task types sharing similar cognitive demands have been identified (Corbett et al., 2009; Jefferies & Lambon Ralph, 2006). In AD, Tchakoute et al. (2017) demonstrated that most of the different types of semantic tasks correlate highly or moderately. The semantic disorder is considered to be the most explicit in SD leading to strong correlations between semantic tasks (Chapman et al., 2020; Corbett et al., 2009; Jefferies & Lambon Ralph, 2006).

Semantic impairments are commonly caused by damage to the anterior temporal lobes but damage to distributed modality-specific cortices may also lead to semantic impairment (Lambon Ralph, Jefferies, Patterson & Rogers, 2017). Damage to distributed brain areas in different patient groups may be represented by a divergence in semantic task performance. Thus, semantic impairment may be undetected if specific neuropsychological tasks are not used. In the following, the most common semantic task types are presented to clarify the processes required for successful performance in the tasks.

First, a frequently used task in assessing semantic impairment is a confrontation naming task because anomia is a typical symptom of semantic impairment (e.g. Jefferies & Lambon Ralph, 2006; Mason-Baughman & Wallace, 2013; Reilly, Peelle, Antonucci & Grossman, 2011). Nevertheless, it is challenging to define whether anomia arises from the deterioration of the semantic system or from deficits in other functions needed in speech production (Laine & Martin, 2006). One of the most famous picture naming tasks, BNT (Boston Naming Test; Kaplan, Goodglass, & Weintraub, 1983) is based on the perception that patients producing lexical-phonemic or semantic errors could have a corresponding disorder. In contrast, Rohrer et al. (2008) argue that many patients producing a collection of different naming errors and specific errors can also occur as a result of different underlying impairments.

Second, semantic verbal fluency (SVF; also called category fluency) is used to assess semantic processing but a poor performance may also stem from an impairment in executive functions (e.g. Reverberi, Cherubini, Baldinelli, & Luzzi, 2014; Troyer, Moscovitch & Winocur, 1997; Whiteside et al., 2016). In SVF, items belonging to a specific semantic category (e.g. animals) are produced in a fixed time frame, typically one minute (Strauss, Sherman & Spreen, 2006a). SVF tasks are often included in neuropsychological assessment (Rabin, Barr & Burton, 2005), and they are sensitive to cognitive impairment in many diseases, such as AD (Verma & Howard, 2012). Considering the confrontation naming and SVF, it is noteworthy that deficits in these tasks may also be due to impairments in speech production.

Third, semantic association tasks are used in assessing the integrity of semantic knowledge, for example the Pyramids and Palm Trees test (PPT; Howard & Patterson, 1992) and the Camel and Cactus Test (CCT; Bozeat, Lambon Ralph, Patterson, Garrard & Hodges, 2000). The PPT includes different types of associations with two response choices from the same category (Howard & Patterson, 1992). In the CCT, there are four same-category response choices and it was developed in order to create a more sensitive task than the PPT (Bozeat et al., 2000). Theoretically, the PPT and the CCT are based on the common view that the representations within semantic memory are organized into a network of associations sharing similar features (e.g. Snowden, 2015). Thus, semantic impairments cause difficulties understanding the connections between different concepts. Other types of tasks to assess the integrity of semantic knowledge are odd-one-out tasks (Westfall & Lee, 2021) and category judgement/sorting tasks (Adlam, Patterson, Bozeat & Hodges, 2010). In the picture versions of semantic association, odd-one-out and category judgement tasks, visual impairment may also underlie poor performance. In these tasks (including the PPT and the CCT), there are often two versions of the tasks: the items are presented as either pictures (non-verbal versions) or words (verbal versions). In the word versions, non-semantic language deficits may cause difficulties.

Fourth, word comprehension assessment often includes spoken and written word-picture matching (WPM; e.g. Cole-Virtue & Nickels, 2004;). WPM tasks are included in several assessment batteries such as the Cambridge Semantic Memory Battery (CSM; Adlam et al., 2010), the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA; Kay, Coltheart & Lesser, 1992) and the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1972). The CSM is a collection of tasks using a set of 64 items from six subcategories assuming the network structure of semantic memory. The PALPA is based on an assumption of modular structure of language processing where impairments in modules or in routes between them can be discriminated. The BDAE is a diagnostic tool for assessing aphasia and does not provide in depth information of specific components of language processing. Overall, difficulties in WPM tasks may be caused by deficits in executive function (multiple response choices) or in processing spoken or written input in addition to semantic impairments. This notion needs to also be considered in the association and odd-one-out tasks discussed above.

Fifth, for assessing word comprehension, synonym and category judgement tasks can be used. An example of a synonym judgement is a task where a synonym pair has to be defined out of three or more words (Martin, Schwartz & Kohen, 2006; Jefferies et al., 2009). Theoretically, the representations of synonyms can be thought to stand close to each other in the hierarchical structure of semantic memory (e.g. Snowden, 2015). The integrity of the semantic network should yield to the ability of connecting synonymic words. The synonym judgement tasks are only applicable in a verbal format. However, abstract words can also be assessed and thus, the synonym judgement tasks are thought to be more sensitive than some other task types that do not require speech production.

Finally, there are some types of semantic tasks used mostly in experimental study designs, such as a noun and phrase identification task (Mason-Baughman & Wallace, 2013) and a semantic feature verification task (Antonucci, 2014). Many of these tasks are considered to be more difficult as multiple cognitive processes are needed for a successful performance. Thus, the tasks presented earlier in the text are considered to be closer to “a pure semantic task”. However, all the presented tasks offer a limited assessment of semantic memory because of the nature of semantic knowledge and therefore, multiple tasks are needed (Callahan et al., 2010; Ohman et al., 2020).

To summarize, although different semantic tasks are widely used in the research literature, to our knowledge, the factor structure has not been investigated in a wide range of semantic tasks in a healthy population. Research generally confirms that semantic impairments are present in AD (Verma & Howard, 2012) and in semantic aphasia (Thompson et al., 2015), and previous research suggests that semantic impairments are qualitatively divergent in different patient populations (e.g. Reilly et al., 2011). There are many different types of semantic tasks used in assessment but a more comprehensive view of the most sensitive tasks in detecting semantic impairment in different diseases is needed. In clinical settings, the current task batteries for assessing semantic memory functions are culture-specific, they do not assess multiple aspects of semantic function and they are time-consuming to administer (e.g. CSM; Adlam et al., 2010; Italian battery for the assessment of semantic memory; Catricalà et al., 2013; The Nombela 2.0 Semantic Battery; Moreno-Martínez & Rodríguez-Rojo, 2015).

As semantic tasks currently used for clinical assessment have limitations, we developed new semantic tasks to study the semantic function of healthy older adults, patients with stroke aphasia and patients with AD. The aims of this data-driven study were (1) to examine the factor structure in a wide range of semantic tasks in healthy older adults, (2) to examine the possible differences between healthy controls, patients with AD and patients with stroke aphasia in factor variables, and (3) to assess the clinical applicability of factor variables and tasks in differentiating the two patient groups from healthy controls.

Methods

Participants

Three groups of older adults took part in the study: healthy older adults (n = 59, 33 female), patients with AD (n = 13, 6 female) and patients with stroke aphasia (n = 14, 8 female). All participants volunteered to take part in the study. An informed consent was obtained from each participant before any study procedures. In addition to the participant's consent, the patient's closest proxy gave their informed consent in the cases of AD. The study was approved by the Ethics Committee of the Hospital District of Southwest Finland.

Healthy participants were recruited from activity groups of retired people. Patients with AD and aphasia were recruited from public healthcare, adult day-care centers and from dementia and aphasia associations. Participants completed background questionnaire and were interviewed to determine their eligibility for the study. For the patient groups, the diagnosis was verified from their medical records. Exclusion criteria for all groups included: (a) significant loss of hearing and/or vision, (b) history of neurological disease (other than AD or stroke for the patient groups) or dyslexia, and (c) mother tongue other than Finnish.

Healthy participants were screened using the Mini-Mental State Examination and a score 28–30 was required for participation (MMSE; Folstein, Folstein & McHugh, 1975). In addition, participants reporting atypical subjective memory symptoms in the background questionnaire and interview were excluded. Patients with AD were required to have a minimum of 18 points on the MMSE and to be community dwelling in order to ensure sufficient cognitive capacity. MMSE was included in a more extensive test battery in patients with AD; the test conducted were: the Trail Making Test (TMT; see e.g. Strauss, Sherman & Spreen, 2006b), the memo-BNT (Karrasch et al., 2010), the CERAD Word List Memory, and the CERAD Word List Delayed Recall (Welsh et al., 1994). Patients with aphasia had a left hemisphere stroke diagnosis. The severity of aphasia symptoms was determined using the Western Aphasia Battery (WAB; Kertesz, 1982). Patients with very severe aphasia were excluded from the study (WAB Aphasia Quotient [AQ] >30).

According to the original research plan, we aimed to recruit more participants into the AD group. Because of the COVID-19 situation, the recruitment had to be suspended in March 2020. The small number of AD patients and the diversity of background variables in the three study groups complicated the matching of the groups. We matched the groups for educational background and thus, the group of aphasic patients in this study is also relatively small. We also aimed to match the three groups for age but the AD group is older than the two other groups (see Table 1). However, semantic processing is considered to be preserved in aging (Toepper, 2017). Demographic data and performance in MMSE and WAB are presented in Table 1. For MMSE, there was a statistically significant difference between the healthy and the AD group in MMSE (Kruskal-Wallis test: $\chi 2= 32.99$, $p < .001$, $df = 1$, $\eta 2=.457$).

Table 1

Demographic Characteristics of the Three Study Groups and Group Comparisons for Background Variables

	Healthy group (*n*=59) *M* (*SD*)	AD group (*n*=13) *M* (*SD*)	Aphasia group (*n*=14) *M* (*SD*)	*df*	F	partial η2	*p*	Healthy vs. AD *p*	Healthy vs. aphasia *p*	AD vs. aphasia *p*
Age (yrs)	69.0 (6.2)	74.3 (6.1)	68.9 (4.7)	2	4.37	.095	.016	.013	1.000	.055
Education (yrs)	13.0 (4.1)	11.4 (3.5)	12.5 (2.7)	2	.93	.022	.400	.371	.907	.735
MMSE	28.9 (0.9)	24.2 (2.13)	-	-	-	-	-	-	-	-
WAB AQ	-	-	79.5 (17.2)	-	-	-	-	-	-	-

Note. For age and education, groups were compared using one-way ANOVAS.

Materials and Procedure

The participants were tested in a quiet environment at different locations: at the Department of Psychology and Speech-Language Pathology at the University of Turku, in the participant’s home, or in an adult day-care center. The background questionnaire was collected and the participants were interviewed in order to assess their neurological anamnesis and suitability for the study.

All tasks were carried out during two to four 60 to 90-minute sessions depending on the subjects’ characteristics and health related factors. The task instructions were carefully detailed as the tasks were administered by doctoral and master level students in speech pathology. The tasks and their administration order for each study group are provided in Table 2. The subjects did not receive any feedback during the sessions.

Table 2

Tasks in the Administration Order

	Healthy group	AD group	Aphasia group
Tasks	1. MMSE	1. MMSE	1. WAB
	2. Semantic fluency tasks	2. CERAD / Word List	2. BNT
	3. BNT	3. Trail Making Test	3. Semantic fluency tasks
	4. New semantic task battery[a]	4. CERAD / Word List Delayed	4. New semantic task battery[a]
		5. Memo-BNT	
		6. Semantic fluency tasks	
		7. New semantic task battery[a]	

[a] Within the new semantic task battery the administration order of the tasks was pseudorandomized.

Semantic Verbal Fluency Tasks

In a semantic verbal fluency (SVF) task, words belonging to a specific semantic category (e.g. animals, fruits, tools) are produced typically in a 60-second time span (e.g. Strauss, Sherman & Spreen, 2006a). In this study, the instruction in the task was given as follows: “I ask you to name as many words as possible belonging to a specific category in one minute. First, we will do a practice task. Name as many kitchen utensils as possible. Begin.” In the practice task, if the participant was not able to produce any words or made mistakes, the instructor encouraged the participant and gave examples belonging to the category. After practicing, the instruction was given as follows: “Do you have any questions? We will start the task. Name as many animals/clothes as possible. Begin.” In this study, animals and clothes were used as categories. The order of the two categories was randomized for the participants. The number of correct items were counted. Intrusions, repetitions, proper names, paraphasias (the meaning of the word is unclear) and grammatical variations were not accepted (See Lehtinen et al., 2021).

Boston Naming Test

The Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983; in Finnish Laine, Koivuselkä-Sallinen, Hänninen & Niemi, 1997) is a 60-item visual confrontation naming task using line drawings of objects as stimuli. For the purposes of this study, a score for the advanced level BNT (starting from item 30) was calculated leading to maximum score of 31. Participants were required to respond within 20 seconds. After that, phonemic or semantic cues were given if necessary. The total score was calculated by the number of correct responses and the correct responses produced after semantic cues.

New Semantic Task Battery

We created nine new semantic memory tasks (Luotonen & Renvall, unpublished) for semantic comprehension, using both picture and word stimuli to obtain a deeper understanding of the effect of the number of stimuli in semantic processing. The picture stimuli were photographs obtained from the Bank of Standardized Stimuli (BOSS; Brodeur, Dionne-Dostie, Montreuil & Lepage, 2010). Visually ambiguous photographs were avoided. The two task types (Semantic Association tasks and Category Judgement tasks) included tasks with picture and word stimuli. The picture versions of the task were first created and then the tasks were converted into written words. As the primary stimuli were pictures, we were not able to control for word length in the verbal tasks. In addition to the new tasks, the semantic task battery includes the formerly created Synonym Judgement tasks (Renvall, unpublished). Basic information of the tasks is described below and, for the sake of brevity, further details and examples are represented in Appendix A.

Within the new semantic task battery, the tasks were administered in pseudorandomized order, thus the same task type (e.g. Semantic Association task) did not appear one after the other. Every semantic task began with two to four practice items to ensure the subject understood the instructions of the task. There was no time limit to complete the semantic tasks.

Semantic Association tasks (Luotonen & Renvall, unpublished). The Semantic Association tasks consisted of 60 items. Semantic Association tasks '1+2 pictures' and '1+2 words' contained a target picture/word on the top of two response choices. Semantic Association tasks '1+5 pictures' and '1+5 words' contained a target picture/word on the top of five response choices. Participants were asked to point to the response choice that best matched to the target picture/word.

Odd-One-Out task (Luotonen & Renvall, unpublished). The 80-item Odd-One-Out task consisted of four 20-item sections: three pictures, four pictures, five pictures and six pictures. For each item, participants were asked to point to that picture.

Word-Picture Matching tasks (Luotonen & Renvall, unpublished). The 80-item Word-Picture Matching tasks consisted of four 20-item sections: three pictures, four pictures, five pictures and six pictures. For each item, participants were asked to point to the picture that matches to the spoken word (Spoken Word-Picture Matching task) or to the written word (Written Word-Picture Matching task).

Category Judgement tasks (Luotonen & Renvall, unpublished). In the 72-item Category Judgement tasks, items were presented on cards. The stimuli were either pictures (Category Judgement task / pictures) or words (Category Judgement task / words). The tasks consisted of three sections: First, all 72 items were sorted into two semantic categories (living and man-made). Second, the 36 living items were sorted into four semantic subcategories (fruits, vegetables, mammals and birds). Third, the 36 non-living items were sorted into four semantic subcategories (tools, household items, transportation, clothes). Participants were asked to place the cards into the right category.

Synonym Judgement tasks (Renvall, unpublished). The 80-item Synonym Judgement tasks consisted of a word pair for each item, the word pairs being either synonyms or non-synonyms. The words were controlled for imageability and familiarity. The participant were asked to decide whether the two words are synonyms or not.

Analysis

All statistical analyses were performed with the R software (R Core Team, 2019) with packages psych (Revelle, 2020), GPArotation (Bernaards & Jennrich, 2014), rstatix (Kassambara, 2021), REdas (Maier, 2015), heplots (Fox, Friendly, Monette & Chalmers, 2021) and cutpointr (Thiele, 2021). For the semantic verbal fluency tasks, the BNT and the tasks in the Semantic task battery, number of correct items was used as variable. In addition, for the tasks in the Semantic task battery task completion time was used as variable. For the task completion times, we used converted scores (2500 - "task completion time in seconds") as a shorter time indicates better performance in tasks. For the converted scores, 2500 was chosen as it was the first round number exceeding the poorest performance in the data.

Principal components analysis (PCA) with an Oblique rotation was used to explore possible clustering pattern of the different semantic measures on the non-clinical subjects' sample. For this purpose, we calculated Z-scores for all variables centering the scores to the healthy sample mean. The feasibility of the PCA data was viewed by Bartlett's test of sphericity (p <.05) and the Kaiser-Meyer-Olkin (KMO) Test (values >.70). The number of components was determined using the point of inflection in scree plot and eigenvalues over the Kaiser's criterion of 1. For each component, variables with factor loadings < .40 were selected. Using this variable selection, we created mean sum scores for all the factors, that is all the variables loading on a specific factor were summed and divided by the number of variables.

We compared the performance between healthy older adults, patients with AD and patients with aphasia based on the clustering of the tasks and thus, the factor scores were used. The data were analyzed using Analysis of Covariance (ANCOVA) controlling for age and education. For the effect size, the partial eta squared (partial $\eta 2$) was calculated. For post-hoc comparisons, a Tukey HSD test was used.

The receiver operating characteristic (ROC) analysis was performed for AD versus healthy older adults and for aphasia versus healthy older adults. First, we used the factor scores to determine which factor is the best in discriminating clinical cases from non-clinical cases. Second, we used the raw scores of all task variables to discover which task within each factor was the best in discriminating clinical cases from non-clinical cases. We chose to use the raw scores (number of correct items and converted task completion times) instead of centered scores in the ROC analysis for clinical relevance.

We calculated the values of the area under the curve (AUC) to evaluate discrimination of clinical cases from non-clinical cases in all factors and separate variables. Furthermore, we calculated five indicators of test performance: (1) sensitivity (the likelihood of true positives), (2) specificity (the likelihood of false negatives), (3) Youden's index, (4) positive predictive value (PPV; the probability that the disorder is present when the test is positive), and (5) negative predictive value (NPV; the probability that the disorder is not present when the test is negative). In addition, we determined a cut-off point score for which both sensitivity and specificity are maximal using the Youden's index.

Results

Principal Components Analysis

Principal components analysis (PCA) with an Oblique rotation was used for component extraction, allowing the components to correlate in the healthy older adults' sample. All variables were converted into Z-scores prior to the analysis to allow a comparison of the variables. We needed to exclude the following seven variables in order to run the PCA as they had Kaiser-Meyer-Olkin (KMO) values lower than .70: Semantic verbal fluency / clothes, Category Judgement task pictures: score variable, Category Judgement task pictures: time variable, Category Judgement task words: score variable, Category Judgement task words: time variable, Spoken Word-Picture Matching task: score variable, and Written Word-Picture Matching task: score variable. After excluding these variables, the KMO measure (KMO = .81) verified the sampling adequacy for the remaining 17 variables, and the Bartlett's test of sphericity $\chi 2$ (136) = 590,305, $p < .001$ indicated that the correlations between items were adequate for PCA. The determinant of the correlation matrix was .000011 showing no problems with the multicollinearity.

Table 3

Summary of the Principal Components Analysis Results for the Semantic Tasks

Item	Oblique rotated component loadings				h^2	u^2
	C1: Semantic association factor	C2: Time factor	C3: Verbal factor	C4: Synonym factor		
Semantic Association task 1+2 pictures	**.88**	-.04	.18	-.15	0.78	0.22
Semantic Association task 1+5 pictures	**.85**	.14	-.36	.13	0.81	0.19
Odd-One-Out task	**.80**	.02	.07	.13	0.79	0.21
Semantic Association task 1+2 words	**.76**	-.06	.11	.01	0.63	0.37
Semantic Association task 1+5 words	**.51**	-.07	.35	.29	0.71	0.29
Spoken Word-Picture matching / Time	-.13	**.92**	-.09	-.07	0.78	0.22
Semantic Association task 1+5 pictures / Time	.25	**.78**	-.13	.11	0.73	0.27
Odd-One-Out task / Time	.07	**.73**	.14	-.15	0.60	0.40
Written Word-Picture matching / Time	-.03	**.67**	.32	-.07	0.66	0.34
Semantic Association task 1+2 pictures / Time	-.05	**.58**	.27	.28	0.64	0.36
Semantic Association task 1+2 words / Time	.05	**.57**	.26	.21	0.63	0.37
Semantic Association task 1+5 words / Time	.02	.21	**.82**	.00	0.83	0.17
Written Synonym Judgement task / Time	.10	.12	**.64**	.31	0.74	0.26
Boston Naming Test	**.47**	.02	**.58**	-.02	0.69	0.31
Semantic verbal fluency (animals)	.06	-.02	**.47**	.23	0.35	0.65
Spoken Synonym Judgement task	.00	.01	0.07	**.88**	0.82	0.18
Written Synonym Judgement task	.01	-.05	-0.05	**.88**	0.76	0.24
Eigenvalues	6.73	2.73	1.30	1.18		
% of variance	21.13	20.18	15.55	13.38		

Note. Component loadings over .40 appear in bold. C1 = Component 1; C2 = Component 2; C3 = Component 3; C4 = Component 4

An initial analysis was run to obtain eigenvalues for each component in the data. Four components had eigenvalues over the Kaiser's criterion of 1 and they explained 70 % of the variance. Thus, four components were retained for the final analysis. Table 3 shows the component loadings after rotation. The variables loading to component 1, the Semantic association factor, were characterized by tasks that require processing of items sharing similar properties. Component 2, the Time factor, included variables of task completion time for six semantic tasks. Component 3, the Verbal factor, consisted of two tasks that require word finding and speech production and time variables for the two verbal semantic tasks. Component 4, the Synonym factor, consisted of the two Synonym Judgement tasks.

Group Comparisons

A One-way Analysis of Covariance was conducted to examine whether the factor scores differed between healthy older adults, patients with AD and patients with aphasia controlling for age and education. Table 4 shows that significant differences between the three groups were found in all four factors. Age was a significant covariate for the Semantic Association factor and the Time factor. Education was a significant covariate for the Verbal factor.

In post-hoc comparisons, the Tukey HSD test showed that a significant difference was found between patient groups in the Verbal factor (see Table 4). Healthy older adults differed from both patient groups in the Semantic association factor, the Time factor and the Verbal factor, and from aphasia group also in the Synonym factor (see Table 4). For a qualitative examination of the scores, we provide means, standard deviations, and ranges of all factor and task variables in all three study groups in Appendix B.

Table 4

ANCOVA Statistics of the Four Factors Using Age and Education as Covariates in the Three Study Groups and Post-Hoc Comparisons of the Groups

	Sum of Squares	*df*	Mean Square	*F*	*p*	Partial η2	Post-hoc comparisons
The Semantic association factor							healthy > aphasia, healthy > AD
Group	15.84	2	7.92	17.05	<.001	0.253	
Age	1.99	1	1.99	4.29	.041	0.035	
Education	1.51	1	1.51	3.24	.076	0.038	
The Time factor							healthy > AD, healthy > aphasia
Group	21.98	2	10.99	32.18	<.001	0.404	
Age	4.46	1	4.46	13.07	<.001	0.139	
Education	0.04	1	0.04	0.12	.730	0.001	
The Verbal factor							healthy > AD, healthy > aphasia, AD > aphasia
Group	30.06	2	15.03	51.13	<.001	0.540	
Age	0.50	1	0.50	1.69	.197	0.010	
Education	1.42	1	1.42	4.82	.031	0.056	
The Synonym factor							healthy > aphasia,
Group	12.13	2	6.10	9.15	<.001	0.169	
Age	2.36	1	2.36	3.56	.062	0.031	
Education	1.25	1	1.25	1.89	.173	0.023	

Note. In the post-hoc comparisons, the groups differ at the $p = .01$ level in the Tukey HSD test.

ROC Analysis

The receiver operating characteristic (ROC) analysis was performed for patients with AD versus healthy older adults and for patients with aphasia versus healthy older adults. The four factors from the PCA were used in the analysis in addition to individual task variables. One variable (BNT) had a loading over .40 for two factors. It was included only in the Verbal factor as the loading of the variable was higher than for the semantic association factor. For the factor scores, the mean of the variable Z-scores loading to the factor were used. For the individual task variables, raw scores were used to obtain clinical relevance and thus, the cut-off scores of individual task variables are reported. In this context, we considered AUC values over .90 as excellent.

In the AD group, the AUC of the Semantic association factor, the Time factor and the Verbal factor were over .90. In addition, four task variables reached an AUC value over .90.

Table 5 presents the AUC values, sensitivity, specificity, Youden's index, predictive values and cut-off scores of individual task variables from the ROC analysis for the AD group.

Table 5

ROC Analysis of the Screening Ability of Semantic Tasks for AD Patients

	AUC	Sensitivity	Specificity	Youden's index	PPV	NVP	Cut-off score
Semantic association factor[1]	**0.931**	0.847	1.000	0.847	1.000	0.591	-0.310
Semantic Association task 1+2 pictures	0.810	0.763	0.769	0.532	0.938	0.417	58
Semantic Association task 1+5 pictures	**0.925**	0.831	0.923	0.754	0.980	0.545	55
Odd-One-Out task	**0.918**	0.814	0.923	0.737	0.980	0.522	72
Semantic Association task 1+2 words	0.737	0.627	0.692	0.319	0.902	0.290	59
Semantic Association task 1+5 words	0.855	0.576	1.000	0.576	1.000	0.342	58
Time factor[1]	**0.905**	0.746	0.923	0.669	0.978	0.444	-0.146
Spoken Word-Picture Matching / Time	0.839	0.797	0.846	0.643	0.959	0.478	2197
Semantic Association task 1+5 pictures / Time	**0.928**	0.864	0.846	0.710	0.962	0.579	2116
Odd-One-Out task / Time	0.826	0.847	0.846	0.693	0.962	0.550	1934
Written Word-Picture Matching / Time	0.747	0.898	0.615	0.513	0.914	0.571	2198
Semantic Association task 1+2 pictures / Time	0.817	0.898	0.615	0.513	0.914	0.571	2219
Semantic Association task 1+2 words / Time	0.846	0.763	0.846	0.609	0.957	0.440	2256
Verbal factor[1]	**0.920**	0.949	0.923	0.872	0.982	0.800	-0.542
Semantic Association task 1+5 words / Time	**0.934**	0.847	0.923	0.770	0.980	0.571	2106
Written Synonym Judgement task / Time	0.898	0.831	0.846	0.677	0.961	0.524	2223
Boston Naming Test	0.874	0.627	1.000	0.627	1.000	0.371	27
Semantic verbal fluency / animals	0.867	0.780	0.846	0.626	0.958	0.458	19
Synonym factor[1]	0.688	0.780	0.538	0.318	0.885	0.350	-0.086
Spoken Synonym Judgement task	0.709	0.475	1.000	0.475	1.000	0.295	80
Written Synonym Judgement task	0.651	0.831	0.538	0.369	0.891	0.412	77

Note. AUC = area under the curve; PPV = positive predictive value; NVP = negative predictive value.
AUC values over .90 appear in bold.
[1] For the factor scores, the mean of the variable Z-scores loading to the factor was used.

In the aphasia group, the Verbal factor had an AUC value over .90, and the Semantic association factor and the Time factor over .80. In addition, five task variables reached an AUC value over .90. Table 6 shows the AUC values, sensitivity, specificity, predictive values and cut-off scores from the ROC analysis for the aphasia group.

Table 6

ROC Analysis for the Screening Ability of Semantic Tasks for Aphasia Patients

		AUC	Sensitivity	Specificity	Youden's index	PPV	NVP	Cut-off score
Semantic association factor[1]		0.816	0.627	0.929	0.556	0.974	0.371	-0.052
	Semantic Association task 1+2 pictures	0.719	0.763	0.643	0.406	0.900	0.391	58
	Semantic Association task 1+5 pictures	0.682	0.949	0.286	0.235	0.848	0.571	53
	Odd-One-Out task	0.703	0.814	0.571	0.385	0.889	0.421	72
	Semantic Association task 1+2 words	0.837	0.898	0.643	0.541	0.914	0.600	58
	Semantic Association task 1+5 words	0.772	0.695	0.714	0.409	0.911	0.357	57
Time factor[1]		0.898	0.627	1.000	0.627	1.000	0.389	0.006
	Spoken Word-Picture matching / Time	0.840	0.797	0.786	0.583	0.940	0.478	2197
	Semantic Association task 1+5 pictures / Time	0.809	0.814	0.714	0.528	0.923	0.476	2124
	Odd-One-Out task / Time	0.754	0.627	0.786	0.413	0.925	0.333	2038
	Written Word-Picture matching / Time	**0.909**	0.881	0.857	0.738	0.963	0.632	2203
	Semantic Association task 1+2 pictures / Time	0.848	0.729	0.929	0.658	0.977	0.448	2267
	Semantic Association task 1+2 words / Time	**0.949**	0.881	0.929	0.81	0.981	0.650	2236
Verbal factor[1]		**0.965**	0.864	1.000	0.864	1.000	0.636	-0.232
	Semantic Association task 1+5 words / Time	**0.970**	0.898	0.929	0.827	0.981	0.684	2082
	Written Synonym Judgement task / Time	**0.925**	0.831	0.857	0.688	0.961	0.545	2223
	Boston Naming Test	0.772	0.949	0.714	0.663	0.933	0.769	22
	Semantic verbal fluency / animals	**0.958**	0.881	0.929	0.81	0.981	0.650	18
Synonym factor[1]		0.760	0.763	0.714	0.477	0.918	0.417	-0.068
	Spoken Synonym Judgement task	0.775	0.475	1.000	0.475	1.000	0.311	80
	Written Synonym Judgement task	0.696	0.78	0.643	0.423	0.902	0.409	78

Note. AUC = area under the curve; PPV = positive predictive value; NVP = negative predictive value.
AUC values over .90 appear in bold.
[1] For the factor scores, the mean of the variable Z-scores loading to the factor was used.

Discussion

Different types of semantic tasks are widely used in clinical assessment of a range of neurological disorders in which the nature of the semantic impairment might vary depending on the diagnosis. A variety of semantic tasks can reveal different aspects of semantic cognition. However, little is known about the clinical relevance and sensitivity of the tasks in diverse clinical populations. In the present study, we created a battery of new semantic tasks and found four theoretically valid factors underlying the performance of healthy adults. The result provides evidence of semantic tasks tapping into many cognitive subcomponents. Examining the clinical applicability of factors and tasks in AD and aphasia, the results supported the prior evidence that demonstrates deficits in a wider set of semantic tasks in AD than in aphasia.

The four-factor solution of the semantic task variables showed the following, separate factors: the Semantic association factor, the Time factor, the Verbal factor, and the Synonym factor. The presence of these factors indicated that the cognitive subcomponents underlying a performance vary across different task types. Generally, the performance of healthy adults in semantic tasks has been associated to a somewhat steady performance across the tasks, following ceiling effects (Catricalà et al., 2013). In the semantic tasks employed in the current study, ceiling effects were not frequently found in the healthy control population, indicating that these tasks vary in difficulty and may thus be more sensitive to milder impairments. Previous studies have seldom included measures of task completion time, which emerged from the results as loading on a separate factor and could be an important aspect to measure in semantic tasks.

These four factors can be discussed in relation to the existing theoretical frameworks of semantic processing. Often, current frameworks include processing of different types of stimuli (e.g. verbal and non-verbal; see e.g. Zannino et al., 2013), provide views of the organization of semantic information (e.g. concrete and abstract; see e.g. Shallice & Cooper, 2013), or discuss the distinction between an amodal semantic ‘hub’ and modality-specific ‘spokes’ (e.g. Patterson & Lambon Ralph, 2016). As far as we know, the current frameworks do not offer a comprehensive explanation of our suggested factor structure. From a clinical point of view, this four-factor structure provides additional support to the clinical use of semantic association tasks, task completion time, naming and fluency tasks and abstract verbal tasks. In the following, we will discuss each factor in detail.

The Semantic association factor included the scores (number of correct responses) of the four Semantic association tasks and the Odd-one-out task. The result supports the role of widely tested and used semantic association tasks (e.g. Pyramids and Palm Trees test) in assessing semantic deficits (Adlam et al., 2010; Bozeat et al., 2000; Jefferies & Lambon Ralph, 2006). All tasks in the Semantic association factor likely employ the integrity of semantic knowledge as the subject is required to process items that share basic conceptual properties (e.g. shirt and trousers) while simultaneously trying to find associative links between a dissimilar item. It seems that this required process determines this factor instead of the domain of stimuli (verbal vs. non-verbal). This notion is supported by theoretical views of the general hub component of semantic memory that integrates information from different modalities (Patterson, Nestor & Rogers, 2007; Patterson & Lambon Ralph, 2016).

Variables of the task completion time of the six semantic tasks loaded to the Time factor. Only two of the time variables were not included in the Time factor but included in the Verbal factor instead, and these are discussed below. A popular explanation of the role of task completion time is that general processing speed can cause a decline in performance in a range of neuropsychological tasks (Sleimen-Malkoun, Temprado & Berton, 2013). In the healthy group, it is noteworthy that there is more variation in the task completion times than in the total scores. The participants were relatively old which increases the role of their general processing speed. In the context of semantic tasks, reaction times have been studied in experimental study designs but there is a lack of evidence as regards the role of clinically relevant task completion times.

The Verbal factor consists of the two tasks that require word finding and speech production (BNT and SVF) and time variables of the two semantic tasks that we consider requiring the longest processing times of the semantic tasks. In one task, the Semantic Association task 1+5 words, a simultaneous processing of six written words at the same time is required, and in the other task, the Written Synonym Judgement task, the word pairs include, for example, the processing of abstract words. Compared to other time variables' in the Time factor, the available evidence suggests that these two tasks require more verbal working memory capacity thus leading to a larger variation in task completion times. We suggest verbal working memory as being one of the cognitive subcomponents defining the Verbal factor (see e.g. Acheson & MacDonald, 2009 for a review of verbal working memory).

Lastly, the score variables of the two Synonym Judgement tasks clustered on the Synonym factor. As stated above, these tasks differ from the other semantic tasks in the use of abstract verbal items as the other tasks only include the processing of verbal or non-verbal concrete items. This result provides confirmatory support for the arguments of Shallice and Cooper (2013) demonstrating the possible discrimination of concrete and abstract semantic systems.

Regarding the factor structure, it is noteworthy that we needed to exclude seven variables in order to run the PCA. The ceiling effect caused the exclusion of the score variables for the Word-Picture Matching tasks. In the Category Judgement tasks, the ceiling effect was evident but in addition, some of the healthy participants categorized living items (e.g. fruits) to man-made items which lead to the exclusion of both score and time variables. In addition, the Semantic verbal fluency task with the category ‘clothes’ was excluded as the performance of healthy older adults showed large variation. The covariation of these tasks and other semantic tasks could not be analyzed and conclusions regarding the cognitive subcomponents required in the performance could not be made.

The results of the ANCOVA and ROC analysis provide evidence for the usefulness of the factor structure in clinical assessment as group differences in all four factor variables were found. In the post-hoc analysis, the AD group differed from the healthy group in the Semantic association factor, the Time factor and the Verbal factor. In addition, these three factors demonstrated outstanding discrimination for the AD group indicating a general semantic impairment in line with previous literature (e.g. Hodges et al., 1992; Verma & Howard, 2012).

The aphasia group differed from the healthy group in all four factors but only the Verbal factor demonstrated outstanding discrimination. In addition, the patient groups differed in the Verbal factor. These results provide confirmatory evidence that patients with aphasia have a more specific language impairment compared to the general semantic impairment of patients with AD, however, the semantic impairment in these two patient groups has not been compared to our knowledge (see e.g. Chapman et al., 2020; Corbett et al., 2009). The small group sizes may be the reason why we did not find other differences between the patient groups despite the differences shown in the discriminative ability of the factors in the ROC analysis. Further data collection is needed to determine whether there is a differential impairment of semantic processing in these patient groups.

In addition to the factor variables, the results of the ROC analysis showed evidence for the discriminative ability of a wide range of semantic task variables in both patient groups as AUC values over .70 are generally considered to have an acceptable ability of discrimination (Mandrekar, 2010). In the AD group, the Semantic Association task 1+5 pictures had outstanding discrimination for both the score and the time variable. Consistent with the previous studies, this result confirms the clinical relevance of the commonly used task type of semantic association for AD patients (e.g. PPT; Howard & Patterson, 1992; CCT; Bozeat et al., 2000). The number of stimuli in this new task has been increased compared to the current tasks, possibly leading to an even better discrimination of semantic deficits while at the same time increasing the role of working memory during the task performance. Interestingly, the AUC values of these variables exceeded the values of the SVF tasks which are generally considered to offer superior differentiation between patients with AD and healthy controls (see e.g. Verma & Howard, 2012). In the aphasia group, the SVF task and the time variables of the tasks requiring verbal processing (reading or producing words) showed the best discriminative ability. These results further endorse the existing evidence of verbal processing being the core deficit in aphasia. In this study, the group of aphasia patients had relatively mild deficits of comprehension and thus, we did not expect to see severe semantic deficits. Contrary to our expectations, the score variables of non-verbal tasks showed acceptable discrimination and the time variables even excellent discrimination. These results might indicate the possibility of semantic deficits in this patient group. However, the role of other cognitive skills (e.g. executive processing) and the previously suggested semantic control mechanism needs to be considered (Thompson et al., 2018).

Considering the background variables, the results of ANCOVA showed that age was a significant covariate for the Time factor. Semantic memory is broadly thought to be preserved in healthy aging (Toepper, 2017) but recent studies have expanded perspectives on the nature of semantic cognition in aging (Hoffman & Morcom, 2018). As discussed above, general processing speed might have an impact on the Time factor, and aging is related to a slower processing speed (Salthouse, 1996). Education is known to have a significant effect on performance in many neuropsychological tasks, for example semantic association tasks (Callahan et al., 2010). In this study, education did not affect the performance in semantic tasks.

Overall, the data indicates that certain semantic tasks are superior in distinguishing healthy adults from patient groups. Importantly, different tasks have dissimilar abilities in discriminating AD and aphasia patients from healthy controls. However, the patient groups were small and therefore, although we were able to match the three groups by education it was not possible to match them by age. This limitation is evidence of our difficulty in collecting data on patients with AD during the COVID-19 pandemic, but we aim to increase the group size when it is possible. As the focus of the study was on the semantic tasks, it is possible that different evaluations would have arisen if the focus had been on a wider set of neuropsychological tasks (e.g. executive functions). Another limitation is the lack of comparison of the new semantic tasks with the existing tasks (e.g. Pyramids and Palm Trees test). In order to develop the semantic task battery further, the single items could be reviewed for their level of difficulty.

When developing semantic tasks for this study, we aimed to advance the currently available clinical tasks by using standardized photographs (BOSS). A similar development has recently been made to the Camel and Cactus Test (modified CCT; Moore et al., 2020) and the Nombela 2.0 semantic battery (Moreno-Martínez & Rodríguez-Rojo, 2015). To further develop these semantic tasks for clinical use, the psychometric properties of the task battery should be studied (e.g. face validity, inter-rater reliability, test-retest reliability, internal consistency). Additionally, as suggested by Moore et al. (2020) and Ohman et al. (2020), the number of tasks and the number of items in each task should be reviewed in order to find the most effective ways to assess semantic functions in a clinical setting.

In sum, the results of the present study indicate that semantic tasks tap into many subcomponents. In this study, patients with AD had a deficit in all subcomponents but patients with stroke aphasia showed a severe deficit in the Verbal factor. From a clinical perspective, these results provide information on choosing the most efficient diagnostic tasks for assessing semantic impairment. For patients with AD, we recommend applying a semantic association task with several response choices as both score and time variables showed outstanding discrimination. For patients with aphasia, we propose the use of task completion time of semantic tasks with written words in clinical assessment. Utilizing careful task selection, the clinicians can efficiently differentiate semantic impairments.

References

Acheson, D.J., & MacDonald, M.C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135*(1), 50–68. https://doi.org/10.1037/a0014411

Adlam, A.-L.R., Patterson, K., Bozeat, S., & Hodges, J.R. (2010). The Cambridge Semantic Memory Test Battery: Detection of semantic deficits in semantic dementia and Alzheimer's disease. *Neurocase, 16,* 193–207. https://doi.org/10.1080/13554790903405693

Antonucci, S.M. (2014) What matters in semantic feature processing for persons with stroke-aphasia: Evidence from an auditory concept–feature verification task, *Aphasiology, 28*(7), 823–839. https://doi.org/10.1080/02687038.2014.913769

Bernaards, C., & Jennrich, R. (2014). *GPArotation: GPA Factor Rotation.* https://CRAN.R-project.org/package=GPArotation

Bozeat, S., Lambon Ralph, M.A., Patterson, K., Garrard, P., & Hodges, J.R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia, 38*(9), 1207–1215. https://doi.org/10.1016/s0028-3932(00)00034-8

Brodeur, M.B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE 5*(5): e10773. https://doi.org/10.1371/journal.pone.0010773

Callahan, B.L., Macoir, J., Hudon, C., Bier, N., Chouinard, N., Cossette-Harvey, M., Daigle, N., Fradette, C., Gagnon, L., & Potvin, O. (2010). Normative data for the Pyramids and Palm Trees Test in the Quebec-French population. *Archives of Clinical Neuropsychology, 25*(3), 212–217. https://doi.org/10.1093/arclin/acq013

Catricalà, E., Della Rosa, P.A., Ginex, V., Mussetti, Z., Plebani, V., & Cappa, S.F. (2013). An Italian battery for the assessment of semantic memory disorders. *Neurological Sciences, 34*, 985–993. https://doi.org/10.1007/s10072-012-1181-z

Chapman, C.A., Hasan, O., Schulz, P.E., & Martin, R.C. (2020). Evaluating the distinction between semantic knowledge and semantic access: Evidence from semantic dementia and comprehension-impaired stroke aphasia. *Psychonomic Bulletin & Review, 27*, 607–639 https://doi.org/10.3758/s13423-019-01706-6

Cole-Virtue, J., & Nickels, L. (2004). Spoken word to picture matching from PALPA: A critique and some new matched sets. *Aphasiology, 18*, 77–102. https://doi.org/10.1080/02687030344000346

Corbett, F., Jefferies, E., Ehsan, S., & Lambon Ralph, M. (2009). Different impairments of semantic cognition in semantic dementia and semantic aphasia: evidence from the non-verbal domain. *Brain, 132*(9), 2593–2608. https://doi.org/10.1093/brain/awp146

Goodglass, H., & Kaplan, E. (1972). *Boston Diagnostic Aphasia Examination.* Philadelphia, PA: Lea & Febiger.

Folstein M.F., Folstein S.E., & McHugh P.R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Fox, J., Friendly, M., Monette, G., & Chalmers, P. (2021). *heplots: Visualizing Hypothesis Tests in Multivariate Linear Models.* https://cran.r-project.org/package=heplots

Hodges, J.R., Salmon, D.P., & Butters, N. (1992). Semantic memory impairment in Alzheimer's disease: Failure of access or degraded knowledge? *Neuropsychologia, 30,* 301–314. https://doi.org/10.1016/0028-3932(92)90104-T

Hoffman, P., & Morcom, A. M. (2018). Age-related changes in the neural networks supporting semantic cognition: A meta-analysis of 47 functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews, 84,* 134–150. https://doi.org/10.1016/j.neubiorev.2017.11.010

Howard, D., & Patterson, K. (1992). *The Pyramids and Palm Trees: A test of semantic access from words and pictures.* Bury St Edmunds, UK: Thames Valley Test Company.

Jefferies, E., & Lambon Ralph, M.A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain, 129*, 2132–2147. https://doi.org/10.1093/brain/awl153

Jefferies, E., Patterson, K., Jones, R. W., & Lambon Ralph, M. A. (2009). Comprehension of concrete and abstract words in semantic dementia. *Neuropsychology, 23*(4), 492–499.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test.* Philadelphia, PA: Lea & Fibiger.

Karrasch, M., Myllyniemi, A., Latvasalo, L., Söderholm, C., Ellfolk, U., & Laine, M. (2010). The diagnostic accuracy of an incidental memory modification of the Boston Naming Test (memo-BNT) in differentiating between normal aging and mild Alzheimer's disease. The *Clinical neuropsychologist,* *24*(8), 1355–1364. https://doi.org/10.1080/13854046.2010.521982

Kassambara, A. (2021). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests.* https://cran.r-project.org/package=rstatix

Kay, J., Lesser, R., & Coltheart, M. (1992). *PALPA: Psycholinguistic Assessments of Language Processing in Aphasia.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Kertesz, A. (1982). *Western Aphasia Battery.* New York: Grune & Stratton.

Laine, M., Niemi, J., Koivuselkä-Sallinen, P. & Tuomainen, J. (1997). *Bostonin diagnostinen afasiatutkimus (BDAT)*. Helsinki: Psykologien Kustannus.

Laine, M., & Martin, N. (2006). *Anomia: theoretical and clinical aspects.* Hove: Psychology Press.

Lambon Ralph, M.A., Jefferies, E., Patterson, K., & Rogers, T.T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience, 18*, 42–55. https://doi.org/10.1038/nrn.2016.150

Lehtinen, N., Luotonen, I. & Kautto, A. (2021). Systematic administration and analysis of verbal fluency tasks: Preliminary evidence for reliable exploration of processes underlying task performance. *Applied Neuropsychology. Adult*, 1–13.

Luotonen, I., & Renvall, K. (unpublished, a). *Semantic Association tasks*. Department of Psychology and Speech-Pathology, University of Turku, Finland.

Luotonen, I., & Renvall, K. (unpublished, b). *Odd-One-Out task*. Department of Psychology and Speech-Pathology, University of Turku, Finland.

Luotonen, I., & Renvall, K. (unpublished, c). *Word-Picture Matching tasks*. Department of Psychology and Speech-Pathology, University of Turku, Finland.

Luotonen, I., & Renvall, K. (unpublished, d). *Category Judgement tasks*. Department of Psychology and Speech-Pathology, University of Turku, Finland.

Maier, M.J. (2015*). REdaS: Companion Package to the Book 'R: Einführung durch angewandte Statistik'*. https://cran.r-project.org/package=REdaS

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology,* *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Martin, N., Schwartz, M.F., & Kohen, F. (2006). Assessment of the ability to process semantic and phonological aspects of words in aphasia: A multi-measurement approach. *Aphasiology*, *20*(2), 154–166. https://doi.org/10.1080/02687030500472520

Mason-Bauhman, M., & Wallace, S.E. (2013). Semantic feature knowledge in persons with aphasia: The role of commonality, distinctiveness and importance. *Aphasiology 27*(3), 364–380. https://doi.org/10.1080/02687038.2012.730602

Moore, K., Convery, R., Bocchetta, M., Neason, M., Cash, D. M., Greaves, C., Russell, L. L., Clarke, M., Peakman, G., van Swieten, J., Jiskoot, L., Moreno, F., Barandiaran, M., Sanchez-Valle, R., Borroni, B., Laforce, R., Jr, Doré, M. C., Masellis, M., Tartaglia, M. C., Graff, C., … Genetic FTD Initiative, GENFI* (2020). A modified Camel and Cactus Test detects presymptomatic semantic impairment in genetic frontotemporal dementia within the GENFI cohort. *Applied neuropsychology: Adult*, 1–8. Advance online publication. https://doi.org/10.1080/23279095.2020.1716357

Moreno-Martínez, F.J., & Rodríguez-Rojo, I.C. (2015). The Nombela 2.0 semantic battery: an updated Spanish instrument for the study of semantic processing. *Neurocase, 21*(6), 773–785. https://doi.org/10.1080/13554794.2015.1006644

Ohman, A., Sheppard, C., Monetta, L., & Taler, V. (2020). Assessment of semantic memory in mild cognitive impairment: The psychometric properties of a novel semantic battery. *Applied Neuropsychology: Adult*, 1–7. https://doi.org/10.1080/23279095.2020.1774885

Patterson, K., & Lambon Ralph, M. (2016). The hub-and-spoke Hypothesis of semantic Memory. In G. Hickok and S.L. Small (Eds.) *Neurobiology of Language*, 765–775. Academic Press.

Patterson, K., Nestor, P.J., & Rogers, T.T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*, 976–987. https://doi.org/10.1038/nrn2277

R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Rabin, L.A., Barr, W.B., & Burton, L.A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20*, 33–65. https://doi.org/10.1016/j.acn.2004.02.005

Reilly, J., Peelle, J. E., Antonucci, S. M., & Grossman, M. (2011). Anomia as a marker of distinct semantic memory impairments in Alzheimer's disease and semantic dementia. *Neuropsychology*, *25*(4), 413–426. https://doi.org/10.1037/a0022738

Renvall, K. (unpublished). *Synonym Judgement task: Spoken and written version*. Department of Psychology and Speech-Pathology, University of Turku, Finland.

Revelle, W. (2020). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University. https://CRAN.R-project.org/package=psych

Reverberi, C., Cherubini, P., Baldinelli, S., & Luzzi, S. (2014). Semantic fluency: Cognitive basis and diagnostic performance in focal dementias and Alzheimer's disease. *Cortex, 54*, 150–164. doi://doi.org/10.1016/j.cortex.2014.02.006

Rohrer, J. D., Knight, W. D., Warren, J. E., Fox, N. C., Rossor, M. N., & Warren, J. D. (2008). Word-finding difficulty: a clinical analysis of the progressive aphasias. *Brain*, *131*, 8–38. https://doi.org/10.1093/brain/awm251

Salthouse, T.A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428. https://doi.org/10.1037/0033-295x.103.3.403

Shallice, T., & Cooper, R.P. (2013). Is there a semantic system for abstract words? *Frontiers in Human Neuroscience*, *7*(175), 1–10. https://doi.org/10.3389/fnhum.2013.00175

Sleimen-Malkoun, R., Temprado, J. J., & Berton, E. (2013). Age-related dedifferentiation of cognitive and motor slowing: insight from the comparison of Hick-Hyman and Fitts' laws. *Frontiers in Aging Neuroscience*, *5*, 62. https://doi.org/10.3389/fnagi.2013.00062

Snowden, J.S. (2015). Semantic memory. In J.D. Wright (Ed.), *International Encyclopedia of the Social and Behavioral Sciences* (pp. 572–578). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.51059-9

Strauss, E., Sherman, E., & Spreen, O. (2006a). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary (3rd ed.)* (pp. 499–525). New York: Oxford University Press.

Strauss, E., Sherman, E., & Spreen, O. (2006b). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary (3rd ed.)* (pp. 655–677). New York: Oxford University Press.

Tchakoute, C.T., Sainani, K.L., & Henderson, V.W. (2017). Semantic memory in the clinical progression of Alzheimer disease. *Cognitive and Behavioral Neurology*, 30(3), 81–89.

Thiele, C. (2021). *cutpointr: Determine and Evaluate Optimal Cutpoints in Binary Classification Tasks.* https://cran.r-project.org/package=cutpointr

Thompson, H.E., Robson, H., Lambon Ralph, M.A., & Jefferies, E. (2015). Varieties of semantic 'access' deficit in Wernicke's aphasia and semantic aphasia. *Brain, 138*(12), 3776–3792. https://doi.org/10.1093/brain/awv281

Thompson, H. E., Almaghyuli, A., Noonan, K. A., Barak, O., Lambon Ralph, M. A., & Jefferies, E. (2018). The contribution of executive control to semantic cognition: Convergent evidence from semantic aphasia and executive dysfunction. *Journal of Neuropsychology, 12*(2), 312–340. https://doi.org/10.1111/jnp.12142

Toepper, M. (2017). Dissociating normal aging from Alzheimer's disease: A view from cognitive neuroscience. *Journal of Alzheimer's Disease, 57*(2), 331–352. https://doi.org/10.3233/JAD-161099

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology, 11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138

Tulving, E. (1972) Episodic and semantic memory. In E. Tulving E. and W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.

Verma, M., & Howard, R. J. (2012). Semantic memory and language dysfunction in early Alzheimer's disease: a review. *International Journal of Geriatric Psychiatry, 27*(12), 1209–1217. https://doi.org/10.1002/gps.3766

Welsh, K. A., Butters, N., Mohs, R. C., Beekly, D., Edland, S., Fillenbaum, G., & Heyman, A. (1994). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology, 44*(4), 609–614. https://doi.org/10.1212/wnl.44.4.609

Westfall, H.A., & Lee, M.D. (2021). A model-based analysis of the impairment of semantic memory. *Psychonomic Bulletin and Review*. Online publication. https://doi.org/10.3758/s13423-020-01875-9

Whiteside, D.M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal Fluency: Language or Executive Function Measure? *Applied Neuropsychology: Adult, 23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Zannino, G.D., Perri, R., Monaco, M., Caltagirone, C., Luzzi, S., & Carlesimo, G. A. (2013). The special status of verbal knowledge in semantic memory: Evidence from performance of semantically impaired subjects on verbalizable and non-verbalizable versions of the object decision task. *Brain & Language, 128,* 9–17. https://doi.org/10.1016/j.bandl.2013.11.003

Funding

This work was financially supported by the University of Turku Graduate School wages awarded to the first author and the Finnish Association of Speech and Language Therapists.

Acknowledgments

We thank the University of Turku Speech-Language Pathology students (Lotta Behm, Reetta-Maija Henttinen, Vilma Kaajakari, Meri Kraft, Terhi Laakkonen and Juulia Vehkaoja) for their assistance in data collection. We are grateful to the participants for their effort in taking part in this study.

Disclosure of Interest

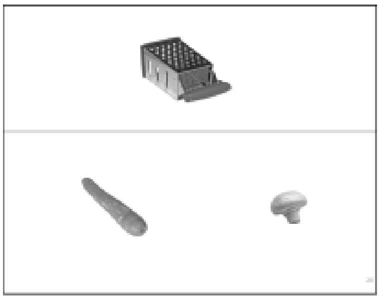
The authors do not have conflicts of interest regarding this research study.

Appendix A.
The Semantic Task Battery

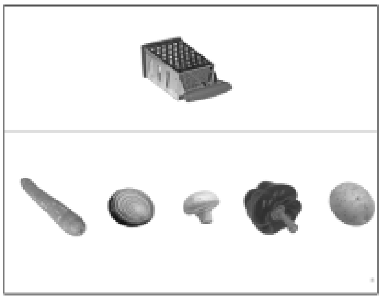
1. Semantic Association Tasks

- Task instruction: The participant is asked to decide which of the response choices could best be matched or is most closely related to the target stimulus.
- Each page in the test booklet contains the target stimulus at the top of the page and the response choices below.
- The response choices are from the same superordinate category (e.g. vegetables). The target stimulus and the response choices are semantically related (associated) but neither synonyms nor superordinate category.
- The tasks '1+2' and '1+5' differ between the number of response choices (2 vs. 5).
- The tasks 'pictures' and 'words' differ between the stimulus modality (pictures vs. written words).
- Each task consists of 60 items. The items are presented in a pseudorandomized order (the superordinate category of the response choices is never sequentially the same). The order is different in each task.

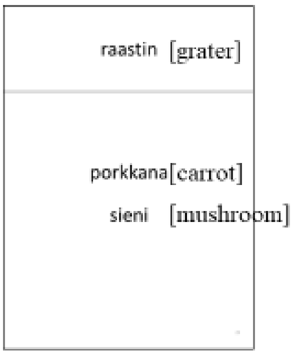
Examples of the Semantic Association Tasks.



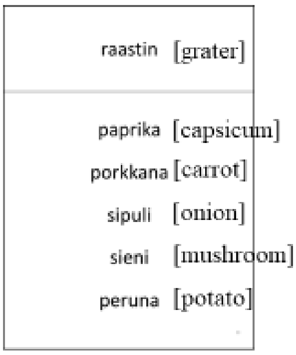
Semantic Association task 1+2 pictures



Semantic Association task 1+5 pictures



Semantic Association task 1+2 words



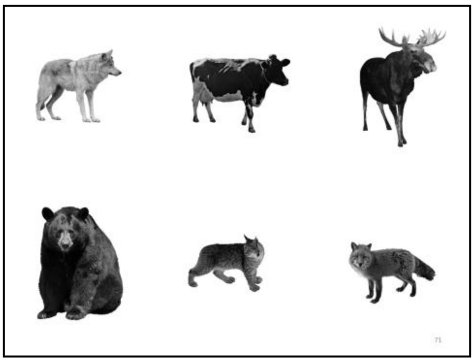
Semantic Association task 1+5 words

2. Odd-One-Out Task

- Task instruction: The participant is asked to decide which one of the response choices is the odd-one-out.
- The task contains 80 items that are divided to four sections: 3, 4, 5 and 6 pictures. Each section contains 20 items. The sections are presented in a randomized order.
- Within the sections, the items are presented in a pseudorandomized order (the superordinate category of the target stimulus is never sequentially the same).

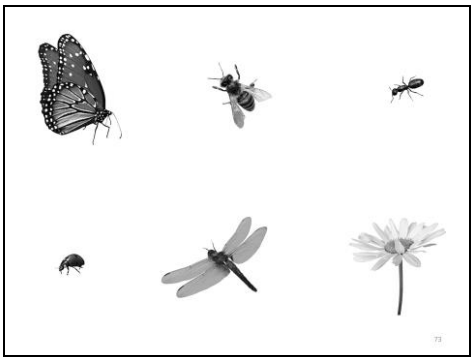
- A half of the target stimuli (40 items; 10 out of each section) represent the same superordinate category than the other response choices. For example, in the category ‘animals’ the target stimulus is ‘moose’ and other response choices are farm animals.

Example of the Odd-One-Out Task (6 pictures). All response choices represent the same superordinate category.



- The other half of the target stimuli (40 items; 10 out of each section) represent a different superordinate category than the other response choices. For example, the target stimulus is ‘daisy’ and other response choices are insects.

Examples of the Odd-One-Out Task (6 pictures). Response choices represent a different superordinate category that the target stimulus.



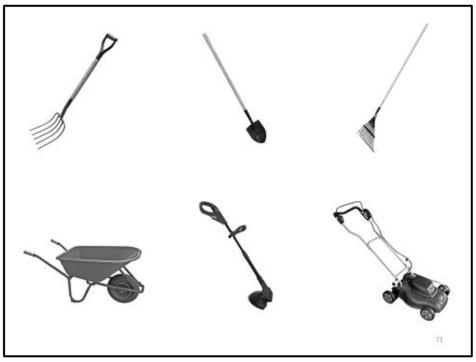
3. Word-Picture Matching tasks

- Task instruction: The participant is asked to point at the picture that matches the spoken/written word.
- The task contains of 80 items that are divided to four sections: 3, 4, 5 and 6 pictures. Each section contains 20 items. The sections are presented in a randomized order.
- Within the sections, the items are presented in a pseudorandomized order (the superordinate category of the stimuli is never sequentially the same). The order is different in the two tasks (Word-Picture Matching task / spoken words and Word-Picture Matching task / written words).
- All response choices represent the same superordinate category. The order of the response choices in randomized for the two tasks.

Examples of the Word-Picture Matching task (4 pictures and 6 pictures)

Spoken: "television"
Written: TELEVISION

Spoken: "rake"
Written: RAKE



4. Category Judgement tasks

- Task instruction: The participant is asked to place cards into the right category.
- The stimuli are presented as pictures (Category Judgement task / pictures) or words (Category Judgement task / words) in 72 cards. The categories (written on the sheets of paper) were placed in front of the participant.
- First, all 72 items are sorted into two semantic categories (living and man-made). Second, the living items (36) are sorted into four semantic subcategories (fruits, vegetables, mammals, birds). Third, the non-living items (36) are sorted into four semantic subcategories (tools, household items, transportation, clothes).

- The items are presented in a randomized order (the cards are shuffled prior to the task).
- In the Category Judgement task / words, the item is presented written and spoken at the same time.

5. Synonym Judgement tasks

- Task instruction: The participant is asked to decide whether the two spoken/written words are synonyms.
- The tasks consist of 80 word pairs. The word pairs are either synonyms or non-synonyms.
- In the spoken version of the task (Synonym Judgement task / spoken words), word pairs are presented from a recording. After each word pair, the participant answers “yes” or “no” during a three-second gap between the items. If the participant is not able to answer in three seconds, the recording is paused.
- In the written version of the task (Synonym Judgement task / written words), word pairs were presented as a list. The participant was asked to mark a cross to “yes” or “no” box.
- Imageability and familiarity of the word pairs are controlled and following four groups of 20 items: (1) high imageability and high familiarity, (2) high imageability and low familiarity, (3) low imageability and high familiarity, and (4) low imageability and low familiarity.

Appendix B.
Descriptive statistics of semantic tasks.

Descriptive Statistics of Factor Scores in the Three Study Groups

Variable		Healthy group (*n*=59)	AD group (*n*=13)	Aphasia group (*n*=14)
Semantic association factor				
	Mean (*SD*)	0.00 (0.40)	-0.90 (0.61)	-0.95 (1.46)
	Range	-1.44–0.53	-2.00– -0.32	-4.64–0.25
Time factor				
	Mean (*SD*)	0.00 (0.40)	-1.00 (0.83)	-1.16 (1.05)
	Range	-1.58–0.60	-2.94–0.08	-3.18–0.00
Verbal factor				
	Mean (*SD*)	0.00 (0.34)	-0.89 (0.54)	-1.52 (1.09)
	Range	-1.10–0.65	-1.97–0.15	-3.67– -0.26
Synonym factor				
	Mean (*SD*)	0.00 (0.45)	-0.51 (0.91)	-0.99 (1.66)
	Range	-1.79–0.32	-2.20–0.25	-4.68–0.25

Descriptive Statistics of Task Scores in the Three Study Groups

Variable		Healthy group (*n*=59)	AD group (*n*=13)	Aphasia group (*n*=14)
BNT				
	Mean (*SD*)	26.58 (3.33)	20.15 (5.84)	15.29 (10.86)
	Range	14–31	8–26	0–31
Semantic verbal fluency / animals				
	Mean (*SD*)	23.64 (5.58)	15.54 (5.46)	10.14 (6.34)
	Range	13–37	9–30	0–20
Semantic verbal fluency / clothes				
	Mean (*SD*)	18.10 (5.50)	13.39 (3.10)	8.93 (4.96)
	Range	5–28	9–17	0–16
Semantic Association task 1+2 pictures				
	Mean (*SD*)	58.32 (1.60)	55.15 (3.60)	54.21 (10.11)
	Range	53–60	46–59	20–60
Semantic Association task 1+5 pictures				
	Mean (*SD*)	56.75 (3.48)	51.39 (2.90)	51.07 (13.65)
	Range	38–60	47–56	9–59
Semantic Association task 1+2 words				
	Mean (*SD*)	58.83 (1.22)	57.23 (2.24)	55.43 (4.24)
	Range	54–60	53–60	47–59
Semantic Association task 1+5 words				
	Mean (*SD*)	57.42 (2.29)	53.39 (3.50)	51.71 (8.27)
	Range	51–60	45–57	32–60
Odd-one-out task				
	Mean (*SD*)	74.22 (4.24)	62.92 (7.72)	70.21 (5.93)
	Range	61–80	49–74	59–78
Spoken Word-Picture Matching task				
	Mean (*SD*)	79.25 (0.92)	76.54 (3.80)	75.07 (7.88)
	Range	77–80	65–80	56–80
Written Word-Picture Matching task				
	Mean (*SD*)	79.42 (0.95)	77.39 (3.36)	77.64 (3.39)
	Range	77–80	67–80	67–80
Category Judgement task pictures				
	Mean (*SD*)	139.78 (5.66)	138.39 (4.59)	132.79 (17.45)
	Range	117–144	126–144	93–144
Category Judgement task words				
	Mean (*SD*)	139.00 (8.01)	137.92 (4.96)	131.21 (21.38)
	Range	107–144	131–144	80–144
Spoken Synonym Judgement task				
	Mean (*SD*)	78.14 (3.00)	75.85 (5.34)	69.93 (13.62)
	Range	66–80	63–79	39–79
Written Synonym Judgement task				
	Mean (*SD*)	77.86 (3.21)	73.92 (6.59)	73.21 (9.32)
	Range	64–80	63–80	49–80

Descriptive Statistics of Task Completion Times in the Three Study Groups

Variable		Healthy group (*n*=59)	AD group (*n*=13)	Aphasia group (*n*=14)
Semantic Association task 1+2 pictures / time				
	Mean (*SD*)	2275.46 (57.20)	2181.46 (101.75)	2177.71 (112.06)
	Range	2063–2361	1967–2307	1866–2297
Semantic Association task 1+5 pictures / time				
	Mean (*SD*)	2203.90 (77.11)	1983.62 (182.76)	2031.57 (171.39)
	Range	2010–2319	1470–2168	1787–2274
Semantic Association task 1+2 words / time				
	Mean (*SD*)	2278.27 (52.12)	2176.39 (130.26)	2075.00 (188.90)
	Range	2063–2346	1892–2302	1650–2256
Semantic Association task 1+5 words / time				
	Mean (*SD*)	2184.05 (88.59)	1911.39 (197.41)	1690.79 (472.52)
	Range	1918–2315	1530–2184	506–2120
Odd-One-Out task / time				
	Mean (*SD*)	2209.93 (48.90)	2125.69 (110.45)	2137.21 (61.07)
	Range	2028–2304	1813–2230	2035–2222
Spoken Word-Picture Matching task / time				
	Mean (*SD*)	2235.73 (41.52)	2175.92 (71.42)	2113.86 (141.85)
	Range	2090–2301	2057–2267	1688–2238
Written Word-Picture Matching task / time				
	Mean (*SD*)	2034.10 (106.13)	1808.62 (242.26)	1819.43 (286.71)
	Range	1703–2174	1288–2123	1110–2124
Category Judgement task pictures / time				
	Mean (*SD*)	2111.36 (86.67)	1954.92 (97.21)	1992.71 (142.78)
	Range	1853–2242	1753–2067	1642–2148
Category Judgement task words / time				
	Mean (*SD*)	2113.95 (81.99)	1901.15 (131.86)	1947.57 (164.32)
	Range	1830–2277	1612–2050	1553–2115
Written Synonym Judgement task / time				
	Mean (*SD*)	2276.41 (69.69)	2103.77 (155.92)	2001.79 (307.66)
	Range	2012–2379	1724–2286	1282–2268