Name agreement and naming latencies for typed picture naming in aging adults

Reference:
Paesen Lise, Leijten Mariëlle.- Name agreement and naming latencies for typed picture naming in aging adults
Clinical linguistics and phonetics - ISSN 0269-9206 - Philadelphia, Taylor & francis inc, (2019)19 p.
Full text (Publisher's DOI): https://doi.org/10.1080/02699206.2019.1590734
To cite this reference: https://hdl.handle.net/10067/1582180151162165141

Name agreement and naming latencies for typed picture naming in aging adults

Paesen, Lise & Leijten, Mariëlle

Management, University of Antwerp, Antwerp, Belgium

Name agreement and naming latencies for typed picture naming in aging adults

This baseline study aimed to create a coherent set of images that can be used to describe language decline found in healthy elderly and to compare this to the language change found in the early stages of Alzheimer's disease. To this extend, a typed picture naming task was created, in which visual complexity, age-of-acquisition, frequency and name agreement were controlled for. 76 healthy elderly participated in the test; their data will be used in follow-up studies to compare with cognitively impaired patients. The entire typing process was logged with keystroke logging tools Inputlog and Scriptlog; the obtained results were analysed in light of the typing product (name agreement and object recognition) and the writing process (naming latencies and interkey latencies). Results showed that the latencies increased with age and that the older participants had longer latencies for images with a lower frequency and higher age-of-acquisition. Hence, our results indicate the need to take both the latencies and the typing product into consideration.

Keywords: language decline; keystroke logging; picture naming

Introduction

Due to the gradual language impairment in the early stages of Alzheimer disease, research into the nature of those language changes and the difference with normal language decline could provide valuable insight into the onset of the disease and into normal cognitive health of the aging elderly. To that effect, researchers have often focussed on language production which, in contrast to language comprehension, shows signs of specific age-related impairments (Burke & Mackay, 1997; Shafto, Burke, Stamatakis, Tam, & Tyler, 2007). Older people often complain about their inability to come up with a specific word, even though they are certain they know that word (Burke & Mackay, 1997; Clark-Cotton, Williams, Goral, & Obler, Loraine, 2007; Shafto et al., 2007; Shafto & Tyler, 2014; Silagi, Bertolucci, & Ortiz, 2015). These so-called tip-of-the-tongue errors are not the result of a decline in semantic skills since the word is almost always retrieved somewhat later; they are due to an impaired access to phonological and/or orthographical information (Burke & Mackay, 1997; Clark-Cotton et al., 2007; Shafto et al., 2007). To measure the extent of the changes in word retrieval, researchers have often relied upon picture naming tasks and/or studied the effects of certain word/image characteristics on naming speed. These tasks provide valuable insight in the language production processes that are required to name the prompted image as fast as possible. By manipulating the given prompts, researchers hope to influence naming latencies and naming agreement. In this study we aimed to create a coherent set of images that can be used in follow-up studies to compare the language decline found in healthy aging controls and the clinical language decline found in the early stages of AD. Therefore, the main aim of this study was to determine how the naming accuracy and latencies are affected by age and variable characteristics so as to create a baseline for further research.

It is important to note that the studies on which we will base ourselves have made use of spoken naming, mostly disregarding the written language production. Nevertheless, studying written language production is crucial to fully grasp both the naming latencies and the cognitive processes underlying the picture naming. Therefore, it is necessary to test the foregoing theories on the written product, keeping in mind that the theories developed for spoken picture naming should not uncritically be applied to written picture naming for reasons of spelling and the change of communicative medium (Torrance et al., 2018). Since the onset of AD is not characterised by motor abnormalities (Cummings & Benson Boston, 1983), follow-up studies will be able to correctly compare the data of AD patients with healthy age-matched controls. In what follows we will translate the existing theories on oral picture naming to an alternative that makes use of written – more specifically typed – picture naming. We will do so by using keystroke logging tools; computer programs that “log and time stamp keystroke activity to reconstruct and describe text production processes” (Leijten & Van Waes, 2013, p359). Our baseline study will focus on the difference in image characteristics, on the language changes that might occur upon aging, and on the coherence of the results in a longitudinal study. Two participant groups, an age group (AG) with younger adults (AG 50 - 64) and a group with older adults (AG 65+), were tested two times with a time interval of three months (moment1 and moment 2) using a within and between subject design.

Overview of word characteristics

Previous studies have pinpointed four eminent word characteristics that influence naming latencies: age of acquisition, frequency, familiarity and name agreement. A first word characteristic which influences naming accuracy and latencies is the *age-of-acquisition* (AoA) of words. Early-acquired words can be triggered more rapidly and thoroughly than later-acquired words (Brysbaert, Stevens, De Deyne, Voorspoels, & Storms, 2014; Rodríguez-Ferreiro, Davies, González-Nosti, Barbón, & Cuetos, 2009). Even in healthy age-matched controls early acquired words are named with a higher speed and more accuracy than those words that are acquired later on in life (Brysbaert, Warriner, & Kuperman, 2014; Gilhooly & Logie, 1980; Kremin et al., 2001). This effect can be contributed to the organisation of the mental lexicon, in which earlier-acquired words tend to have more connections than later-acquired words. Nevertheless, that is only in the condition that those early-acquired words are also frequently used throughout someone’s life and are not only used early on in life (Brysbaert, Stevens, et al., 2014; Lété & Bonin, 2013). Therefore, age-of-acquisition is often studied in interaction with *frequency*, the second determining word characteristic (Brysbaert, Stevens, et al., 2014). Although some studies did not find any frequency effect on the reaction times (Bonin, Fayol, & Chalard, 2001), more recent studies argue that frequency effects can indeed be found, if the AoA is lower (Barry, Morrison, & Ellis, 1997; Severens, Lommel, Ratinckx, & Hartsuiker, 2005). More specifically, reaction times have been found to improve with words that have an early AoA and a high frequency (Scaltritti, Arfé, Torrance, & Peressotti, 2016). Studies that focus on AD showed similar results: AoA had an effect on the naming accuracy, resulting in faster naming latencies when presented with an early-acquired word (Garrard, Maloney, Hodges, & Patterson, 2005; Rodríguez-Ferreiro et al., 2009). Frequency did predict picture naming in AD. Naming accuracy improve when words occur more in the language (Garrard et al., 2005; Kremin et al., 2001), and words with a low frequency tend to lead to more tip-of-the-tongue errors for patients (Astell & Harley, 1996; Garrard et al., 2005). These characteristics are closely related to *familiarity*. Research in both healthy and cognitively impaired patients revealed that familiarity to a concept helps to speed naming (Brysbaert, Stevens, et al., 2014; Garrard et al., 2005; Kremin et al., 2001; Rodríguez-Ferreiro et al., 2009; Scaltritti et al., 2016; Zannino et al., 2010). However, this effect can be minimalised due to the concept's highly relatedness to AoA and frequency (Brysbaert & Cortese, 2011; Brysbaert, Stevens, et al., 2014). A very familiar concept is often a concept learned at a very young age and that is encountered frequently in daily life (Brysbaert, Stevens, et al., 2014). A fourth influencing characteristic is *name agreement,* which refers to the extent to which different people agree on the name of a certain thing. In studies that focus on either healthy or cognitively impaired participants, higher naming accuracy and faster latencies were found for words with a higher naming agreement (Garrard et al., 2005; Kremin et al., 2001; Rodríguez-Ferreiro et al., 2009; Scaltritti et al., 2016; Severens et al., 2005).

Overview of image characteristics

Naming latencies in picture naming tasks are not only influenced by the characteristics of the targeted word. They are also influenced by the image characteristics, such as visual complexity, colour and what is depicted. In general, researchers argue that both healthy and cognitively impaired patients have greater difficulty naming living things than man-made objects (Adlington, Laws, & Gale, 2009; Duarte & Robert, 2014), even though AD patients are thought to have a deficit in both categories (Adlington et al., 2009). Some studies attribute the advantage for non-living things to the lower number of features that are needed to distinguish between the different man-made things (Duarte & Robert, 2014). The naming of living things would benefit from more features and surface detail (Adlington et al., 2009). In that respect colour showed to have a positive effect on naming latencies, especially if the colour was the diagnostic characteristic of the object (e.g.: an orange orange or a yellow banana). Therefore, researchers argue that the recognition of living things benefits from the addition of colour; non-living things have a lower colour diagnosticity and do not need the addition of colour to be easily recognised (Adlington et al., 2009; Moreno-Martinez & Rodriguez-Rojo, 2015; Rodríguez-Ferreiro et al., 2009; Zannino et al., 2010). Since cognitively impaired patients often suffer from a visual impairment, an increase in colour contrasts even aids AD patients leading to comparable results with healthy controls (Adlington et al., 2009). Nevertheless, this effect cannot be replicated when adding photographic detail (Zannino et al., 2010). For the remainder of this article we will discuss how we interpreted the existing theories for the purpose of our study, how we collected the data and what the language product and process data looked like.

Selection of images for the picture naming task

For the selection of the images, we opted for images that meet strict criteria, based on existing theories described in the previous section. Those images are divided into two categories: the first category, ‘highly relatable images’ (HR), will contain ‘easier’ images; the second, ‘lower relatable images’ (LR), will contain more complex images in order to account for both floor and ceiling effects in follow-up studies with AD patients (Forbes-McKay, Shanks, & Venneri, 2014). Even though the names HR and LR do not do justice to the depth and extend of those categories, those terms will be used in the remainder of this article for convenience’s sake.

For the image selection, we used the picture naming norms described by Severens and colleagues (2005) for 590 pictures in Dutch. We subjected the images of their study to several elimination rounds, always distinguishing between the HR and LR categories. We first selected only those images that are easily nameable (1) and imageable in order to allow for an increased naming speed and accuracy. Then we selected images on AoA (2) and the frequency (3) with which they appear in Dutch. For the remainder of the images we evaluated the mean spoken reaction times (4). We selected those images that could be found within two standard deviations of the mean latencies per category to create a coherent set of images and to discard outliers. Keeping in mind that some AD patients in our follow-up study could have been subjected to other picture naming studies (e.g. Boston naming test), we decided to eliminate images that appear in other tests (5) for the LR category and keep the images of the HR category, since we presume that participants are more accustomed to the latter. An overview of the five selection criteria is provided in table 1.

[Insert table 1 about here]

After this first selection of images, we decided to eliminate words that are difficult to type (6) such as words containing a diaeresis (e.g.: ‘pinguïn’ - penguin in

Dutch). The remaining images were divided into several sub-categories, distinguishing between living things and man-made objects (7); details are provided in Table 2.

[Insert table 2 about here]

The resulting selection consisted of 50 images; these images were redrawn for reasons of visual complexity (8) and coherence in a longitudinal design. We deleted actions from the images, deleted 'decorative' objects, made images of the same sub-category 'look' in the same direction and added colour (9). A few examples of the new images are provided below (cf. figures 1 to 3). After carefully pre-testing those images in a typed picture naming experiment on 200 students and healthy elderly, we deleted two images due to insufficiently high name agreement (<50%). All images that were created for the purpose of this study will be placed in the 'Open Linguistic Picture Database' or OLPD (for more information: www.olpd.eu).

[insert figure 1, 2 and 3 about here]

Aim and hypotheses

This study aimed to create a coherent image set that can be used to distinguish healthy language decline from pathological language decline upon aging. These images were studied in light of the naming accuracy (name agreement & object recognition) and latencies (naming latencies & interkey latencies).

Naming accuracy

- We hypothesize that name agreement & object recognition will be equal for the HR and LR categories.
- We hypothesize that name agreement & object recognition will not differ between the two age groups (AG 50 – 64 and AG 65+).

Latencies

- We hypothesize that naming & interkey latencies will be longer for LR than HR nouns.
- We hypothesize that naming & interkey latencies will be longer for the older age group (AG 65+).
- We hypothesize that naming & interkey latencies will remain the same with a three-month interval.

Naming accuracy x latencies

We believe that name agreement can be used to predict naming & interkey latencies due to its high correlation with the latencies.

Method

Materials and procedure

Each participant followed the same procedural steps during both test moments. Firstly, they had to complete the (1) typing test (Van Waes, Leijten, Mariën, & Engelborghs, 2017), followed by the (2) picture naming test, a (3) questionnaire, the (4) Montreal Cognitive Assessment (MoCA), and the (5) Geriatric Depression Scale (GDS). The typing test and picture naming test were administered on the researcher’s computer on which two keystroke logging tools were installed: tools are programs that log and time stamp every keyboard activity. The tools were respectively Inputlog 7.05 (Leijten & Van Waes, 2013) and Scriptlog (Frid, Johansson, Johansson, Wengelin, & Johansson, 2014).

(1) Typing test

In order to account for interpersonal differences in typing speed, we decided to subject all of our participants to a typing test. One specific module of Inputlog 7.0 is the ‘copy task’, which consists of different typing tests in which specific words/letters/sentences – made up of specific bigram combinations - need to be copied. The assignment is presented left on the screen; hence the letters/words/sentences can be copied and do not need to be remembered, keeping the cognitive load to a minimum (Leijten & Van Waes, 2013; Van Waes et al., 2017). Seven consecutive assignments need to be fulfilled: (1) Repetition of two letters for a time span of 15 seconds, (2) a sentence repetition task for a time span of 30 seconds, (3 to 6) copying a combination of two/three words seven times, (7) copying four blocks of six consonants. A final questionnaire enquired after the handedness of the participant.

(2) Picture Naming Test

A picture naming test was created, using 48 coloured images as described in the introduction. The keystroke logging tool Scriptlog (Frid et al., 2014) was used to log and time stamp the entire writing process. Before commencing the picture naming task, the participants were instructed to name the prompted picture by typing. They were told it was crucial that they used only one word to name the image (no adjectives, no articles) and that they had to name the image as fast as they could. Every participant was given three images as a practise session to make sure they understood the instruction and afterwards a random combination of the 48 images (not including those three practise images) followed.

(3) Questionnaire

With the use of a questionnaire we enquired after the background of the participants. Participants had to answer questions on their studies, job, possible visual impairments, language impairments, bilingualism and possible disorders.

(4) MoCa

The Montreal Cognitive Assessment (Nasreddine et al., 2005) is a cognitive screening tool developed to differentiate healthy cognitive aging from Mild Cognitive Impairment (MCI) - the clinical stage before dementia. We decided to opt for this test rather than its famous counterpart the MMSE (Folstein, Folstein, & McHugh, 1975) for three main reasons: the shorter administration time, its higher sensitivity to the earliest stages of AD and its high reliability in re-testing within a period of three months (Nasreddine et al., 2005). During the test, tasks are given on short-term memory recall, visuospatial abilities, executive functions, attention, concentration, working memory, language and orientation to time and place (Nasreddine et al., 2005). We administered the Dutch version 7.1 on the first encounter with our participants and version 7.2 on the second. In order to be considered cognitively healthy, a participant needs to achieve a score of 26 out of 30 or higher. A score between 18 and 25 out of 30 indicates that a participant might suffer from mild cognitive impairment.

(5) GDS

The Geriatric Depression Scale is a questionnaire developed for screening especially elderly on possible signs of a depression (Sheikh & Yesavage, 1986). In order to keep the total test time for our study to a minimum, we decided to select the version containing 15 questions. A score of 6 out of 15 or more indicates that the participant shows signs of a depression.

Participants

To form the base-line of this study, we recruited 76 healthy participants; 68 of them completed the two contact moments (moment 1 and moment 2 which was approximately 3 months later). They were contacted through several elderly organisations, the network of other participants and the researcher's own network. Participants were native Dutch speakers, aged 50 or older on the moment of the first test and had to be sufficiently computer literate. Participants were excluded if they had a history of neurological/psychological illnesses and they were not allowed to have a writing disability. Participants were tested individually at a location of their choice – mostly their home. After careful selection of the participants, we used the data of 62 participants. Note that eight of those 62 participants had a lower MoCA score on one of the two test moments, leading to usable data of 58 participants for moment 1 and 60 for moment 2 (see table 3). Participants with both a positive and a negative MoCA score on the two test moments will be excluded for some analyses (cf. statistical analysis). We divided the participants into two age groups based on the retirement age in Belgium, which is 65: a younger age group (AG 50 - 64) with people aged between 50 and 64 and an older age group (AG 65+) with people aged 65+.

[insert table 3 about here]

Statistical analysis

The data that was logged with Scriptlog (Frid et al., 2014) was transformed into a file that can be analysed with the keystroke logging tool Inputlog 7.05 (Leijten & Van Waes, 2013; free download for research purposes on www.inputlog.net). The latter tool allows for more detailed analysis of the naming latencies via the so-called *Word Pause analysis,* which enabled us to define the words that were written, the naming latencies (pauses before the nouns) and the interkey latencies (pauses between letters). We selected the relevant variables (cf. Table 4) and used SPSS for further analyses.

[insert table 4 about here]

Data reduction

In total, we expected 6912 nouns to be typed over the course of the two picture naming tasks; 48 images were named by 76 participants during the first moment and by 68 participants the second time. However, due to technical errors, only 6854 nouns were typed. These data were also produced by participants whose MoCA or GDS scores violated the aforementioned requirements on one or both test moments. We deleted their data and used the resulting data of 58 healthy elderly on the first moment and 60 on the second moment. Furthermore, not all of these nouns were target nouns; some participants added adjectives (e.g.: *black* cat) and even entire descriptions (e.g.: pirate *with wooden leg and eye patch*), thus increasing the word count. Therefore, nouns were excluded if they were proceeded by an adjective or description. Nouns that were followed by a description were kept for some analyses since there were no proceeding words to influence the naming latency of the target noun. The remaining data consisted of 5469 nouns. These nouns were divided into six categories (cf. Table 5):

(1) Named correctly: images that were named correctly without typing/spelling mistakes.

(2) Named correctly – with typo correction: images that were named correctly; a typing/spelling mistake was made during the writing process and this error was corrected

(3) Named correctly – containing typo: images that were named correctly; a typing/spelling mistake can be found in the final word product (e.g.: ‘sharrk’ instead of ‘shark’)

(4) Synonyms: images that were recognised correctly but given an alternative name or morphological variant (e.g.: cat = ‘kat’ or ‘poes’ or ‘katje’)

(5) Empty answers: images that were not or barely named due to technical errors, by pressing the ‘enter’ button too quickly by accident or by pressing the ‘enter’ button while correcting a mistake in the middle of the word

(6) Incorrect names: images that were given the wrong name (e.g.: typing ‘fish’ instead of ‘shark’) or adjectives

[insert table 5 about here]

Depending on the analyses, we in- or excluded some of these scores:

- When looking at the product variable ‘Name Agreement’, we only took words into consideration that were named correctly, independent of typing/spelling errors (85.9% of words included);
- The product variable ‘Object recognition’ was be calculated using the data from the correctly named words (independent of errors) and synonyms (94.81% of words included);
- The naming latencies for the different words (before word pause), indicating how long it took participants to start typing the word, were only measured for words that were named correctly, independent of typing/spelling errors (85.9% of words included);
- The time it took participants to write a certain word (interkey latencies) – taking into account their personal typing speed and the number of letters per word – could only be calculated for words that were named and typed completely correct. Words that contain(ed) typing/spelling errors were discarded (76.43% of words included);

- The difference in naming latencies before words between the two tests moments was only be calculated for participants that enrolled in both test moments and had a sufficient MoCA score both moments. Words that were named correctly, independent of spelling, were included (85.9% of words included);
- The difference in naming latencies within words per letter between the two different tests moments was only be calculated for participants that enrolled in both moments and had a sufficient MoCA score both moments. Words that were named correctly – without typing/spelling mistakes in the process and/or product – will be included (76.43% of words included).

Results

Naming accuracy

In order to measure the naming accuracy for a certain object/animal/human/vehicle we made a distinction between two variables. On the one hand, there is name agreement where we only allowed those objects that were named completely correctly, independent of typing/spelling errors. On the other hand, there is object recognition, a variable in which synonyms were also taken into consideration.

[insert table 6 about here]

'Name agreement will be equal for the two categories HR and LR'. We presumed that name agreement would be equal for both categories since it was one of the selection criteria. Results affirm this hypothesis, $p = 0.627$ (cf. table 6). More specifically, a Mann-Whitney test showed no differences within the age groups for the HR and LR words; AG 50 - 64: $U = 188.50$, $p = 0.065$, $r = 0.267$; AG 65+: $U = 243.00$, $p = 0.491$, $r = 0.099$, (cf.: table 7).

'Name agreement will not differ between the two age groups, for neither HR and LR'. A

paired-samples t-test indicated that there was no difference in name agreement between the younger age group and the older age group, $t(18) = 1.63$, $p = 0.120$, $d = 0.37$ for HR images. However, a significant difference for LR images was found between the first age group and the older age group, $t(28) = 2.367$, $p = 0.025$, $d = 0.44$. This result indicates that the 65+ year olds did have more difficulty in correctly naming the LR images compared to their younger counterparts.

[insert table 7 about here]

'Object recognition will be equal for the two categories HR and LR'. The general object recognition scores as shown in table 6 indicate that there a significant difference in object recognition between HR images and LR images, $p = 0.006$. In addition to those results, nearly but no significant differences were found in within each age group: AG 50 - 64 ($U = 187.500$, $p = 0.053$, $r = 0.280$) and a significant difference between high and low frequency words was found for AG 65+ $U = 131.000$, $p = 0.002$, $r = 0.448$ (cf.: table 8).

'Object recognition will not differ between the two age groups, for neither HR and LR'. With the paired-samples t-test, no significant differences were found for LR words $t(28) = 1.046$, $p = 0.304$, $d = 0.194$. However, contrary to the results of the name agreement, there was a significant difference in naming accuracy for HR images $t(18) = -3.327$, $p = 0.004$, $d = 0.764$.

[insert table 8 about here]

Latencies

When comparing the naming and interkey latencies in a between-subject design, it is important to take the computer literacy and hence the personal typing speed of all participants into consideration. The analysis of the typing task, particularly the analysis of bigram intervals as provided by Inputlog 7.05 (Leijten & van Waes, 2013), provided a measure that exemplifies the typing literacy of the participants (cf. table 9).

[insert table 9 about here]

A Mann-Whitney test revealed that there is no significant difference in typing speed between the two groups and the two test moments; moment 1: $U = 421.5$, $p = 0.173$, $r = 0.169$; moment 2: $U = 436.0$, $p = 0.242$, $r = 0.145$. Since the typing speed of the two groups is not significantly different, we decided not to take the typing speed into consideration in further analyses of the latencies.

'Naming & interkey latencies will be longer for LR than HR nouns'. Significant differences in naming latencies were observed within both AG 50 - 64 $U = 660650.000$, $p <0.001$, $r = 0.122$ and AG 65+ $U = 445890.500$, $p <0.001$, $r = 0.163$. Furthermore, a significant difference in interkey latencies was observed within AG 50 - 64 $U = 438273.000$, $p <0.001$, $r = 0.258$ and AG 65+ $U = 360877.000$, $p < 0.001$, $r = 0.175$ (cf. table 10).

[insert table 10 about here]

'Naming & interkey latencies will be longer for the older age group'. When looking at the naming latencies, results of the Mann-Whitney test show that there is a significant difference for the LR images $U = 740842.000$, $p <0.001$, $r = 0.227$ and HR images $U = 340865.500$, $p <0.001$, $r = 0.169$ (cf. table 10). Looking at the data of interkey latencies, we see the following results with a Mann-Whitney test: in the naming of LR images, the interkey latencies differed significantly $U = 714278.500$, $p <0.001$, $r = 0.072$; the same goes for HR images $U = 314575.00$, $p <0.001$, $r = 0.124$ (cf. table 10).

'Naming & interkey latencies will remain the same with a three-month interval'. Since the previously described results might be induced by our repeated testing with an

interpose of three months, the data of the two test moments will also be compared with a Mann-Whitney test (cf. table 11). For AG 50 - 64, there was no significant difference in naming latencies neither for LR nouns U = 289713.000, p = 0.264, r = 0.028 nor for HR nouns U = 120557.500, p = 0.390, r = 0.027. Similar results can be found for AG 65+, with the difference that for both LR images U = 198359.500, p = 0.057, r = 0.053 and HR images U = 83612.000, p = 0.071, r = 0.062 the results are nearly significant.

[insert table 11 about here]

A Mann-Whitney test revealed that there is no significant difference in interkey latencies in AG 50 - 64 for both LR images U = 224496.00, p = 0.135, r = 0.040 and HR images U = 103869.500, p = 0.951, r = 0.002. The opposite is true for AG 65+, for both LR images U = 143902.500, p = 0.013, r = 0.074 and HR images U = 71981.500, p = 0.026, r = 0.078 (cf. table 11).

Naming accuracy x latencies

In order to measure if latencies can be predicted with the use of the name agreement, a Pearson correlation coefficient was computed.

'Naming & interkey latencies can be predicted using name agreement'. There was a negative correlation between the two variables for naming latencies, r = -0.397, n = 96, $p < 0.001$. This result is summarized with a scatterplot (cf. figure 4). No significant correlation was found for interkey latencies: r = -0.139, n = 96, p = 0.177.

[insert figure 4 about here]

Discussion

With this study we aimed to create a coherent set of images that can be used to distinguish healthy cognitive aging from pathological aging. Therefore, the influence of age and image characteristics on naming accuracy and naming latencies was studied as a baseline for further research.

Naming accuracy

We studied naming accuracy by distinguishing between name agreement and object recognition; two variables that account for the amount of images that are named correctly in either a strict sense (name agreement) or a wider sense (object recognition) in which synonyms and morphological variants were allowed. Since we controlled for name agreement upon selecting the images from previous studies, we presumed that neither image characteristics nor increasing age would trigger differences in naming accuracy. Image characteristics did not influence the name agreement nor object recognition scores within the younger age group. Accordingly, they did not influence name agreement in the older age group. However, this older age group showed lower object recognition scores for LR images compared to HR images. We attribute this effect to the higher amount of synonyms that were found for the HR images. With respect to the influence of age, the results are more ambiguous to interpret. Whereas the name agreement scores did not differ for HR images, the object recognition scores indicated that the younger age group did have more difficulty with those images than the older age group. Surprisingly, the name agreement scores for LR images showed greater difficulty for the older age group compared to the younger, a result that was evened out when looking at object recognition. When disregarding the results of the object recognition scores, our results are in line with previous studies where healthy aging elderly tended to have more tip-of-the-tongue errors in low frequent words than in high frequent words (Clark-Cotton et al., 2007). Moreover, our study adds that it is only with increasing age that word retrieval issues for lower relatable (and thus lower frequent) words arise. Since AD patients in the early stages of the disease show an increased impairment in word retrieval, we believe that the discrepancy in naming accuracy between the HR and LR images will only grow. The issues encountered in object recognition scores can be contributed to the differences in synonyms and morphological variants used. Therefore, we would suggest that name agreement is a more stable determiner of differences in picture naming than object recognition.

Latencies

We presumed that both image characteristics and age would influence the latencies (naming and interkey), with a stable result between different test moments. As expected, image characteristics were correctly predicted to influence latencies, with faster reaction times for the easier images (HR vs LR). Furthermore, latencies were also correctly hypothesized to increase with age, even though the typing speed of the two groups was comparable. This finding is consistent with previous findings on spoken picture naming and is further proof of the increasing difficulty healthy aging elderly have with lexical retrieval (Kemper & Sumner, 2001; Shafto & Tyler, 2014). Furthermore, our results are also in line with studies that have found an effect of AoA and frequency on the naming latencies in picture naming tasks (Barry et al., 1997; Boukadi, Zouaidi, & Wilson, 2016; Levelt, Roelofs, & Meyer, 1999; Scaltritti et al., 2016). With respect to the change in latencies over time, our results are less straight forward. The younger age group showed no difference in naming and interkey latencies between the two moments. Accordingly, the older age group showed similar naming latencies, indicating that naming latencies are a stable measure to use in longitudinal research. The interkey latencies, however, triggered faster results in the older age group for the second test moment. Possibly, we triggered a learning effect; once the required word form is retrieved, the processing within the word takes less time because the word has been used before in the previous test session.

With respect to future research, we tested the correlation between latencies and name agreement in order to find out if name agreement can be used to predict naming latencies. Our results indicate that name agreement can be used as a predictor for naming latencies, with faster reaction times to words with a higher name agreement. Naming latencies are also found to be influenced by name agreement in other studies; with faster reaction times for words that were both HR and had a low AoA, which we manipulated in our image selection (Scaltritti et al., 2016). This finding could not be replicated for interkey latencies, possibly triggered by the ongoing search for the correct orthographical form for the retrieved word.

Implications

This baseline study set out to create a coherent set of images for a typed picture naming task that can be used in future research to differentiate healthy cognitive aging from pathological aging. To that extend, our results imply that even in healthy aging, differences in typed picture naming can be found. Therefore, we find it necessary never to treat the healthy controls group as a whole, but rather as a set of individuals with different ages. By logging the entire writing process we were able to provide additional and more in-depth details to the cognitive processes that are required to name the prompted images. These latencies are needed in addition to the naming accuracy data. We believe that for studies with a longitudinal design, naming latencies provide a stable measure in differentiating between AD patients and healthy controls. However, the change in interkey latencies must be studied more thoroughly in future studies, since it provides more details on the ongoing writing process and the difficulty encountered during this process. It is important to stress the exploratory nature of this study. We found differences in latencies both between the different image categories and age groups. This implies that there is a change in language capacities in healthy aging elderly; further studies with cognitively impaired patients are needed to define more clearly which language change is still to be considered healthy and which is pathological.

Acknowledgements

This research was supported by the research fund BOF-DOCPRO 2015, PS ID: 31661 of the University of Antwerp. We would also like to thank Eric Van Horenbeeck, the technical coordinator of Inputlog, and Luuk Van Waes, co-founder of Inputlog and the typing task. We thank Johan Frid and Victoria Johansson, respectively the technical coordinator and co-founder of Scriptlog for creating a Scriptlog module especially for our research purposes. Further thanks go to Peter Mariën and Dorien Vandenborre for their input during the design of the study. Iris Schrijver receives thanks for the revision of this article. Last but not least we would like to thank the elderly who enthusiastically participated in this study.

Declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Bibliography

Adlington, R. L., Laws, K. R., & Gale, T. M. (2009). Visual processing in Alzheimer's disease: Surface detail and colour fail to aid object identification. *Neuropsychologia*, *47*(12), 2574–2583. https://doi.org/10.1016/j.neuropsychologia.2009.05.004

Astell, A. J., & Harley, T. A. (1996). Tip-of-the-tongue states and lexical access in dementia. *Brain and Language*, *54*(2), 196–215. https://doi.org/10.1006/brln.1996.0071

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart Pictures: Effects of Age of Acquisition, Frequency, and Name Agreement. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *50*(3), 560–585. https://doi.org/10.1080/783663595

Bonin, P., Fayol, M., & Chalard, M. (2001). Age of Acquisition and Word Frequency in Written Picture Naming. *The Quarterly Journal of Experimental Psychology Section A*, *54*(2), 469–489. https://doi.org/10.1080/713755968

Boukadi, M., Zouaidi, C., & Wilson, M. A. (2016). Norms for name agreement, familiarity, subjective frequency, and imageability for 348 object names in Tunisian Arabic. *Behavior Research Methods*, *48*(2), 585–599. https://doi.org/10.3758/s13428-015-0602-3

Brysbaert, M., & Cortese, M. J. (2011). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *Quarterly Journal of Experimental Psychology*, *64*(3), 545–549. https://doi.org/10.1080/17470218.2010.503374

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, *150*, 80–84. https://doi.org/10.1016/j.actpsy.2014.04.010

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Burke, D. M., & Mackay, D. G. (1997). Memory, language, and ageing. *Philos Trans R Soc Lond B Biol Sci*, *352*(1363), 1845–1856. https://doi.org/10.1098/rstb.1997.0170

Clark-Cotton, M.R., Williams, R.K., Goral, M., & Obler, L.K. (2007). Language and communication in aging. In J.E. Birren (Ed.), *Encyclopedia of gerontology: Age, aging, and the aged* (2nd ed., pp. 1–8). London: Elsevier

Cummings, J. L., & Benson, D. F. *Dementia: A clinical approach*. Boston, Butterworths, 1983.

Duarte, L. R., & Robert, C. (2014). Semantic richness influences naming pictures in old age and in Alzheimer's disease. *Journal of Neurolinguistics, 31*, 55–68. https://doi.org/10.1016/j.jneuroling.2014.06.004

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*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Forbes-McKay, K., Shanks, M., & Venneri, A. (2014). Charting the decline in spontaneous writing in Alzheimer's disease: a longitudinal study. *Acta Neuropsychiatrica*, *26*(04), 246–252. https://doi.org/10.1017/neu.2014.2

Frid, J., Johansson, V., Johansson, R., Wengelin, Å., & Johansson, M. (2014). Developing a keystroke logging program into a writing experiment environment. Presentation at Writing Across Borders, WRAB, Paris February 2014.

Garrard, P., Maloney, L. M., Hodges, J. R., & Patterson, K. (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain*, *128*(2), 250–260. https://doi.org/10.1093/brain/awh341

Gilhooly, K. J., & Logie, R. H. (1980). METHODS &DESIGNS Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, *12*(4), 395–427. https://doi.org/10.3758/BF03201693

Kemper, S., & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging*, *16*(2), 312–322. https://doi.org/10.1037//0882-7974.16.2.312

Kremin, H., Perrier, D., Wilde, M. De, Dordain, M., Bayon, A. Le, Gatignol, P., … Arabia, C. (2001). Factors predicting success in picture naming in Alzheimer's disease and primary progressive aphasia. *Brain and Cognition*, *46*(1–2), 180–183. https://doi.org/10.1006/brcg.2000.1270

Leijten, M., & Van Waes, L. (2013). Keystroke Logging in Writing Research. *Written Communication*, *30*(3), 358–392. https://doi.org/10.1177/0741088313491692

Lété, B., & Bonin, P. (2013). Does frequency trajectory influence word identification? A cross-task comparison. *Quarterly Journal of Experimental Psychology*, *66*(5), 973–1000. https://doi.org/10.1080/17470218.2012.723725

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1–75. https://doi.org/10.1017/S0140525X99001776

Moreno-Martinez, F. J., & Rodriguez-Rojo, I. C. (2015). On colour, category effects, and Alzheimer's disease: A critical review of studies and further longitudinal evidence. *Behavioural Neurology, 2015,* 1–12. https://doi.org/10.1155/2015/960725

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment , MoCA : A Brief Screening. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Rodríguez-Ferreiro, J., Davies, R., González-Nosti, M., Barbón, A., & Cuetos, F. (2009). Name agreement, frequency and age of acquisition, but not grammatical class, affect object and action naming in Spanish speaking participants with Alzheimer's disease. *Journal of Neurolinguistics*, *22*(1), 37–54. https://doi.org/10.1016/j.jneuroling.2008.05.003

Scaltritti, M., Arfé, B., Torrance, M., & Peressotti, F. (2016). Typing pictures: Linguistic processing cascades into finger movements. *Cognition, 156*, 16–29. https://doi.org/10.1016/j.cognition.2016.07.006

Severens, E., Lommel, S. Van, Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119*(2), 159–187. https://doi.org/10.1016/j.actpsy.2005.01.002

Shafto, M. A., Burke, D. M., Stamatakis, E. A., Tam, P. P., & Tyler, L. K. (2007). On the Tip-of-the-Tongue: Neural Correlates of Increased Word-finding Failures in Normal Aging. *Journal of Cognitive Neuroscience, 19*(12), 2060–2070. https://doi.org/10.1162/jocn.2007.19.12.2060

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health*, *5*(1–2), 165–173. https://doi.org/10.1300/J018v05n01_09

Silagi, M., Bertolucci, P., & Ortiz, K. (2015). Naming ability in patients with mild to moderate Alzheimer's disease: what changes occur with the evolution of the disease? *Clinics*, *70*(6), 423–428. https://doi.org/10.6061/clinics/2015(06)07

Torrance, M., Nottbusch, G., Alves, R. A., Arfé, B., Chanquoy, L., Chukharev-Hudilainen, E., Dimakos, I., Fidalgo, R., Hyönä, J., Jóhannesson, Ó. I., Madjarov, G., Pauly, D. N., Uppstad, P. H., van Waes, L., Vernon, M., & Wengelin, Å. (2018). *Timed written picture naming in 14 European languages. Behavior Research Methods,50(*2), 744–758. doi:10.3758/s13428-017-0902-x

Van Waes, L., Leijten, M., Mariën, P., & Engelborghs, S. (2017). Typing competencies in Alzheimer's disease: An exploration of copy tasks. *Computers in Human Behavior*, *73*, 311–319. https://doi.org/10.1016/j.chb.2017.03.050

Zannino, G. D., Perri, R., Salamone, G., Di Lorenzo, C., Caltagirone, C., & Carlesimo, G. A. (2010). Manipulating color and other visual information influences picture naming at different levels of processing: Evidence from Alzheimer subjects and normal controls. *Neuropsychologia*, *48*(9), 2571–2578. https://doi.org/10.1016/j.neuropsychologia.2010.05.003

Table 1. Overview of picture selection criteria 1 to 5.

		Highly Relatable	**Lower Relatable**
1.	Name Agreement (%)	>50%	>50%
2.	AoA (%)	Young (-6y)	Older (+6y)
3.	Frequency (%)	>30 per million words	<10 per million words
4.	Reaction Times (in ms)	631,5 < X < 981,34	652 < X < 1191,48
5.	Occurrence in other studies	Ignored	Accounted for

Table 2. An overview of the selection criteria based on variable 7.

	Highly Relatable	Lower Relatable
Living	Humans – 0	Humans – 5 (*4)
	Animals – 5	Animals – 10
Non-living	Vehicles – 5	Vehicles – 5
	Objects – 10 (*9)	Objects – 10

(Due to pre-testing results we deleted two images; the affected categories and their new image count are marked with a ‘*’)

Table 3. Description of the demographic data of the participants in mean (sd) on the first contact moment and three months later.

	Moment 1 (*n* = 58)		Moment 2 (*n=60)*	
	AG 50 - 64 (n = 34)	AG 65+ (n = 24)	AG 50 - 64 (n = 30)	AG 65+ (n = 30)
Age (years)	56.88 (*3.22*)	70.63 (*5.36*)	56.80 (*3.20*)	69.77 (*4.96*)
Years of education	14.94 (*2.86*)	13.54 (*2.75*)	15.33 (*2.58*)	13.17 (*2.60*)

Table 4. The selected naming accuracy and latency variables for the analysis of written picture naming.

Variables	Description
Name agreement	The percentage of words that are named correctly in a strict sense, not including variants or synonyms.
Object recognition	The percentage of words that are named correctly in a wider sense: including synonyms, morphological variants (e.g.: dolphin – dolphins), more specific names (e.g.: Dalmatian instead of dog) and more general names (e.g.: boat instead sailboat)
Naming latency nouns[1]	The onset reaction time needed by the participant before typing the target noun. This time (in ms) is the ‘x’ in the examples below
Interkey latency[2]	The latencies between the letters of the target noun, divided by the number of letters to account for differences in word length. This can be exemplified by the latencies between the letters for the noun ‘ball’ (= y1 + y2 + y3) versus ‘wheelbarrow’ (= y1 + ... + y10); those results cannot be compared due to the number of letters that differ. In order to make those results comparable, the interkey latencies must be calculated divided by the number of letters, which leads to an interkey latency for “ball” of ((y1 + y2 + y3)/3) ms and for “wheelbarrow” of ((y1 + ... + y 10)/10) ms.
Examples	_b_a_l_l (x y1 y2 y3) _w_h_e_e_l_b_a_r_r_o_w (x y1 y2 y3 y4 y5 y6 y7 y8 y9 y10)

[1] Inputlog terminology: before word pause

[2] Inputlog terminology: within word pause

Table 5. The percentage of words on the total of 5469 words - per score, test moment and age group.

		AG 50 - 64		AG 65+		
		Moment 1	Moment 2	Moment 1	Moment 2	*Total*
1	Named correctly	23.42	19.91	16.15	16.95	*76.43*
2	Named correctly - with typo correction	2.10	1.98	1.24	1.35	*6.67*
3	Named correctly – containing typo	0.84	0.71	0.75	0.49	*2.80*
4	Synonyms	2.58	2.08	2.21	2.03	*8.91*
5	Empty answers	0.15	0.46	0.33	0.35	*1.28*
6	Incorrect names	1.02	0.38	1.63	0.88	*3.91*

Table 6. Overview of the naming accuracy results.

	HR	**LR**	***p***
Name Agreement (%) (N = 4759)	82.47 (15.18)	84.54 (13.78)	0.627
Object recognition (%) (N = 5246)	95.09 (1.28)	91.73 (4.93)	0.006*

Table 7. Name agreement scores both within and between the two age groups in %.

	AG 50 - 64	AG 65+	*p*
HR	81.33 (*14.33*)	83.82 (*16.87*)	0.120
LR	86.32 (*12.55*)	82.44 (*16.44*)	0.025
P	0.065	0.099	

Table 8. Object recognition between the two age groups in %.

	AG 50 - 64	AG 65+	*p*
HR	94.16 (*1.15*)	96.20 (*2.43*)	0.004*
LR	92.30 (*3.39*)	91.06 (*7.84*)	0.304
P	0.053	0.002*	

Table 9. A measure of typing speed in ms.

	AG 50 - 64	AG 65+	*p*
Moment 1	207 (*53*)	240 (*84*)	0.242
Moment 2	207 (*52*)	240 (*85*)	0.173

Table 10. The naming and interkey latencies in ms contrasting the two age groups and image categories.

		AG 50 - 64	AG 65+	*p*
Naming latencies	HR	1611 (833)	1901 (1506)	0.000
	LR	1845 (1322)	2261 (1747)	0.000
Interkey latencies	HR	212 (140)	241 (137)	0.124
	LR	254 (107)	276 (141)	0.072

Table 11. Statistics per age group, per moment, per frequency type in ms.

			Moment 1	Moment 2	*p*
Naming latencies	AG 50 - 64	HR	1728 (891)	1769 (2181)	0.390
		LR	1961 (1422)	1844 (1003)	0.264
	AG 65+	HR	2166 (1751)	2004 (1564)	0.071
		LR	2635 (2784)	2542 (4158)	0.057
Interkey latencies	AG 50 - 64	HR	207 (90)	217 (180)	0.135
		LR	260 (115)	247 (96)	0.951
	AG 65+	HR	244 (115)	239 (151)	0.026
		LR	283 (139)	271 (142)	0.013

Figure 1. The dolphin (dolfijn) ©

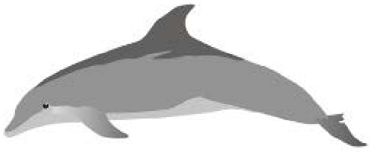

Figure 2. The shark (haai) ©

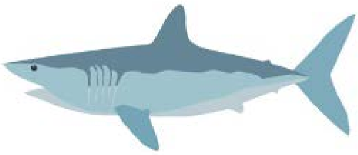

Figure 3. The pirate (piraat) ©

Figure 4. The correlation between name agreement and naming latencies

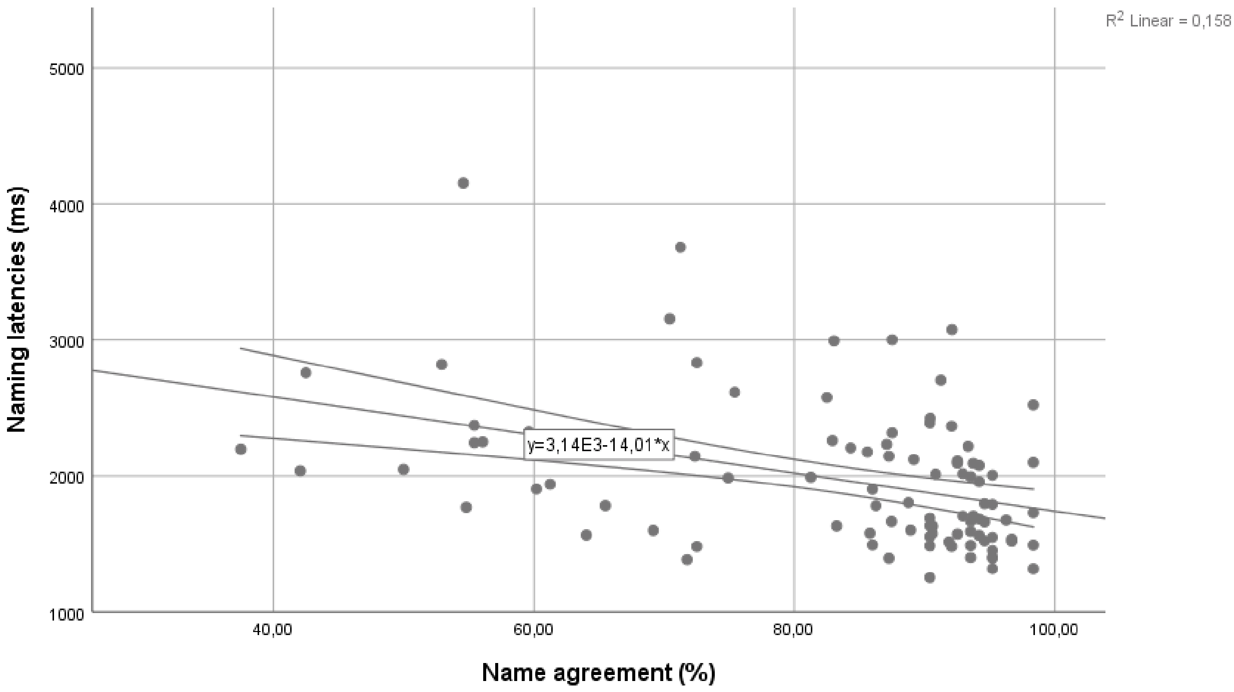

Figure 1. The dolphin (dolfijn) ©

Figure 2. The shark (haai) ©

Figure 3. The pirate (piraat) ©

Figure 4. The correlation between name agreement and naming latencies