Puzzling Out the Best Method

On the Influence of Medium and Researcher Presence

A bachelor thesis for Linguistics
consisting of 8888 words
by Mattanja Kalle, 4162560
guided by Shalom Zuckerman
and also checked by Brigitta Keij

Summary: Zuckerman (2015) came up with a new visual word recognition method: the word search puzzle. He conducted his experiment on paper. In this thesis, the effects of medium (paper or computer) and presence of the researcher (present or absent) are investigated in order to check whether further research can be performed digitally, as that would yield many advantages. It turns out that conducting the experiment digitally without a researcher present is not less good than on paper with a researcher present, as long as a couple of aspects are taken into consideration. A few improvements to the method and the digital version are suggested as well.
Key words: visual word recognition, medium, presence, age-of-acquisition, frequency, concreteness

1 Introduction

Many psycholinguists are interested in visual word recognition as a way of discovering how we process language and therefore have used various types of word recognition experiments, such as immediate and delayed naming, lexical decision, semantic categorisation and eye-tracking (e.g. Morrison & Ellis, 1995; Brysbaert, Lange & Van Wijnendaele, 2000; Bowers, Davis & Hanley, 2005; Vitu, O'Regan & Mittau, 1990; and Juhasz & Rayner, 2003). In many of these experiments, however, the words are presented in isolation. Of course, this does not render significant results useless, but we have to admit that the setting is quite artificial: in a way, there is a context of there not being any lexical context. At least for certain types of research it would be interesting to look at a more natural setting of reading as well. Yet, handing the participants a complete context could have unintentional consequences, as texts easily can include interfering aspects you are not aware of. A text of consisting of nonwords could be helpful in this respect, but encountering a real word would then be unusual, causing the results to be influenced by the unexpectedness of encountering such a word. Luckily, there is a simple way of introducing real words in a context without a meaningful lexical context in which participants will not be surprised to encounter these real words: a word search puzzle (Zuckerman, 2015).

Yet, before going into more detail, it is useful to point out the general structure of this study. In the next section, the theoretical groundwork will be laid out. Subsequently, the methodology will be explained and the following section will be dedicated to the results. After that, the consequences of those results will be examined in the discussion, whereupon the conclusions will be summarized in the final section.

2 Theoretical framework

2.1 Previous word recognition methods

In order to fully appreciate the benefit of this specific visual word recognition method, it is of course necessary to first review previous ones. Therefore, this section will include a short summary of immediate and delayed naming, lexical decision, semantic categorisation and eye-tracking.

2.1.1 Immediate and delayed naming

The idea of immediate naming, a method used by Morrison and Ellis (1995) and Brysbaert, Lange and Van Wijnendaele (2000), among others, is that the participant is presented with words in isolation and is then to directly read them out loud as fast as possible. The time it takes the participant to respond to the appearance of the word, the response time, is taken as a measure for how fast the participant processes and thus recognises the word. The procedure for delayed naming is exactly the same, except for a short delay of reading the stimulus out loud: the participant has to wait with giving his answer until a certain cue is given. Then, the response time is measured from cue to response, indicating the time it takes to produce the word, but not to access it, as this will already have taken place before the cue is given. Therefore, the difference in results between immediate and delayed naming should indicate whether differences between words are due to the processing or producing of the word. Using this method, Morrison and Ellis (1995) found an AoA effect, but no frequency effect. However, Brysbaert, Lange and Van Wijnendaele (2000), did find both an AoA as a frequency effect.

2.1.2 Lexical decision

In a lexical decision task, also carried out by Morrison and Ellis (1995) and Brysbaert, Lange and Van Wijnendaele (2000) and many others, the participant has to identify the stimulus presented to him (in isolation) as either being an existing word or a nonword. This ensures that the participant has to actually access the word in his mental lexicon, to be certain the accessing is included in the response time (as with immediate naming, measured from the moment the stimulus appears to the moment the participant responds). Again, the response time is a measure for the speed at which the participant has recognised the word.

With the lexical decision task, Morrison and Ellis (1995) found both an AoA as a frequency effect, as did Brysbaert, Lange and Van Wijnendaele (2000).

2.1.3 Semantic categorisation

Semantic categorisation, used by e.g. Bowers, Davis and Hanley (2005), is essentially the same as the lexical decision task, but now the participant is not asked to decide whether the stimulus is a word or nonword, but rather in which semantic category it belongs (e.g. natural vs. artefact). In this way, the participant has not only recognized the word as existing in his mental lexicon, but also has to have accessed its meaning. The difference in processing time between different stimuli is, once more, the response time.

Through this method, Bowers, Davis and Hanley (2005) found out that neighbourhood size (thus the amount of words similar to the stimulus, only differing from it with one letter) influences visual word recognition: more neighbours and thus more competitors, makes it harder to recognize a word.

2.1.4 Eye-tracking studies

It is also possible to research word recognition with eye-tracking, as did Vitu, O'Regan and Mittau (1990) and Juhasz and Rayner (2003), among others. As the eye-tracking device registers the location of the gaze, it is not necessary to present the words in isolation, although it is common to do so, in order to eliminate other possible influences.

With the eye-tracking, there are various possible measures, such as the first-fixation duration is the time the participant first focuses on the word, the single-fixation duration is the same as the first-fixation duration, but only for words only once being focused on, the gaze duration is the total of fixation on the word while reading it for the first time and total-fixation is the total amount of time spent on reading a word, including rereads.

Vitu, O'Regan and Mittau (1990) found effects of word length and frequency on gazing time, both with words in isolation as during reading and Juhasz and Rayner (2003) found effects of frequency, subjective familiarity, word length, concreteness and AoA fixation durations.

It can thus be concluded that there are visual word recognition methods that are considered reliable and that have proven effects of AoA, concreteness, frequency and length, which are factors that will be looked into in this research as well.

2.2 The word search puzzle

The design of the word search puzzle method is rather simple: you hand the participants a grid filled with letters, of which some constitute words and the rest are just random letters. Essentially, it's a word search puzzle, but without the list of words they are supposed to be looking for. All the participants have to do, is finding as many words as possible in a limited amount of time. As soon as they spot a word, they have to select it. Not all participants will recognize all words (at least within a certain time limit), either because they simply don't know it is a word (as with lexical decision), or because they fail to notice the word. The general tendency is that the more familiar[1] the word, the faster it is recognized (Morrison, Ellis & Quinlan, 1992; Morrison & Ellis, 1995; Brysbaert, Lange & Wijnendaele, 2000; Zuckerman, 2015), so, for instance, effects of age-of-acquisition (AoA), frequency and concreteness are expected to arise.

Zuckerman (2015) has tested the word search puzzle method for those effects by calculating the detection rate, the so-called *d-value*, of each word and comparing them to the corresponding AoA-, frequency- and concreteness-values. While the development of this method is still in its infancy, the method has already been proven to work: he found significant correlations of d-value with word frequency, age-of-acquisition, word length and concreteness.

Still, there is more to this method, because it allows for a variety of uses. First of all, it is possible to score which words each participant has found, as did Zuckerman (2015). Then it should be easy to test for potential influences on word recognition, as it would result in significantly more (or less) findings of certain words. These potential influences can be word-specific, such as the abovementioned AoA, frequency, length and concreteness (e.g. Zuckerman, 2015), but also participant-specific, such as age, gender, education, etc. (Keuleers, Stevens, Mandera & Brysbaert, 2015). Secondly, the position of the words can be chosen specifically as well, in order to research, for example, the segmentation of text into processing units: would a compound word or its components be found first and what happens if two words overlap? Would all words be found, or would the finding of one word inhibit the finding of the other word? Thirdly, if the order in which the words are found and the amount of time that has passed in between are recorded, priming effects could be studied more extensively. As the position of the relevant words can be modified, the researcher could determine the exact distance between prime and target. In conclusion, the combination of all these possibilities with its biggest strength, the absence of a lexical context without presenting words in complete isolation, make this method a relevant addition to the research of visual word recognition.

2.3 The medium

Although solving such puzzles is increasingly done digitally, it is probable that many participants still are more used to and possibly even prefer a paper version. Most of the discussed possibilities of this research method are achievable for the paper version, except for measuring the exact moment at which a participant finds a word. On a computer,

[1] In this case, *familiar* is used to generalise over possible factors: *known longer* (lower age-of-acquisition), *encountered more often* (higher frequency), *more 'real'* (higher concreteness), *easier to visualize mentally* (imageability), etc.

however, it is not that difficult to start a timer and mark the moment at which a word is found. A digital version would have even more advantages, for example, automatically gathering the data in an Excel-file and a built-in time limit, immediately disabling the possibility of marking words. If the puzzle is timed by hand, a few moments may pass before the participant has noticed the time limit has passed and stops marking words.

As said before, Zuckerman (2015) has proven that his method works and even though this new method still needs to become more and more refined, its first results look promising. The refinement of aspects such as the distance between de characters, the type of the gridlines and the size of the word search puzzle would be easier to figure out if the less time-consuming, digital version of the experiment could be used and thus studying the possibilities of a digital version has priority over the other facets.

Since Zuckerman's research (2015) was carried out using only printed word search puzzles, it is necessary to investigate the possible influences of the type of medium prior to replacing (or combining the use of) the printed version with the more efficient digital one. For instance, Shibata, Takano & Omura (2014; Takano, Shibata, & Omura, 2015) found that their participants (age 20 to 40) read faster from a paper than from a screen, but this is mainly due to moving the papers of a document. Furthermore, they point out that with paper, participants use their finger to indicate where they are in the text more often than when reading on a computer, causing them to lose track in digital documents more often. Moreover, Podolsky and Soiferman (2014) found a strong preference for reading off of paper, even though participating students use digital devices frequently. Their reasons were that adding notes is easier, their eyes and head get less sore, they are used to it, they can keep track of the lines more effortlessly and they were less distracted.

With a word search puzzle, there is no need to move pages or add notes and because of the short time limit, the strain on eyes and head will not be of relevance, but keeping track of where you are and less distractions may cause paper to be preferred. Therefore, it should be checked whether the medium influences the results of the word search puzzle, before the digital version can be accepted as a useful equivalent to the printed one.

2.4 Presence of the researcher

If approved, we would like to take full advantage of the digital version and not have the requirement of meeting each participant, so we could gather a lot of data in just a short amount of time by simply distributing the link to the puzzle. In that case, it is necessary to see if there is any difference between presence and absence of the researcher. According to Webster (1997), the presence of the researcher does have an effect on survey responses: the surveys were filled in more completely (even more so when the importance of the survey for either the society or a specific person ('sponsor') or group was emphasized), but the participants also tended to give more socially desirable answers. With the word search puzzle method, however, there are no 'socially more desirable' answers, unless socially undesirable words are included in the word search puzzle. Nonetheless, it could still determine how seriously the participants fill in the puzzle and as research would fall into the category of the social utility or help-the-sponsor appeal, whether the researcher is present should definitely be taken into account.

2.5 Generation of the participant

Not only the setting of the experiment will be studied: the data will also be used to gain some insights into participant-specific influences on visual word recognition: the effects of generation and gender will be considered. Especially because digital versions will be used, it is reasonable to suspect that the generation may have some influence. When the participants are split up in adolescents (< 25), adults (25 - 39), mature adults (40 - 64) and seniors (> 65), we have the youngest generation, having grown up with modern technologies, the adults, who witnessed the growing use of computers, the mature adults, who have started using computers later in life and the oldest generation, who may not be completely competent to use the modern technologies to their full extent. Also, the influence of AoA may differ for the generations.

2.6 Gender of the participant

As said before, differences due to gender will be checked as well. Gender has shown to affect the vocabulary, as reported by Keuleers, Stevens, Mandera and Brysbaert (2015), who have researched many different factors. They did find a difference in vocabulary size, where the men had a larger vocabulary by 0,5% on average, but with the associated effect size being very small. Yet, the words found by each gender did differ clearly, with differences in the percentages found per gender up to 40% (as reported on by marc, 2014, June 13).

The word search puzzle method does not rely solely on the knowledge of the word, but also on recognizing it within the time limit. Therefore, the difference in the size of the vocabulary may not matter as much for this research, but the words found by the participants of each gender may indeed differ.

2.7 Research questions and hypotheses

This study will be centered around the two context-related factors: the medium (on paper or digital) and the presence of the researcher (present or absent). In order to determine whether these factors yield any significant effect, the results of the four conditions (*paper-and-researcher-present*, PP; *paper-and-researcher-absent*, PA; *digital-and-researcher-present*, DP; and *digital-and-researcher-absent*, DA) will be compared. If both factors do not influence the results significantly, future research can be conducted by simply sharing a link to the experiment. Therefore, the two main research questions are:

RQ1: Are there any important differences between the conditions?
RQ2: In what way could the word search puzzle method be further improved?

If there would be any differences, the abovementioned articles would suggest a preference for paper as well as presence of the researcher, but most of the influencing aspects do not apply to this experiment. Because of the nature of the second question, it is not yet possible to make any predictions with regard to the improvement of the method.

In order to answer the main questions, several aspects will be studied: the number of words found, the effects of participant- and word-specific factors and whether there are words that are specific to a subset (the conditions, versions, generations and genders).

To investigate the number of testwords found, one-way ANOVA and two-way ANOVA with Fisher's LSD post hoc test will be used, except for when there are only two variables, in which case a t-test will be performed. For the amount of words found, the following subquestions are formulated:

SQ1: Is there any difference in the number of words found between the conditions?
SQ2: Is there any difference in the number of words found between the versions?
SQ3: Is there any difference in the number of words found between the generations?
SQ4: Is there any difference in the number of words found between the genders?

If there would be any difference, it is possible more words will be found on paper than digitally, because it would be the preferred medium according to abovediscussed articles, but the conditions are not expected to vary significantly. There are not supposed to be any real differences between versions and no distinctions for generations are suspected either, although the possibility is not rejected. If there would be any contrast, the oldest generation may find less words, due to potentially reduced eyesight and hence slower reading. With regard to gender, the women could very well find more words, as they are supposedly better at language.

To look into the influence of the participant-specific factors (condition, version, generation, gender and location) and word-specific factors (AoA, frequency, concreteness and length), generalized linear mixed models will be run. The following subquestions concern these word-specific factors:

SQ5: Which factors influence word recognition?

It is expected that the word-specific factors, AoA, frequency, concreteness and length will influence word recognition, but most of the participant-specific factors, condition, generation and gender, not. Location will of course be a predictor variable, but when combined with version, the effect should be (at least partly) cancelled.

To find any words specific to a certain subset, a partial correlation of the d-values of the words with the context-related (condition and version) and participant-specific (generation, gender and testword count) factors will be calculated for each testword. With regard to this aspect, the subquestions are the following:

SQ6: Are there any words that do not correlate with the testword count?
SQ7: Are there any words specific to a certain condition?
SQ8: Are there any words specific to a certain version?
SQ9: Are there any words specific to a certain generation?
SQ10: Are there any words specific to a certain gender?

It is not likely that any word should not correlate with the testword count, as finding more testwords increases the chances of finding each testword. It is not expected either that there are words specific to any condition or version, but certain words might be found more by one of the generations. Regarding gender, ten words that were explicitly chosen to be either more feminine or masculine were included and it is probable for at least those words to be specific to a gender.

The next section will discuss the method in more detail.

3 Method

3.1 Pilot

Before the real experiment was carried out, a pilot[2] was conducted by students as assignment for a psycholinguistics course. It was essentially the same as the experiment itself and the

[2] The pilot was intended to be part of the data gathering process of the experiment, but as a result of not all students fully understanding the design of the experiment, the reliability of the results was compromised and thus it was decided to not include them as part of the real experiment, but only as a pilot. For example, the different conditions did not receive equal amounts of participants (PP: n = 65;

results were used to check for certain aspects of its design. From the results, a couple of conclusions could be drawn.

A two-way ANOVA revealed no main effect for medium (F(1,151) = 0.52; p = .474) or for researcher presence (F(1,151) = 0.00; p = .955), nor was an interaction-effect of medium and presence found (F(1,151) = 3.59; p = .060). The means and standard deviations for medium and researcher presence can be found in **table 1**. These results suggest that at least for the number of testwords found, the conditions are equivalent.

Table 1
Means (Standard Deviations) per Medium and Researcher Presence

	Researcher present		Researcher absent	
Medium	N	M (sd)	N	M (sd)
Paper	65	18.58 (6.05)	15	20.87 (7.65)
Computer	22	20.05 (6.87)	53	17.62 (6.51)

A one-way ANOVA showed no main effect of version on the number of testwords found (F(3,151) = 1.39; p = .249). The means and standard deviations for the different versions can be found in **table 2**. These results suggest that the versions are sufficiently similar to not interfere with the number of testwords found.

Table 2
Means (Standard Deviations) per Version

Version	N	M (sd)
A	38	20.24 (7.45)
B	39	19.13 (5.95)
C	39	17.82 (5.42)
D	39	17.59 (6.51)

Another two-way ANOVA[3] revealed a main effect for gender (F(1,148) = 4.68; p = .032) and generation (F(2,148) = 4.74; p = .010) on the number of testwords found. The effect of generation lies between the adolescents and the mature adults (p = .015). However, there was no interaction-effect of gender and generation (F(2,148) = 0.87; p = .420). The means and standard deviations for gender and generation can be found in **table 3**. These results suggest that differences between genders and generations are worth looking into.

Table 3
Means (Standard Deviations) per Gender and Generation

	Men		Women	
Generation	N	M (sd)	N	M (sd)
Adolescents	41	18.39 (5.16)	78	20.22 (7.19)
Adults	7	15.86 (4.98)	3	19.67 (8.02)
Mature adults	10	12.30 (2.67)	15	17.73 (6..89)

Various generalized linear mixed models (GLMM) were run, in order to find out what factors, both word-specific and participant-specific, influenced the probability of the word being found. The fixed factors of the different models were AoA, concreteness , frequency and

PA: n = 15; DP: n = 22; DA: n = 53) and some remarks of participants suggested that they had received the list of words that were placed in the puzzle, so they could find the words as if it were a normal word search puzzle, instead of relying on ‘accidentally’ spotting the words.

[3] The data of the participant who fell into the senior generation (N = 1) were excluded from all analysis (concerning a categorisation by generation) for its lack of members.

length for the word-specific ones and location, condition (medium*presence), version, generation and gender for the participant-specific factors, and as location surely will be a predictor variable, the combinations with location and the other participant-specific factors are included as well (condition*location, version*location, generation*location and gender*location).

The one with location and version as the fixed factors and participant and testword as the random factors is the GLMM that best fits the data ($F(40, 6.16) = 3.34$; $p < .001$). This means that the words higher up in the puzzle indeed are more likely to be found than those near the bottom, but this effect is partly cancelled by the use of different versions, thus proving that the use of the different versions helps controlling for the location-effect.

It is, however, unexpected that there are no effects of AoA, frequency, concreteness and length, although this could be due to the limited amount of data. Nonetheless, not finding an effect of condition is desirable.

Because of the compromised reliability of the pilot, no partial correlations were performed to look for outliers, as the differences could easily be due to sampling fluctuations or interference of malexecution of the experiment. However, some other observations are worth mentioning.

One is that the amount of times the word *uur* (meaning *hour* in Dutch) was found, was very low (it is found by only 10.3%, whereas 46.7% is the average *d-percentage*[4] and the second lowest percentage being the one of *snor*, meaning *moustache*: 18.7%), probably due to the random letter *v* standing in front of it, constituting the word *vuur* (meaning *fire*; found by 28,4%). For the real experiment, the matrices should thus be checked more carefully for these accidental words overlapping with the intended words.

When it comes to the time limit, three minutes appears to be a valid time limit, as the participants had enough time to find quite a lot of words, while not finding every word. The average amount of words per participant is 18.68 (sd = 6.51), with a minimum of 6 and a maximum of 37. The participants have to find enough words to make sure that if they do not spot a certain word, it means something, while at the same time they can't have enough time to spot all the words, because in that case, we would not have obtained any useful data. This method relies on the balance between finding certain words and not others.

For the experiment, new versions of the pilot word search puzzles (version A1, B1, C1 and D2) were made (version A2, B2, C2 and D2) with other randomly generated letters in between, checked carefully for the presence of random letters interfering with the finding of the testwords (such as *v* in front of *uur*). All the letters that either constituted a word with the whole testword or a part of it, were then, of course, replaced with random letters that did not cause any interference.

3.2 Participants

A total of 181 participants responded, of which 69 were male and 112 were female, all within an age range of 14 to 82 . In order to make sure none of the participants knew what the test was actually measuring, linguists, linguistic students and other people who knew the goal of this experiment, were not allowed to participate. With the digital word search puzzle, not all trials were fully completed. Only the data of fully completed digital tests were included.

---

[4] What I will refer to as *d-percentages*, short for *detection-percentages*, named analogous to *d-values*, are the percentages of the d-value relative to the amount of times a testword could have been detected, calculated over all the data or over a certain subset (e.g. the different conditions, versions, generations or genders, or any other sort of subset).

It was intended to have equal amounts of men and women per version (A2, B2, C2 and D2) per condition. For both paper conditions, this was more or less achieved, but this was not the case for the digital ones. To obtain enough participants for the DP condition, the experiment was carried out in an office location, so multiple computers were available at the same time, while the researcher was present. Unfortunately, most employees working in that office were male. For the DA condition, however, most participants were female.

Furthermore, as the version was assigned automatically (with equal weight per version), the assignment of the versions was not manipulated over factor (present or absent) or gender. The full distribution of participants can be found in **appendix A**.

3.3 Materials

The set of words in the word search puzzle consisted of words of which AoA, frequency and concreteness values were available. Both the AoA and concreteness values were obtained from Brysbaert, Stevens, De Deyne, Voorspoels & Storms (2014) and those of the frequency and length came from the SUBTLEX-NL database (Keuleers, Brysbaert & New, 2010). For AoA, words were only selected if the AoA had been based upon at least two studies, of which none were marked as known by less than hundred percent of children of that age. The AoA, frequency, concreteness and word length had to be spread across the range rather equally. Of this data set, 30 words were semi-randomly[5] selected.

In order to enlarge a possible effect by gender, 10 words (5 considered more masculine, 5 more feminine) were added to result in the total of 40 word search puzzle words. They were selected as follows: of the same combined data set with AoA, frequency, concreteness and length values, another 117 words were selected, which then were rated on masculinity or femininity by six participants (a man and woman for each of three generations), with a score of 1 being most masculine and 5 being most feminine. It was calculated which words were considered most masculine and most feminine and of both categories, five words were chosen that matched with the words of the other category in terms of AoA, frequency, concreteness and length. A list of the 40 selected words, from this moment onwards referred to as *testwords*, can be found in **appendix B**, including a translation, the word-specific factors and an indication for the possibly gender-specific words whether they are considered more masculine or feminine.

Then, the 40 words were horizontally placed in a letter grid, completed with randomly generated letters. Four different versions (A2, B2, C2 and D2) of this letter grid were made, only differing in the order of the rows, in order to account for the location of the words, as most participants read left-to-right, top-to-bottom. All versions can be found in **appendix C**.

Of the very same versions, a digital version was made by Hans Kalle. Before the experiment was started, the usability was tested by several of his fellow software developers and a few non-programmers as well. The instructions for both the digital and paper versions were more or less the same, but the digital one was slightly more elaborate: it also included more specific instructions regarding how to fill in a digital version, as well as a test word search puzzle to practice selecting words.

---

[5] First, the dataset was reduced randomly by only selecting every thirtieth word or so, except for words with a low AoA, as there are less of those words and even less are still used at a later age. For instance, the word *ikke* (meaning *I*) is only used by young children that cannot yet pronounce *ik* (the real word). For the low AoA words, not many were suitable, so they had to be hand-picked, not so very randomly. For the higher AoA words, however, there were more options to eliminate a certain word if it was a compound word consisting of words that could be mistaken as separate words etc. Notwithstanding, it was almost impossible to find longer words that were not compound words.

3.4 Procedure

To obtain enough participants for the first three conditions (PP, PA and DP), I went to places where I knew were people willing to participate, so the experiment was conducted at various locations, varying from an office space to their own homes. This should not be a problem, since the experiment in the DA condition would not be performed in a laboratory setting either, as the participants can complete test on whatever locations they would choose. The instructions for the paper and digital versions were kept as similar as possible. The instructions can be found **appendix D**.

For the PA condition, I would sent them away with the paper version (including the written instructions) and a timer, but for the other two conditions, the participants would fill in the word search puzzle, either on paper or on my laptop, while in my presence.

If any participant had questions about what was considered a correct word, the only comment I made, was that they should mark a word as soon as they saw it and considered it a word, to ensure I did not influence their 'marking behaviour' in any way.

To transform the spreadsheet from the experiment into the Excel-files to be imported in SPSS, a Jupyter[6] notebook were used. It was used to exclude non-completed sessions and a few strange sessions, e.g. the session of a participant who selected the words letter by letter, resulting in strings of just two letters and missing first or last letters of words with an odd number of letters. In addition, the data was merged into one matrix with an entry for each testword per participant, marked for being found or not. All word-specific and participant-specific data was included in every entry. For more detail, the complete notebooks can be found in **appendix E**.

Presently, we will analyze the results.

4 Results

4.1 Number of testwords found

The number of selections made by the participants made may include some combinations that are no real words. For a selection to be counted as a testword, it had to be the exact same string. As soon as a letter was missing (e.g. *nachtegaa* instead of *nachtegaal*) or another was included (e.g. *klimmenr* instead of *klimmen*), it was not counted as a testword, because we cannot make such assumptions, especially in the cases that were less clear. Besides, most of the times, the participants re-selected the correct word.

Before looking into the differences of the number of selections and testwords found for various factors, it is useful to state the minimum, maximum and mean scores, which also can be found in table 3: no one found less than 8 or more than 36 testwords (M = 21.21; sd = 6.54). The number of selections found is higher, with a minimum of 12, a maximum of 56 (M = 33.01; sd = 8.75), because many participants could find words within the testwords or in a string of random letters.

A two-way ANOVA revealed a main effect for medium ($F(1,176) = 6.76$; $p = .010$), but not for researcher presence ($F(1,176) = 2.37$; $p = .126$), nor was an interaction-effect of medium and presence found ($F(1,176) = 0.09$; $p = .769$). Thus it can be concluded that when

[6] Jupyter notebooks are files for programming with the programming language Python, which can easily be used to e.g. make calculations, transform spreadsheets and make various types of graphs. The most important advantages of using such a program to construct the needed spreadsheet, is easily keeping track of all the transformations and the simplicity of adjusting a step. If you have forgotten a certain step, you do not have to manually carry out all the other steps again: you simply add the step and run the programme on the original input again.

the experiment was carried out digitally, the number of testwords found was higher, but it is not dependent on the presence of a researcher. The means and standard deviations for medium and researcher presence can be found in **table 4**.

Table 4
Means (Standard Deviations) per Medium and Researcher Presence

	Researcher present		Researcher absent	
Medium	N	M (sd)	N	M (sd)
Paper	32	18.28 (6.91)	30	20.20 (5.07)
Computer	32	21.31 (6.26)	86	22.62 (6.64)

A one-way ANOVA showed no main effect of version on the number of testwords found ($F(3,176) = 0.26$; $p = .858$). The means and standard deviations for the different versions can be found in **table 5**.

Table 5
Means (Standard Deviations) per Version

Version	N	M (sd)
A	39	22.00 (6.47)
B	52	21.13 (7.48)
C	40	20.83 (4.84)
D	49	20.98 (6.88)

To check for gender and generation effects on the number of testwords found, another two-way ANOVA[7] was performed. It revealed no main effect for gender ($F(1,171) = 1.97$; $p = .162$), nor generation ($F(2,171) = 1.95$; $p = .145$). There was no interaction-effect of gender and generation either ($F(2,171) = 0.60$; $p = .550$). The means and standard deviations for gender and generation can be found in table 6.

Table 6
Means (Standard Deviations) per Gender and Generation

	Men		Women	
Generation	N	M (sd)	N	M (sd)
Adolescents	16	22.19 (6.69)	58	22.31 (6.41)
Adults	34	20.59 (5.79)	34	21.79 (7.33)
Mature adults	17	17.82 (5.25)	18	21.06 (6.91)

4.2 Effects of word- and participant-specific factors

As with the data of the pilot, various GLMMs were run, in order to find out what factors, both word-specific and participant-specific, influenced the probability of the word being found. The fixed factors of the different models were AoA, concreteness , frequency and length for the word-specific ones and location, condition (medium*presence), version, generation and gender for the participant-specific factors, with the various possible combinations of participant-specific factors with location as well, in order to control for the effect of the

[7] Of three participants, the age was marked as missing and thus their data was excluded from all analysis concerning a categorisation by generation. The data of the only participant of the senior generation was excluded from the same analyses as well, for its lack of members and thus not yielding meaningful results.

location of the testwords (condition*location, version*location, generation*location and gender*location).

The GLMM that fits the data the best, is the one with location and version as the fixed factors and participant and testwords as the random factors ($F(117, 7.04) = 5.28$; $p < .001$). This means that the words higher up in the puzzle are more likely to be found than those near the bottom, but the use of different versions partly cancels this effect.

The expected effects of AoA, frequency, concreteness and length were thus not found, and likewise, there were no effects of the different conditions.

4.3 Words specific to a subset

Partial correlations were performed to see how well finding each word correlates with the different participant-specific factors (generation, gender and number of testwords found) and the context-related ones (condition and version). When looking at the correlation of the findings of the words and one dependent variable, the other possible factors were controlled for.

It is no surprise that almost all of the words significantly correlate with the testword count, as finding more words immediately raises the chance of finding a specific word. Only the words *nat* ($r = .14$; $p = .065$) and *uur* ($r = .04$; $p = .594$) do not correlate.

With regard to the conditions, only *pijn, snor* and *xylofoon* yielded a significant result. When taking a look at the d-percentages per condition, it can be concluded that *pain* was found most in the DA condition and much less in the paper conditions (PP: 15.6%; PA: 13.3%; DP: 25.0%; DA: 33.7%), *snor* was found least in the DA condition and most in the PA condition (PP: 37.5%; PA: 46.7%; DP: 40.6%; DA: 31.4%) and *xylofoon* was found most in the DA condition and least in the PP condition (PP: 34.4%; PA: 56.7%; DP: 65.6%; DA: 77.9%).

When comparing the versions, it appears that *haast, hard, jaar, jullie, salon, schepper, school, voorzitter* and *xylofoon* correlate and thus are not as easily found in the different versions. The d-percentages of all the testwords for every version can be found in **table 7**.

Table 7
D-percentages of Significantly Correlating Testwords per Version

	Version A	Version B	Version C	Version D
haast	66.7%	50.0%	70.0%	75.5%
hard	69.2%	78.8%	45.0%	59.2%
jaar	71.8%	80.8%	65.0%	46.9%
jullie	56.4%	40.4%	75.0%	61.2%
salon	56.4%	50.0%	65.0%	71.4%
schepper	84.6%	65.4%	42.5%	85.7%
school	84.6%	90.4%	87.5%	61.2%
voorzitter	64.1%	63.5%	40.0%	40.8%
xylofoon	59.0%	53.8%	72.5%	73.5%

No words appear to be significantly correlated with the generations and the only words that are specific to a gender, are *geweer*, *jurk*, *nachtegaal* and *salon*. According to the d-percentages, *geweer* is found more often by the women (men: 44.9%; women: 64.0%), *jurk* more often by the men (men: 55.1%; women: 42.3%), just as *nachtegaal* (men: 70.0%; women: 57.7%) and *salon* (men: 66.7%; women: 56.8%).

The results of all the partial correlations can be found in **table 8**.

Table 8
Words specific to a subset – partial correlation overview

	testword count	conditions	versions	generations	gender
alles	.45***	-.02	.08	.00	-.09
dansen	.37***	-.04	.01	.00	.08
geweer	.43***	.01	-.10	.09	.16*
haast	.42***	-.03	.16*	-.02	-.05
hard	.43***	.03	-.16*	.01	.09
idee	.18*	.02	-.11	.03	-.01
jaar	.26***	-.06	-.21**	.05	.06
jullie	.37***	-.03	.13*	-.03	-.10
jurk	.41***	.08	.14	.02	-.16*
kan	.25**	.04	.10	.00	-.05
kijk	.20**	.10	-.09	-.09	.14
klauw	.39***	.01	-.09	-.04	.10
klimmen	.37***	-.02	-.14	.03	.07
kom	.33***	.14	-.02	-.05	-.02
leger	.45***	-.09	.10	.17	.01
liter	.44***	.10	.09	-.00	.01
morgen	.30***	-.07	-.12	-.08	.14
nachtegaal	.37***	-.05	.13	-.13	-.23**
nat	.14	-.10	-.05	-.02	-.05
nek	.25**	.00	.07	.03	.04
pijn	.20**	-.16*	-.01	.06	.06
priester	.36***	.01	-.01	.06	.03
roman	.38***	.05	.03	.06	.06
salon	.50***	.04	.19*	.05	-.16*
schepper	.28***	-.14	-.01*	-.05	.03
school	.39***	.03	-.24**	-.04	.20
serie	.17*	.11	-.04	-.10	-.02
sjaal	.30***	-.06	-.11	.02	.11
snor	.41***	.17*	.11	.11	-.02
spijt	.38***	.14	-.06	-.04	.03
tabak	.35***	-.10	.03	.04	-.01
uur	.04	.00	-.06	.06	-.04
verminking	.41***	.04	.07	-.04	-.04
vervalsing	.42***	-.01	-.01	-.04	-.15
visum	.23**	.03	.04	.01	-.13
voorzitter	.26**	.00	-.20**	-.10	.09
weg	.33***	-.01	.09	-.02	.07
xylofoon	.33***	-.24**	.16*	-.14	-.09
yoga	.38***	.02	.02	.02	-.05
ziek	.44***	.04	-.01	.05	-.06

Note. * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Before drawing the final conclusion to the main research questions, the implications of the results will be discussed in the following section.

5 Discussion

5.1 Discussion number of testwords found

With respect to SQ1, there are differences between conditions in the amount of testwords being found. Apparently, more testwords are selected in the digital versions than on paper, which dismisses the hypotheses of no difference. Possibly, it is slightly faster to select words via a digital medium, as the manner of selecting words might be slightly more intuitive: dragging instead of circling. It could also be the result of not being able to select words vertically or diagonally in the digital version, as a few participants in the paper conditions did select a few words vertically or diagonally. This does not automatically render the DA condition as being less qualified for this type of experiment, but only shows another advantage of the digital version (being able to limit the directions in which words are selected) and that the time limit may need to be adjusted according to the condition. To prevent participants from finding too many testwords and therefore, not giving meaningful information on the differences between the words, the time limit should be reduced.

The answer to SQ2, that there are no significant differences between the versions, shows that at least in this respect, the versions are indeed equal. As expected, there are no significant differences for the generations either, which answers SQ3. For SQ4, the results are contrary to the hypothesis: the number of testwords found by women does not significantly differ from the amount found by men.

5.2 Discussion effects of word-specific factors

Concerning SQ5, it appears that location still influences the word recognition the most, even though controlling for the version does cancel part of the effect. In further research, this effect should be reduced even more, in order to better find other possible influences. The effect of location could be reduced by using more different versions or reducing the time limit so drastically, that the participants have to resort to “spotting” the words, instead of “reading” through the letter grid. However, the participant has to be aware of the very short time limit, because otherwise, the effect of location will only be even stronger, as most of the participants would still start reading and only cover a smaller part of the puzzle.

Contrary to the hypothesis, the word-specific factors do not appear to account for the variation in d-values between the testwords. This could be due to the limited set of testwords used. When more words are used, the effect of the different word-specific factors could become clearer. However, it is not advisable to put more words in one letter grid, as this would leave less room for random letters in between and thus would raise the density of the testwords in the letter grid, causing it to lose the essence of the word search puzzle. Nonetheless, it is possible to include more testwords by creating word search puzzles with different sets of testwords.

Finding no effects of generation or gender does comply with the hypothesis for the participant-specific factors, as does finding no influence of condition, which supports the main hypothesis stating that the different conditions are equally suitable to be applied to the word search puzzle method. So, even though the effects of aoa, concreteness and frequency are not proven with this experiment, this experiment shows that the digital version should not be any less good than the paper version and as those effects have been proven for the paper version before, it should be possible to replicate those effects with a digital version when carried out on a larger scale, regarding both the number of testwords and participants.

5.3 Discussion words specific to a subset

The words *uur* and *nat* do not correlate with the testword count, rendering them not fit as suitable testwords. For *uur*, it could be caused by the unusualness of starting with a diphthong consisting of twice the same letter or because of the many four-letter words of which the last three are *uur*. The four-letter words would, in a sense, be competitors to *uur*, making the participants (subconsciously) expect a consonant immediately preceding it with which it would constitute a word. This theory also fits with the much more frequent selecting of *vuur* than *uur* in the pilot version. Notwithstanding, no such a theory appears to apply to *nat* but that three-letter words are possibly less easily recognized in a letter grid, as the visual representation of the word is not as fixed or recognizable in a context of random letters as those of longer words. There is a higher chance that random letters generate three-letter words than longer words and thus the visual representation of the words, as stored in the mental lexicon, stand out less. Still, this theory does not fit with the results of *nek* and *weg*, two three-letter words that apparently, per contra, do correlate normally with the testword count.

When it comes to the words more frequently found in certain conditions (*pijn* in the digital conditions, *snor* in the PA condition in contrast to the DA condition, and *xylofoon* in the DA condition opposite to the PP condition), it seems most logical the difference would stem from their meaning, rather than other factors, as they seem to have no word-specific factors in common (as can be seen in **appendix B**). But even then, there seems to be no clear reason for the significant correlations: there are no clear links between their meaning and the setting in which they were found most or least often. Possibly, this effect will be diminished when the experiment is carried out at a larger scale, with more testwords.

Basically the same applies to the words found more or less often in the different versions (*haast, hard, jaar, jullie, salon, schepper, school, voorzitter* and *xylofoon*): there are appear to be no aspects they have in common, nor could their meaning be related to a certain version, as the setting is the same and only the order of the rows differs. There also does not seem to be a certain pattern in the locations of these words in the various versions. If this effect, then, is not diminished by executing this experiment at a larger scale, as was suggested when discussing the differences between the conditions, the only remaining explanation would be priming.

As expected, there are no words specific to a generation. Some are specific to a gender, but not in the way it was predicted: the words considered as more feminine are not specific to women and those considered more masculine are not specific to men. The only words found more often by one gender than the other are *geweer, jurk, nachtegaal* and *salon,* of which *geweer* was considered more masculine, but found more often by the women, and *jurk* and *salon* were considered more feminine, but found more often by the men. *Nachtegaal* was not selected as either feminine or masculine and was found most by the men. Interestingly, the words that were evaluated as specific to a gender were actually more easily recognized by those of the opposite gender. Perhaps, participants find words connected to the other gender more easily, because they encounter those words less often and thus are somehow more aware of reading them (as some sort of novelty effect). However, the other words considered as gender-specific, were not recognized more often by one of the genders.

5.4 Further use of the word search puzzle method

5.4.1 Difference between the conditions

The answer to the first main research question is not entirely consistent with the hypothesis. Based upon all the results, it can indeed be concluded that the DA condition does not seem to be a less suitable word recognition test than the PP one, even though the time limit may have to be shortened. It is also relevant to state that gathering the data for condition PP, PA and DP took about two weeks, while for the DA condition, it took less than five days and within the first three and a half hours after posting the link on only one social media website, 50,6% of the usable DA sessions was already completed.

However, it should be considered that conducting the experiment only digitally could cause a bias in the selection of participants, as Keuleers, Stevens, Mandera and Brysbaert (2015) have pointed out for their own experiment. Even though more potential participants may be reached, they may be primarily the more active users of internet and moreover, only who are interested in helping research will participate.

5.4.2 Improvement of the method

Regarding RQ2, various facets should be considered before further implementing the word search puzzle method. First of all, the role of location could be further decreased by using more versions, reducing the time limit drastically and conducting the experiment at a larger scale, which would be easier to accomplish with the digital version than the paper one. Repeating the experiment at a larger scale would also help to figure out whether the odd results with seemingly random words being found significantly more often in the different conditions or versions are random outliers or are consistently influenced by some underlying factor as priming, for example. At the same time, a larger set-up would cause a more equal distribution of participants over the different versions.

It might also be better to not include any three-letter words as a testword, as *nat* and *uur* strangely did not correlate with the testword count. Removing three-letter testwords from the experiment and informing the participants that they only have to look for words of four letters and more, could highly reduce the number of randomly generated words you have to eliminate from your grid.

Another aspect to be studied is the frequencies of specific letters, which was a strategy mentioned by a couple of the participants: they said they searched for vowels and then looked for words around it. In this way, a word with a diphthong may be more easily found than one with relatively many consonants. This could be based upon the frequencies of the individual letters and the general structure of a syllable of the language in question.

Furthermore, the time limit and its influence should be investigated: is a somewhat shorter time limit sufficient to reduce the effect of location, or is it better to have a radically shorter time limit. When the time limit is drastically shortened, to e.g. 30 seconds, it is even possible to let participants fill in several word search puzzles to acquire about the same amount of data in the same amount of time. The best time limit may also be dependent on what you want to research: if the goal is to reach great similarity to normal reading, a more extended time limit would suit the investigation better than a short one ensuring the participants rely on spotting the words rather than reading the grid.

5.4.3 Improvement of the digital version

Before future research is carried out digitally, there are still a couple of aspects that could be improved. For instance, it could be useful to let the participants indicate which type of device (computer or tablet) they are using.

Also, as the instructions may not be read as carefully as hoped, as becomes clear from the words being selected from right to left or even letter by letter, and as no researcher will be present to point the participant in the right direction, if need be, certain limitations may be implemented to ensure the experiment is used as designed. For example, instead of any three selections in the practice word search puzzle being sufficient to start the real one, a specific task may be made required. This would not only ensure the participant knows exactly how to select the words (instead of e.g. selecting a word letter by letter), but also that the instructions are read, because otherwise, the participant would not know how to continue to the word search puzzle. Furthermore, it may help to place the instructions on the left side of the practice word search puzzle, so it is seen first.

In addition, in order to prevent obscurity on whether the participant intended a selection to be right-to-left, or that the direction is just 'by chance', left-to-right could be made the only possible way of selecting a word. Right-to-left selection could be eliminated, similar to how vertical and diagonal selection were already made inexecutable.

As multiple participants tried to unselect a set of letters and some even expressed the wish for an undo-option, this option has to be considered. If such a function is added, participants do not have to spend time wondering how and if it is possible to unselect a word, but it may cause participants to spend time correcting their selections instead of spending it looking for words. Another option would be to just state in the instructions that it is not possible to unselect something and that it is sufficient to just select the correct word again.

After finishing the experiment, many participants were interested in their ranking based upon the number of words they found, as they were instructed to find as many words as possible. Adding such a score at the end of the test could increase the game-like sensation and it would give competitive participants the possibility to compare scores, which could stimulate sharing via social media and thus resulting in more participants. In order to accomplish this, however, the selections of the participants would have to be checked against a database of existing words, because non-words will have to be excluded.

6 Conclusion

As reviewed in detail above, it can be concluded that the experiment can be just as well conducted digitally, when a couple of aspects are taken into consideration. The use of this medium yields many advantages, as broadly discussed in the theoretical framework, such as the increased ease and speed for gathering the data, but also for the analysis, as the data is instantly available in a spreadsheet format.

The digital version, then, could be used to gather data more efficiently in order to research both improvements of the method (e.g. the best time limit and the influence of the frequencies of letters) and its applications (e.g. typical word recognition topics such as priming and the word-specific factors AoA, concreteness and frequency). To conclude, there is still a lot of interesting research to conduct with this new visual word recognition method: Zuckerman's word search puzzle.

Acknowledgements

This thesis could not have been completed with the much appreciated help from Shalom Zuckerman, who lead me through the whole process of writing this thesis. The great efforts of Hans Kalle are also highly appreciated, without whose programming skills the digital version of the experiment would not even exist. Furthermore, I would like to thank several of Hans

Kalle's colleagues and a couple of family members for testing the digital version on usability and of course all the participants, as without the data provided by them, none of this research could have been done.

Bibliography

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition*, *97*(3), B45-B54, DOI: 10.1016/j.cognition.2005.02.002

Brysbaert, M. (2014, June 13). Words known by men and women. Message posted to http://gnodevel.ugent.be/crr.ugent.be/archives/1628

Brysbaert, M., Lange, M., & Wijnendaele, I. V. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition: Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, *12*(1), 65-85, DOI: 10.1080/095414400382208

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. Data file retrieved from http://crr.ugent.be/archives/1602.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1312, DOI: 10.1037/0278-7393.29.6.1312

Keuleers, E., Brysbaert, M. & New, B. (2010). SUBTLEX-NL: A new frequency measure for Dutch words based on film subtitles. *Behavior Research Methods, 42*(3), 643-650. Data retrieved from http://crr.ugent.be/programs-data/subtitle-frequencies.

Keuleers, E., Stevens, M., Mandera, P. & Brysbaert, M. (2015) Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *The Quarterly Journal of Experimental Psychology, 68*(8), 1665-1692, DOI: 10.1080/17470218.2015.1022560

Garrod, S. (2006). Psycholinguistic Research Methods. Keith Brown (Ed.) Encyclopedia of Language and Linguistics, (251–257).

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 116, DOI: 10.1037/0278-7393.21.1.116

Morrison, C. M., Ellis, A. W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, *20*(6), 705-714, DOI: 10.3758/BF03202720

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic bulletin & review*, *13*(1), 45-52.

Podolsky, T., & Soiferman, K. (2014). Student Academic Reading Preferences: A Study of Online Reading Habits and Inclinations. *Online Submission*.

Shibata, H., Takano, K., & Omura, K. (2014). Comparison of paper and computer displays in reading including frequent movement between pages. In *Proceedings of the 26th*

Australian Computer-Human Interaction Conference on Designing Futures: the Future of Design (pp. 549-558). DOI: 10.1145/2686612.2686700

Takano, K., Shibata, H., & Omura, K. (2015). Effects of paper on cross-reference reading for multiple documents: Comparison of reading performances and processes between paper and computer displays. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction* (pp. 497-505). DOI: 10.1145/2838739.2838745

Vitu, F., O'regan, J. K., & Mittau, M. (1990). Optimal landing position in reading isolated words and continuous text. *Perception & psychophysics*, *47*(6), 583-600, DOI: 10.3758/BF03203111

Webster, C. (1997). Effects of researcher presence and appeal on response quality in hand-delivered, self-administered surveys. *Journal of Business Research*, *38*(2), 105-114, 10.1016/S0148-2963(96)00028-8

Zuckerman, S. (2015). The Word Search Puzzle method for word recognition. Paper presented to: Experimental Psycholinguistics Conference, 7-9 October 2015, Madrid.

Appendices

A. Distribution of participants

Distribution of participants

	PP					*PA*					*DP*					*DA*					**Total**
	A	B	C	D	tot	A	B	C	D	tot	A	B	C	D	tot	A	B	C	D	tot	
Male																					
Adolescent	2		1	1	4			1	1	2	1	1		1	3	1	4	2		7	**16**
Adult		1	3		4	1	2	1		4	1	9	1	5	16	2	1	5	2	10	**34**
Mature adult	2	1		2	5	2	1	3	3	9			1	1	2		1			1	**17**
Senior																			1	1	**1**
No age		1			1																**1**
tot	4	3	4	3	14	3	3	5	4	15	2	10	2	7	21	3	6	7	3	19	**69**
Female																					
Adolescent	2	1	2	3	8		2	2	4	8	3	2	1		6	6	10	4	16	36	**58**
Adult	2	2	1		5	1	1	1		3	1	1	1	1	4	7	4	6	5	22	**34**
Mature adult		2	1	2	5	3				3	1				1		5	3	2	10	**19**
Senior																					
No age						1				1											**1**
tot	4	5	4	5	18	5	3	3	4	15	5	3	2	1	11	13	19	13	23	68	**112**
Total	**8**	**8**	**8**	**8**	**32**	**8**	**6**	**8**	**8**	**30**	**7**	**13**	**4**	**8**	**32**	**16**	**25**	**20**	**26**	**87**	**181**

Note. Versions A, B, C and D are the versions of the actual experiment, thus A2, B2, C2 and D2. 'No age' means that the age was not filled in by the participant. tot = subtotal.

B. Testwords

Table 9

List of Testwords and their Possible Gender-Specificness

word	translation	aoa	conc	freq	length	g-spec
alles	*everything*	5.33	1.9	1708.42	5	
dansen	*to dance*	5.00	4.3	102.91	6	V
geweer	*rifle*	6.34	4.9	55.39	6	M
haast	*haste*	7.07	2.9	72.03	5	
hard	*hard*	5.55	3.4	159.46	4	
idee	*idea*	6.71	1.3	482.99	4	
jaar	*year*	4.45	2.1	762.67	4	
jullie	*you* (plural)	4.75	2.0	2325.71	6	
jurk	*dress*	6.14	4.5	55.75	4	V
kan	*can* (1st, 2nd and 3rd person, singular) or *jar*	5.22	4.2	4025.85	3	
kijk	*look*	4.14	3	1049.77	4	
klauw	*claw*	7.12	4.7	3.27	5	M
klimmen	*to climb*	5.89	3.9	13.17	7	
kom	*come* (1st person, singular) or *bowl*	5.28	4.6	2612.54	3	
leger	*army*	7.05	4.1	107.98	5	M
liter	*litre*	7.69	3.6	8.62	5	
morgen	*tomorrow or morning*	4.65	1.7	423.61	6	
nachtegaal	*nightingale*	8.33	4.8	1.07	10	
nat	*wet*	4.59	4.5	30.57	3	
nek	*neck*	5.19	4.6	57.72	3	
pijn	*pain*	3.79	3.3	266.16	4	
priester	*priest*	6.21	3.8	31.42	8	
roman	*novel*	10.05	4.1	10.98	5	V
salon	*lounge*	7.16	4.3	5.01	5	V
schepper	*creator*	8.25	2.7	4.18	8	
school	*school*	3.61	4.3	246.95	6	
serie	*series*	8.99	2.7	14.7	5	
sjaal	*shawl*	5.36	4.9	5.31	5	V
snor	*moustache*	5.12	4.8	9.95	4	M
spijt	*regret*	6.17	1.7	665.78	5	
tabak	*tobacco*	8.58	4.7	5.26	5	M
uur	*hour*	5.92	3.7	348.69	3	
verminking	*mutilation*	10.85	3.3	1.17	10	
vervalsing	*forgery*	9.78	2.3	2.38	10	
visum	*visa*	12.63	3.6	3.7	5	
voorzitter	*chairman*	9.43	3.3	19.03	10	
weg	*way or away*	4.90	4.1	1481.66	3	
xylofoon	*xylophone*	10.10	4.6	0.18	8	
yoga	*yoga*	11.45	3.3	3.73	4	
ziek	*ill*	4.71	3.2	129.2	4	

C. Versions

Version A

n	b	k	r	r	p	t	j	j	b	d	l	c	i	s	p	e	s	d	m	b	y	n	l
i	a	b	e	s	e	r	i	e	z	g	s	s	k	i	j	k	l	z	c	o	n	v	e
e	m	h	r	q	j	a	a	r	i	a	m	v	o	o	r	z	i	t	t	e	r	g	i
t	o	q	w	j	e	k	t	h	a	r	d	u	l	m	i	p	o	h	l	r	j	x	e
c	s	c	h	o	o	l	x	y	w	u	a	v	f	k	l	a	u	w	s	i	l	l	c
u	i	x	h	g	e	w	e	e	r	v	s	x	i	o	m	a	s	d	l	w	y	j	r
d	k	d	a	n	s	e	n	w	e	k	t	p	r	i	e	s	t	e	r	w	b	t	b
l	y	e	x	q	q	r	l	l	w	z	r	x	q	u	l	b	g	x	d	m	p	p	k
m	u	j	j	t	a	b	a	k	e	z	g	u	v	e	r	m	i	n	k	i	n	g	q
q	x	u	h	k	l	u	h	s	c	h	e	p	p	e	r	p	n	g	q	b	n	q	i
s	e	d	z	i	e	k	d	n	j	b	j	g	x	s	p	i	j	t	l	n	w	r	i
e	k	l	i	m	m	e	n	r	r	e	u	u	r	w	n	f	n	c	q	g	t	p	s
w	d	y	u	r	n	l	k	a	n	k	t	n	a	c	h	t	e	g	a	a	l	y	m
a	f	m	z	k	o	m	o	o	f	q	a	l	l	e	s	r	k	b	p	l	v	i	c
p	y	o	g	a	j	z	i	l	q	i	c	w	w	e	g	p	d	v	j	u	r	k	l
b	k	l	m	l	u	o	h	a	a	s	t	h	u	q	u	l	c	p	l	s	n	o	r
n	c	j	u	l	l	i	e	l	z	k	n	c	m	l	i	t	e	r	j	b	l	g	y
x	x	y	l	o	f	o	o	n	k	i	v	s	a	l	o	n	r	g	b	y	p	q	n
s	m	n	e	k	h	s	r	q	f	v	o	r	s	j	a	a	l	z	f	l	w	u	z
d	c	e	b	d	z	m	c	l	e	g	e	r	x	e	k	d	p	j	p	i	j	n	p
v	m	o	r	g	e	n	c	v	k	o	y	v	e	r	v	a	l	s	i	n	g	e	d
s	a	n	f	f	x	h	l	p	f	p	i	d	e	e	c	c	n	f	y	y	y	x	a
y	v	m	v	i	s	u	m	t	j	y	l	h	r	r	o	m	a	n	u	a	j	x	o
b	x	e	t	o	g	u	p	p	u	m	n	a	t	v	u	w	l	h	g	r	y	r	v

Version B

b	x	e	t	o	g	u	p	p	u	m	n	a	t	v	u	w	l	h	g	r	y	r	v
s	m	n	e	k	h	s	r	q	f	v	o	r	s	j	a	a	l	z	f	l	w	u	z
d	c	e	b	d	z	m	c	l	e	g	e	r	x	e	k	d	p	j	p	i	j	n	p
v	m	o	r	g	e	n	c	v	k	o	y	v	e	r	v	a	l	s	i	n	g	e	d
s	a	n	f	f	x	h	l	p	f	p	i	d	e	e	c	c	n	f	y	y	y	x	a
y	v	m	v	i	s	u	m	t	j	y	l	h	r	r	o	m	a	n	u	a	j	x	o
u	i	x	h	g	e	w	e	e	r	v	s	x	i	o	m	a	s	d	l	w	y	j	r
n	b	k	r	r	p	t	j	j	b	d	l	c	i	s	p	e	s	d	m	b	y	n	l
i	a	b	e	s	e	r	i	e	z	g	s	s	k	i	j	k	l	z	c	o	n	v	e
e	m	h	r	q	j	a	a	r	i	a	m	v	o	o	r	z	i	t	t	e	r	g	i
t	o	q	w	j	e	k	t	h	a	r	d	u	l	m	i	p	o	h	l	r	j	x	e
c	s	c	h	o	o	l	x	y	w	u	a	v	f	k	l	a	u	w	s	i	l	l	c
e	k	l	i	m	m	e	n	r	r	e	u	u	r	w	n	f	n	c	q	g	t	p	s
d	k	d	a	n	s	e	n	w	e	k	t	p	r	i	e	s	t	e	r	w	b	t	b
l	y	e	x	q	q	r	l	l	w	z	r	x	q	u	l	b	g	x	d	m	p	p	k
m	u	j	j	t	a	b	a	k	e	z	g	u	v	e	r	m	i	n	k	i	n	g	q
q	x	u	h	k	l	u	h	s	c	h	e	p	p	e	r	p	n	g	q	b	n	q	i
s	e	d	z	i	e	k	d	n	j	b	j	g	x	s	p	i	j	t	l	n	w	r	i
x	x	y	l	o	f	o	o	n	k	i	v	s	a	l	o	n	r	g	b	y	p	q	n
w	d	y	u	r	n	l	k	a	n	k	t	n	a	c	h	t	e	g	a	a	l	y	m
a	f	m	z	k	o	m	o	o	f	q	a	l	l	e	s	r	k	b	p	l	v	i	c
p	y	o	g	a	j	z	i	l	q	i	c	w	w	e	g	p	d	v	j	u	r	k	l
b	k	l	m	l	u	o	h	a	a	s	t	h	u	q	u	l	c	p	l	s	n	o	r
n	c	j	u	l	l	i	e	l	z	k	n	c	m	l	i	t	e	r	j	b	l	g	y

Version C

n	c	j	u	l	l	i	e	l	z	k	n	c	m	l	i	t	e	r	j	b	l	g	y
x	x	y	l	o	f	o	o	n	k	i	v	s	a	l	o	n	r	g	b	y	p	q	n
w	d	y	u	r	n	l	k	a	n	k	t	n	a	c	h	t	e	g	a	a	l	y	m
a	f	m	z	k	o	m	o	o	f	q	a	l	l	e	s	r	k	b	p	l	v	i	c
p	y	o	g	a	j	z	i	l	q	i	c	w	w	e	g	p	d	v	j	u	r	k	l
b	k	l	m	l	u	o	h	a	a	s	t	h	u	q	u	l	c	p	l	s	n	o	r
y	v	m	v	i	s	u	m	t	j	y	l	h	r	r	o	m	a	n	u	a	j	x	o
b	x	e	t	o	g	u	p	p	u	m	n	a	t	v	u	w	l	h	g	r	y	r	v
s	m	n	e	k	h	s	r	q	f	v	o	r	s	j	a	a	l	z	f	l	w	u	z
d	c	e	b	d	z	m	c	l	e	g	e	r	x	e	k	d	p	j	p	i	j	n	p
v	m	o	r	g	e	n	c	v	k	o	y	v	e	r	v	a	l	s	i	n	g	e	d
s	a	n	f	f	x	h	l	p	f	p	i	d	e	e	c	c	n	f	y	y	y	x	a
c	s	c	h	o	o	l	x	y	w	u	a	v	f	k	l	a	u	w	s	i	l	l	c
u	i	x	h	g	e	w	e	e	r	v	s	x	i	o	m	a	s	d	l	w	y	j	r
n	b	k	r	r	p	t	j	j	b	d	l	c	i	s	p	e	s	d	m	b	y	n	l
i	a	b	e	s	e	r	i	e	z	g	s	s	k	i	j	k	l	z	c	o	n	v	e
e	m	h	r	q	j	a	a	r	i	a	m	v	o	o	r	z	i	t	t	e	r	g	i
t	o	q	w	j	e	k	t	h	a	r	d	u	l	m	i	p	o	h	l	r	j	x	e
s	e	d	z	i	e	k	d	n	j	b	j	g	x	s	p	i	j	t	l	n	w	r	i
e	k	l	i	m	m	e	n	r	r	e	u	u	r	w	n	f	n	c	q	g	t	p	s
d	k	d	a	n	s	e	n	w	e	k	t	p	r	i	e	s	t	e	r	w	b	t	b
l	y	e	x	q	q	r	l	l	w	z	r	x	q	u	l	b	g	x	d	m	p	p	k
m	u	j	j	t	a	b	a	k	e	z	g	u	v	e	r	m	i	n	k	i	n	g	q
q	x	u	h	k	l	u	h	s	c	h	e	p	p	e	r	p	n	g	q	b	n	q	i

Version D

q	x	u	h	k	l	u	h	s	c	h	e	p	p	e	r	p	n	g	q	b	n	q	i
s	e	d	z	i	e	k	d	n	j	b	j	g	x	s	p	i	j	t	l	n	w	r	i
e	k	l	i	m	m	e	n	r	r	e	u	u	r	w	n	f	n	c	q	g	t	p	s
d	k	d	a	n	s	e	n	w	e	k	t	p	r	i	e	s	t	e	r	w	b	t	b
l	y	e	x	q	q	r	l	l	w	z	r	x	q	u	l	b	g	x	d	m	p	p	k
m	u	j	j	t	a	b	a	k	e	z	g	u	v	e	r	m	i	n	k	i	n	g	q
b	k	l	m	l	u	o	h	a	a	s	t	h	u	q	u	l	c	p	l	s	n	o	r
n	c	j	u	l	l	i	e	l	z	k	n	c	m	l	i	t	e	r	j	b	l	g	y
x	x	y	l	o	f	o	o	n	k	i	v	s	a	l	o	n	r	g	b	y	p	q	n
w	d	y	u	r	n	l	k	a	n	k	t	n	a	c	h	t	e	g	a	a	l	y	m
a	f	m	z	k	o	m	o	o	f	q	a	l	l	e	s	r	k	b	p	l	v	i	c
p	y	o	g	a	j	z	i	l	q	i	c	w	w	e	g	p	d	v	j	u	r	k	l
s	a	n	f	f	x	h	l	p	f	p	i	d	e	e	c	c	n	f	y	y	y	x	a
y	v	m	v	i	s	u	m	t	j	y	l	h	r	r	o	m	a	n	u	a	j	x	o
b	x	e	t	o	g	u	p	p	u	m	n	a	t	v	u	w	l	h	g	r	y	r	v
s	m	n	e	k	h	s	r	q	f	v	o	r	s	j	a	a	l	z	f	l	w	u	z
d	c	e	b	d	z	m	c	l	e	g	e	r	x	e	k	d	p	j	p	i	j	n	p
v	m	o	r	g	e	n	c	v	k	o	y	v	e	r	v	a	l	s	i	n	g	e	d
t	o	q	w	j	e	k	t	h	a	r	d	u	l	m	i	p	o	h	l	r	j	x	e
c	s	c	h	o	o	l	x	y	w	u	a	v	f	k	l	a	u	w	s	i	l	l	c
u	i	x	h	g	e	w	e	e	r	v	s	x	i	o	m	a	s	d	l	w	y	j	r
n	b	k	r	r	p	t	j	j	b	d	l	c	i	s	p	e	s	d	m	b	y	n	l
i	a	b	e	s	e	r	i	e	z	g	s	s	k	i	j	k	l	z	c	o	n	v	e
e	m	h	r	q	j	a	a	r	i	a	m	v	o	o	r	z	i	t	t	e	r	g	i

D. Instructions

Paper

Inleiding
Dit is een experiment voor een scriptieonderzoek voor Taalwetenschap. Door uw deelname helpt u ook mee aan de verbetering van onderzoeksmethoden in dat vakgebied.

Uitleg
Het experiment bestaat uit een soort woordzoeker. Bij de woordzoeker zijn echter geen woorden gegeven die u moet vinden, maar het doel is om zoveel mogelijk woorden te vinden binnen drie minuten. De woorden bevinden zich alleen horizontaal, van links naar rechts. Markeer de woorden die u vindt door het woord te omcirkelen.
Eerst vragen we u echter onderstaande vier vragen in te vullen, die ons de benodigde achtergrondgegevens verschaffen (geslacht, leeftijd, moedertaal en de aanwezigheid van de onderzoeker).
Daarna kunt u de pagina omdraaien en beginnen met het invullen van de woordzoeker. Hiervoor heeft u drie minuten de tijd.
Nadat de drie minuten voor het invullen van de woordzoeker verstreken zijn, kunt u eronder eventuele opmerkingen aangeven en uw e-mailadres noteren als u graag de resultaten van het onderzoek zou willen ontvangen.
Alvast hartelijk dank voor uw deelname!

Vragen
Om het onderzoek te kunnen doen, hebben we enkele persoonsgegevens nodig.

Geslacht:
- O Man
- O Vrouw

Leeftijd:

Moedertaal:
- O Nederlands
- O Anders:

Onderzoeker:
- O Aanwezig

Niet aanwezig

On the other side of the page, below the word search puzzle, the following could be found:

Opmerkingen:

E-mailadres:

Digital

Onderzoek

Inleiding

Dit is een experiment voor een scriptieonderzoek voor Taalwetenschap. Door uw deelname helpt u ook mee aan de verbetering van onderzoeksmethoden in dat vakgebied.

Uitleg

Dit experiment kan uitgevoerd worden op een computer of een tablet.

Het experiment bestaat uit een soort woordzoeker. Bij de woordzoeker zijn echter geen woorden gegeven die u moet vinden, maar het doel is om zoveel mogelijk woorden te vinden binnen drie minuten. De woorden bevinden zich alleen horizontaal, van links naar rechts.

Eerst vragen we u vier vragen in te vullen, die ons de benodigde achtergrondgegevens verschaffen (geslacht, leeftijd, moedertaal en de aanwezigheid van de onderzoeker). Als uw moedertaal niet het Nederlands is, kunt u het vinkje weghalen door erop te klikken en dan verschijnt er een veld waar u uw moedertaal in kunt typen.

Daarna krijgt u een pagina te zien met daarin een oefenwoordzoeker, zodat u kunt oefenen met de manier van het markeren van de woorden. Wanneer u een woord heeft geselecteerd, verschijnt er linksboven een gele berichtje 'opslaan...', die verandert in 'opgeslagen', zodat u kunt zien dat uw woorden opgeslagen zijn. Wanneer u genoeg heeft geoefend en ten minste drie woorden heeft geselecteerd in de oefenwoordzoeker, kunt u doorgaan naar de daadwerkelijke woordzoeker.

Nadat de drie minuten voor het invullen van de woordzoeker verstreken zijn, verschijnt een pop-up met een knop om naar de afsluitende pagina te gaan. Daar heeft u de mogelijkheid om opmerkingen aan te geven en eventueel uw e-mailadres in te vullen als u graag de resultaten van het onderzoek zou willen ontvangen.

Alvast hartelijk dank voor uw deelname!

Ja, ik doe mee

Gegevens

Om het onderzoek te kunnen doen, hebben we enkele persoonsgegevens nodig.

De gegevens zijn niet te herleiden tot u als persoon en worden alleen gebruikt om de resultaten te kunnen classificeren.

Geslacht:

Vrouw

Leeftijd:

21

☑ **Nederlandstalig**

☐ **In aanwezigheid van onderzoeker**

Opslaan en verder

Oefen woordzoeker

s	k	s	u	f	a	r	e	e	d
r	k	l	a	p	l	o	p	e	r
q	e	y	e	g	a	s	r	o	q
s	s	a	c	h	z	t	c	i	z
v	m	e	e	t	n	u	a	u	w
s	c	h	r	o	e	v	e	n	d
k	r	o	k	o	d	i	l	e	r
g	e	o	d	r	i	e	m	e	t
f	l	u	i	d	s	p	r	a	k
k	r	a	a	k	e	r	w	o	n

Uitleg

Selecteer woorden in de rechthoek. Bijvoorbeeld het woord klap in de tweede rij.

Sleep over het woord alsof je het doorstreept.

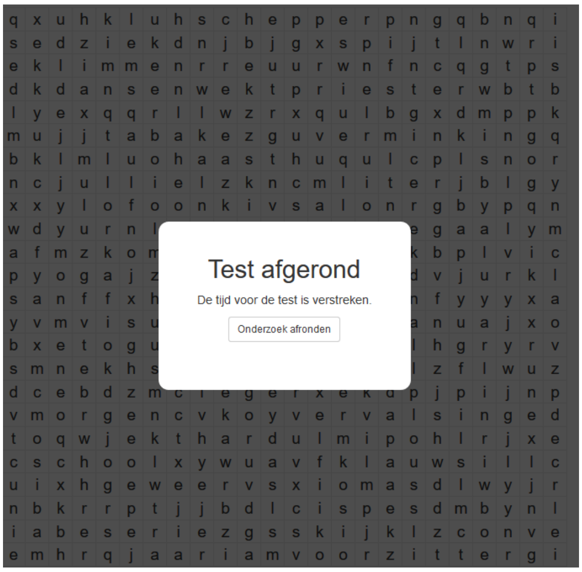
Je kunt ook klikken op de eerste en daarna op de laatste letter van een woord.

Heb je de verkeerde beginletter geselecteerd, klik dan buiten de betreffende rij om de selectie ongedaan te maken. Of druk op Escape.

Geselecteerde woorden worden gemarkeerd en opgeslagen.

Je moet minimaal drie woorden markeren om verder te kunnen.

Ja, ik doe mee



Hartelijk dank

Hartelijk dank voor uw deelname aan het onderzoek.

Als u nog opmerkingen heeft, kunt u die hieronder aan ons kwijt.

Wanneer u de resultaten van het onderzoek wilt ontvangen, vul dan hieronder uw emailadres in. Dit emailadres wordt alleen gebruikt om het onderzoek aan u toe te kunnen sturen. Daarna wordt het emailadres uit onze database verwijderd. Het emailadres is niet te relateren aan een testresultaat.

Opmerkingen:

Emailadres:

Opslaan en afsluiten

Afsluiting

Het onderzoek is afgerond. U kunt nu deze pagina sluiten.

E. Transforming data

These notebooks are used to put the data into usable Excel files. Most of the programming has been done by Hans Kalle, but other parts (mostly adjustments) have been programmed by Mattanja Kalle. Clarifications on what transformations have taken place are part of the notebooks themselves.

Word Search Puzzle - Data Processing and Transformation

June 9, 2016

1 Data Processing Word Search Puzzle

This workboek processes the data gathered with the wordsearch tests.

1.1 Environment

```
In [ ]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import doctest

        pd.set_option('display.max_columns', 80)

        %matplotlib inline
```

1.2 Download the data

```
In [ ]: sessions_digital = pd.read_csv("http://[redacted]/research/services/wordsearch/downloa
        sessions_digital['digital'] = True
        print(sessions_digital.shape)
```

```
In [ ]: sessions_paper = pd.read_csv("http://[redacted]/researchm/services/wordsearch/download/
        sessions_paper['digital'] = False
        print(sessions_paper.shape)
```

```
In [ ]: sessions = pd.concat([sessions_digital, sessions_paper], ignore_index = True)
        sessions.age = pd.to_numeric(sessions.age, errors='coerce')
        sessions.dutch = sessions.dutch == 1
        sessions.mothertongue = sessions.mothertongue.fillna('')
        sessions.present = sessions.present == 1
        sessions.remark = sessions.remark.fillna('')
        sessions.completed = sessions.completed == 1
        sessions.insubset = sessions.insubset == 1
        sessions.head(5)
```

```
In [ ]: words_digital = pd.read_csv("http://[redacted]/research/services/wordsearch/download/wo
        print(words_digital.shape)
        words_paper = pd.read_csv("http://[redacted]/researchm/services/wordsearch/download/wor
        print(words_paper.shape)
        words = pd.concat([words_digital, words_paper], ignore_index = True)
        words.t1 = words.t1.str.replace(',','.')
        words.t1 = pd.to_numeric(words.t1, errors='coerce')
        words.t2 = words.t2.str.replace(',','.')
        words.t2 = pd.to_numeric(words.t2, errors='coerce')
```

```
words.describe()
words.head(5)
```

```
In [ ]: testwords = pd.read_csv("data/testwords_loc.csv")
        testwords.gender = testwords.gender.fillna('')
        print(testwords.shape)
        testwords.head(3)
```

```
In [ ]: boxes = pd.read_csv("data/boxes_experiment.csv", sep=";")
        boxes
```

1.3 Filter correct sessions

Only sessions in testperiod (1-1-2016 .. 19-1-2016).

```
In [ ]: remove = sessions.start.isnull()
        print('Number sessions that never started the test: %d.' % sum(remove))
        sessions = sessions[~remove]
        remove = sessions.digital & (sessions.start < "2016-01-01")
        print('Number of digital sessions before 1-1-2016: %d.' % sum(remove))
        sessions = sessions[~remove]
        remove = sessions.digital & (sessions.start >= "2016-01-20")
        print('Number of digital sessions after 19-1-2016: %d.' % sum(remove))
        sessions= sessions[~remove]
        print(sessions.shape)
```

Sessions used in testing were marked with a negative age or not known and must be deleted.

```
In [ ]: selected = sessions.age > 0
        print('Sessions used in testing (age < 1): %d.' % sum(~selected))
        removed = sessions[~selected]
        sessions = sessions[selected]
        sessions.loc[sessions.age == 100,'age'] = np.NaN
        print(sessions.shape)
        removed.head(10)
```

Complete digital sessions with words that end at more than 170 seconds.

```
In [ ]: suspect_completion = words.id[words.t2 > 170].unique()
        selected = sessions.id.isin(suspect_completion) & ~sessions.completed
        print('Not completed sessions with words selected later than 170 seconds from start: %d.'
              % sum(selected))
        sessions.loc[selected,'completed'] = True
        sessions[selected]
```

Remove not completed sessions.

```
In [ ]: remove = ~sessions.completed
        removed = sessions[remove]
        print('Not completed sessions: %d.' % sum(remove))
        sessions = sessions[~remove]
        removed
```

Remove words not used in usable sessions.

```
In [ ]: selected = words.id.isin(sessions.id)
        print("Number of words not in selected sessions: %d." % sum(~selected))
        words = words[selected]
        print(words.shape)
```

1.4 Conversions to improve usability

Some words were selected backwards. The left-to-right (ltr) version needs to be added to the dataset.

```
In [ ]: %doctest_mode

        def ltr_word(c1, c2, word):
            if c1 > c2:
                return word[::-1]
            return word

        """
        Testing ltr_word
        >>> ltr_word(1,4,'joop')
        'joop'
        >>> ltr_word(4,1,'kjik')
        'kijk'
        >>> ltr_word(1,1,'k')
        'k'
        >>> ltr_word(14,1,'achterstevoren')
        'nerovetsrethca'
        >>> ltr_word(10,1,'voorzitter')
        'rettizroov'
        """

        doctest.testmod()

        %doctest_mode

        words['ltr_word'] = words.apply(lambda row: ltr_word(row.c1, row.c2, row.word), axis=1)
        words['backwards'] = words.apply(lambda row: row.c1 > row.c2, axis=1)
        words[words.backwards].head(5)
```

Mark which words are part of the testswords.

```
In [ ]: words['is_testword'] = words.ltr_word.isin(testwords.word)
        words.tail(10)
```

Add box version to sessions.

```
In [ ]: boxnames = dict(zip(boxes.id, boxes.description))
        sessions['boxname'] = sessions.apply(lambda row: boxnames[row.box][-2:], axis=1)
        sessions.head(10)
```

Define condition (PP, PA, DP, DA) per version.

```
In [ ]: %doctest_mode

        def condition(digital, present):
            if digital:
                if present:
                    return "DP"
                else:
                    return "DA"
            else:
                if present:
```

```
            return "PP"
        else:
            return "PA"

"""
Testing condition
>>> condition(True,True)
'DP'
>>> condition(True,False)
'DA'
>>> condition(False,True)
'PP'
>>> condition(False,False)
'PA'
"""

doctest.testmod()

%doctest_mode

sessions['condition'] = sessions.apply(lambda row: condition(row.digital, row.present), axis=1)
sessions.head(10)
```

Add age category.

```
In [ ]: sessions.age.hist()
```

```
In [ ]: %doctest_mode

        def generation(age):
            if 0 < age < 25:
                return "adolescent"
            elif 25 <= age < 40:
                return "adult"
            elif 40<= age < 65:
                return "mature adult"
            elif 65 <= age <=125:
                return "senior"

        """
        Testing generation
        >>> generation(-5)
        >>> generation(25)
        'adult'
        >>> generation(40)
        'mature adult'
        >>> generation(46)
        'mature adult'
        >>> generation(70)
        'senior'
        """

        doctest.testmod()

        %doctest_mode
```

```
sessions['generation'] = sessions.apply(lambda row: generation(row.age), axis=1)
sessions.head(10)
```

Add wordcounts to sessions.

```
In [ ]: unique_words = words.groupby('id').ltr_word.unique()
        sessions['wordcount'] = sessions.apply(lambda row: len(unique_words[row.id]), axis=1)
        unique_testwords = words[words.is_testword].groupby('id').ltr_word.unique()
        sessions['testwordcount'] = sessions.apply(lambda row: len(unique_testwords[row.id]), axis=1)
        sessions.head(20)
```

2 Filter out strange sessions

```
In [ ]: sessions.testwordcount.hist()
```

```
In [ ]: remove = sessions.testwordcount < 5
        print("Number of session with less then 5 testwords found: %d." % sum(remove))
        removed = sessions[remove]
        sessions = sessions[~remove]
        removed
```

Again filter out words not in usable sessions.

```
In [ ]: selected = words.id.isin(sessions.id)
        print("Number of words not in selected sessions: %d." % sum(~selected))
        removed = words[~selected]
        words = words[selected]
        print(words.shape)
        removed.head(10)
```

```
In [ ]: sessions[sessions.digital].testwordcount.hist()
```

2.1 Combining sessions and words

Combine sessions with testwords.

```
In [ ]: sessions['dummy'] = 1
        testwords['dummy'] = 1
        sessions_testwords = pd.merge(sessions, testwords, on='dummy', suffixes=['_p', '_specific'])
        sessions_testwords.head(10)
```

As it only matters whether the participants have found the word, only the first selections of each word will be included.

```
In [ ]: unique_words = words.groupby(['id','ltr_word']).first().reset_index()
        unique_words.head(20)
```

Combine sessions with words to create combined.

```
In [ ]: combined = pd.merge(sessions, words, on='id')
        print(combined.shape)
        combined.head(3)
```

Combine sessions_testwords with words to create complete.

```
In [ ]: complete = pd.merge(sessions_testwords, words, left_on=['id','word'],
                            right_on=['id','ltr_word'], how='left', suffixes=['','_selected'])
        complete['found'] = ~complete.ltr_word.isnull()
        complete.head(10)
```

Combine sessions_testwords with unique_words to create recognized.

```
In [ ]: recognized = pd.merge(sessions_testwords, unique_words, left_on=['id','word'],
                              right_on=['id','ltr_word'], how='left', suffixes=['','_selected'])
        recognized['found'] = ~recognized.ltr_word.isnull()
        recognized.head(10)
```

Add testword found boolean to sessions.

```
In [ ]: for word in testwords.word.values:
            found_in = words[words.word == word].id.values
            sessions[word] = sessions.apply(lambda row: row.id in found_in, axis=1)
        sessions.head(5)
```

Add location specific to the version (A, B, C or D) of the session to complete.

```
In [ ]: def get_location(row):
            if row.boxname == 'A2':
                return row.location_A
            elif row.boxname == 'B2':
                return row.location_B
            elif row.boxname == 'C2':
                return row.location_C
            elif row.boxname == 'D2':
                return row.location_D
            else:
                return -1

        complete['location'] = complete.apply(lambda row: get_location(row), axis=1)
        complete.head(10)
```

Add location specific to the version (A, B, C or D) of the session to recognized.

```
In [ ]: recognized['location'] = recognized.apply(lambda row: get_location(row), axis=1)
        recognized.head(10)
```

Add percentages found to testwords.

```
In [ ]: def percentage(selection, combined_fieldname, value):
            return len(combined[(combined[combined_fieldname] == value)
                                 & selection].id.unique()) / len(combined[selection].id.unique())

        def add_column(ds, column_name, selection, combined_fieldname, ds_fieldname):
            ds[column_name] = ds.apply(lambda row:
                                 percentage(selection, combined_fieldname, row[ds_fieldname]), axis=1)

        def add_column_testwords_per_word(column_name, selection):
            add_column(testwords, column_name, selection, 'ltr_word', 'word')

        add_column_testwords_per_word('V', combined.gender == 'V')
        add_column_testwords_per_word('M', combined.gender == 'M')
```

```
add_column_testwords_per_word('adolescent', combined.generation == 'adolescent')
add_column_testwords_per_word('adult', combined.generation == 'adult')
add_column_testwords_per_word('mature adult', combined.generation == 'mature adult')
add_column_testwords_per_word('senior', combined.generation == 'senior')
add_column_testwords_per_word('A', combined.boxname == 'A2')
add_column_testwords_per_word('B', combined.boxname == 'B2')
add_column_testwords_per_word('C', combined.boxname == 'C2')
add_column_testwords_per_word('D', combined.boxname == 'D2')
add_column_testwords_per_word('digital', combined.digital)
add_column_testwords_per_word('paper', ~combined.digital)
add_column_testwords_per_word('present', combined.present)
add_column_testwords_per_word('absent', ~combined.present)
add_column_testwords_per_word('DP', combined.digital & combined.present)
add_column_testwords_per_word('DA', combined.digital & ~combined.present)
add_column_testwords_per_word('PP', ~combined.digital & combined.present)
add_column_testwords_per_word('PA', ~combined.digital & ~combined.present)
add_column_testwords_per_word('found', combined.completed)

testwords
```

Write to excel. The recognized tab is the one with all the information, but the other tabs are included in case they are a more practical overview for certain situations.

```
In [ ]: with pd.ExcelWriter('data/results_recognized.xlsx') as writer:
            boxes.to_excel(writer, sheet_name='boxes')
            sessions.to_excel(writer, sheet_name='sessions')
            words.to_excel(writer, sheet_name='words')
            combined.to_excel(writer, sheet_name='combined')
            testwords.to_excel(writer, sheet_name='testwords')
            complete.to_excel(writer, sheet_name='complete')
            recognized.to_excel(writer, sheet_name='recognized')
```