DOES PREDICTABILITY AFFECT GENERATION?

A behavioural study of the generation effect using memory

Rossum, N.J. van (Natasja)
Supervisor: Joost Rommers
Bachelor thesis

Contents

Abstract

Based on the idea that predictions during language comprehension are generated by the language production system, this behavioural study explored the relationship between the prediction and the generation of words. This was done using the generation effect in memory, which is the finding that words that have been generated by participants themselves are better remembered than words that have only been perceived. Participants read sentences ending in a predictable or unpredictable word, which was presented as a picture. They named the picture themselves or only heard its name. Then, they were tested on their memory for the words. The hypothesis was that, if predictable words (but not unpredictable words) are already covertly generated by the production system, then the memory results would show an interaction effect between predictability and generation: the added memory improvement from actual generation should be smaller for predictable words than for unpredictable words. The results showed a clear generation effect, but could not provide evidence to support the interaction effect. As this might be due to a ceiling effect, improvements to the experimental design are suggested.

1. Introduction

So far, language comprehension and language production have mostly been studied separately. The current study, however, will focus on the relationship between comprehension and production. In particular, this behavioural study of the generation effect on memory will examine the link between language generation and language prediction.

The aim of the study is to explore the relationship between prediction and generation with the use of a memory task. The introduction begins by showing evidence that has been found on predictive language processing, followed by the suggestion that the production system is used while predicting words. It will then give a brief overview of the generation effect and the production effect, taking previous research on these topics into account. The introduction ends by giving a motivation and hypotheses for the present experiment. The second section describes the methodology used. The third section presents the findings of the experiment and is followed by an analysis of these results.

Strong evidence has been found on prediction playing a crucial role in language processing. Research has found that people take longer to read garden-path sentences, due to the extra time required for mentally correcting these sentences while reading (Ferreira & Clifton, 1986). A study by Ferreira, Christianson and Hollingsworth (2001) shows that people make predictions when they read garden-path sentences. An example of such a sentence is “While Anna dressed the baby spit up on the bed”. As people read sentences word by word, seeing “Anna dressed the baby” at first seems a plausible combination. However, when finishing the sentence, it becomes clear that Anna is dressing herself and not the baby. Ambiguous garden-path sentences can yield confusion, as the sentences are relatively difficult to read and are often misanalysed. The experiment by Ferreira, Christianson and Hollingsworth (2001) showed that people often incorrectly believe that Anna dressed the baby, but also believe the baby spat up on the bed. It is also possible that participants predict the framework of the sentence instead of the next few words. This is known as the Minimal Attachment preference (Ferreira & Clifton, 1986), in which participants choose the easiest structure which could be applied to the end of the sentence and therefore predict the structure of the sentence. Most importantly, these studies show that prediction plays an important role in language processing.

Other evidence for prediction comes from brain activity. These studies about event-related brain potentials (ERPs) also suggest that people predict while reading sentences. The N400, an event-related brain potential, is elicited by different linguistic and non-linguistic manipulations (for a review, see Kutas & Federmeier (2011)). A study including brain activity measurements concluded that words were pre-activated during language comprehension (DeLong, Urbach & Kutas, 2005). For example, when a sentence such as "The day was breezy so the boy went outside to fly a …" is presented, the predicted word at the end would be "kite". When an unexpected but plausible noun was presented at the end of this sentence, such as "airplane", a stronger N400 was elicited than for the word "kite". The word "kite" at the end of the sentence has a cloze probability of 89%, which means that 89% of people asked to complete the sentence responded with the word "kite". A cloze probability of 86% was measured for the indefinite article "a". When presenting the incorrect indefinite article, a mismatch occurred and a stronger N400 than for a matching article was observed. Both articles could have been a correct continuation of the sentence, but participants seem to favour "a" in the majority of the cases. This shows that we predict specific words when reading sentences.

Moreover, this N400 was also elicited by mismatches in gender in Dutch (van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005) and in Spanish (Wicha, Bates, Moreno & Kutas, 2003; Wicha, Morena & Kutas, 2004). These studies manipulated the gender of articles and other determiners such that they did not agree with the following noun. Besides finding evidence in the ERPs, evidence was also provided from the reading times. When the gender mismatches occurred, the determiners had been read at a slower pace than the sentences without a mismatch. These studies contribute to evidence for prediction when reading sentences.

1.1 Mechanisms underlying prediction

After establishing that language processing is predictive, studies also suggest that the production system is used when predicting words during comprehension. The idea that predictions are generated by the language production system comes, among others, from a framework connecting production, comprehension and acquisition: the P-chain (Dell & Chang, 2014). Such connectionist networks can help to gain a better understanding of language processing. This framework suggests that incremental prediction is involved in language processing and that this is accomplished by using the production system. With the linguistic input that we perceive, we can put ourselves in the speaker's shoes. Not only do we consider what the last word of a sentence could be, but we also covertly produce what we would say if we were the speaker. Similarly, Pickering and Garrod (2007) suggest that the production system creates a model during language processing to anticipate upcoming linguistic elements to support comprehension, thus also claiming that the production system is being used when making predictions.

Multiple studies show that prediction-related effects in ERPs and eye movements are stronger in participants who perform well in verbal fluency tasks (Federmeier, Kutas & Schul, 2010; Hintz, Meyer & Huettig, 2017; Rommers, Meyer & Huettig, 2015). Federmeier, Kutas and Schul (2010) measured the brain activity of younger and older adults when they were reading short phrases. In the verbal fluency task, participants had to name as many words from a particular category as they could within a time limit. This requires access to the mental lexicon and picking the correct words. The predictive sentences were read faster than the rest of the sentences, which means that prediction facilitates language production. The authors concluded that production mechanisms are linked to predictions made during language comprehension. Similar to Federmeier et al. (2010), Rommers et al. (2015) came to the conclusion that multiple mechanisms, such as production, are involved in predictive language processing. In their study, eye movements were observed and strong effects were found linking eye movements and production. A second study looking into eye movement and prediction in verbal fluency tasks concluded that multiple mechanisms are involved in predictive language processing (see also Hintz, Meyer & Huettig, 2017). Participants with strong anticipatory eye movements performed well in these verbal fluency tasks, confirming that prediction and production performance are related.

Moreover, research has shown that both prediction and language processing are left-hemisphere dominant. Several studies examined lateralized differences using ERPs. Federmeier and Kutas (1999) concluded that both hemispheres are sensitive to context, but only the left hemisphere is sensitive to expectations. Wlotko and Federmeier (2007) found that expected endings rather than unexpected endings of sentences activated the left hemisphere. By investigating these findings, it could be surmised that prediction has an influence on other cognitive processes, such as memory. The fact that prediction and production are both dominant in the left hemisphere suggests that these mechanisms are linked and collaborate in language comprehension.

Furthermore, a study by Martin, Branzi and Bar (2018) suggests that the production system is necessary for prediction during language comprehension. This study found that by taxing the production system, prediction was also hindered. To test this, the articulatory system was supressed by participants saying the same utterance repeatedly, in this way reducing the production system's involvement.

The evidence above begins to support the idea that we do not only predict what the following word would be, but we also generate these predictions. Consequently, the production system is used when we predict words. The main question that we will be asking during our experiment is whether prediction is similar to generation. We will be using an experimental approach using memory as a window to these questions. Before detailing the hypothesis, we will examine previous research on the generation effect.

1.2 The generation effect

The generation effect is the finding that words are remembered better if they are self-generated than if they are merely perceived. Slamecka and Graf (1978) have thoroughly experimented with various task rules, for example having participants generate opposites or synonyms. In the generate condition, when being presented with the word "rapid" and given the first letter of the next word (in this case an "f"), participants self-generated the word "fast". In the read condition, "rapid-fast" was shown. Shortly after, a recognition test followed. For each rule, not only were participants more confident that they had encountered the words before, the recognition probability was also higher for generated items than for read items. This experiment did not find a generation effect in the recognition test with the rule that words rhyme. By contrast, for the other rules (associate, category, opposite and synonym), the results clearly showed a generation effect that held in different circumstances.

The participants in the present study will be shown images as a standard production task. Although showing images is not the usual production task used to elicit a generation effect, multiple studies have confirmed that pictures are remembered better than words (Weldon & Roediger, 1987; Zormpa, Brehm, Hoedemaker & Meyer, 2019). Moreover, Zormpa, Brehm, Hoedemaker and Meyer's study (2019) showed a relationship between the generation effect and the picture superiority effect. This study confirms that naming the pictures out loud improves memory. As this interaction effect has been found in a picture-naming task, we also hope to find a generation effect with pictures used at the end of a sentence, instead of solely presenting words or pictures.

We will first be looking at predictability in the present study, as we have discussed evidence that language processing is predictive. We will present predictable and unpredictable sentences with a missing last word to participants. We will ask participants to read sentences that are displayed on a screen. Sentences have been manipulated into two categories: predictability and generation. The sentences either end with a predictable word (i.e. "he swept the floor with a …") or an unpredictable word (i.e. "in the department store she asked for a …"). The last word of these sentences will be presented as an image. Secondly, we will add generation to the experiment. As seen in Slamecka & Graf's experiment (1978), memory benefits have been found for words that are read out loud in comparison to words that are read silently. Depending on the colour of the sentence, participants were asked to either name the picture out loud or to view the picture silently. After a short distraction task, a memory test followed. By using memory as a window into this question, we gained insight into participants' memories and how memory is influenced by production and predictability.

Rommers, Dell and Benjamin (2014) used the same design as the present study, but asked participants to read written words at the end of a sentence out loud or silently. As mentioned earlier, the current study uses pictures instead of words. Not only are words remembered better when self-generated (Slamecka & Graf, 1978), several studies have also found memory benefits for words that are read out loud in comparison to words that are read silently. This finding is known as the production effect. MacLeod, Gopie, Hourihan, Neary and Ozubko (2010) proved that better memory could be found in an experiment using three conditions (reading words out loud, silently and saying non-words out loud, such as "plarc" or "worp," which are words that have no meaning, but could have a meaning in a different language.) Participants indicated whether the words were old or new to them. The study suggested that saying a word out loud may leave "production traces" behind. Later studies, such as Fawcett and Ozubko (2016), argued that production enhances memory not only through recollection and motor movements, but also through becoming familiar with words seen before. Furthermore, by combining the production effect and the generation effect, Zormpa, Brehm, Hoedemaker and Meyer (2019) observed an interaction effect between the production effect and the generation effect. This resulted in both a generation and a production effect. We expect to find both a generation and a production effect in the current study using pictures.

Figure 1 presents a flow-chart of Levelt, Roelofs and Meyer's (1999) well-established theory of the production of words. When reading a word, the participant would start by going through all the words they know in their mental lexicon and would choose the specific word. This is known as lexical selection. The participant would proceed by going through morphological encoding and phonological encoding to retrieve all the information that they possessed on that word. When seeing a picture, however, the word is activated differently as different stages are passed through.

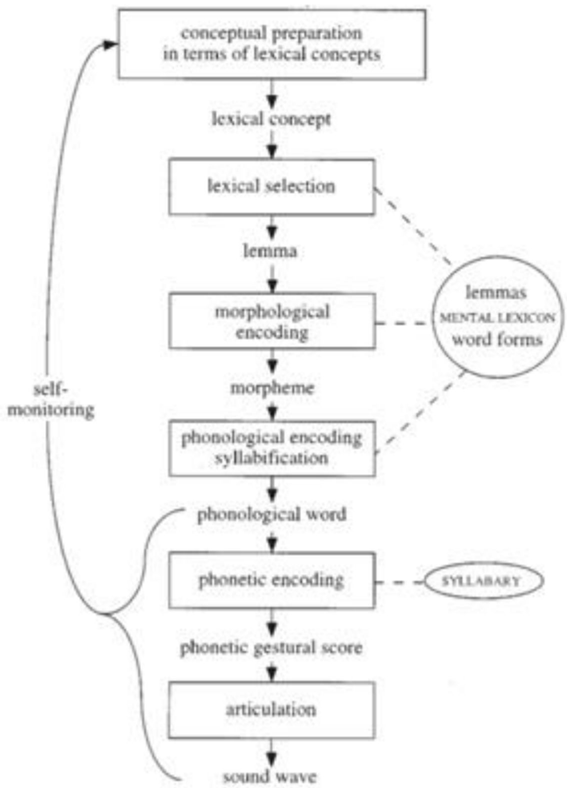

Figure 1: A flow chart of word production based on Levelt, Roelofs and Meyer (1999)

Unlike reading words, where the to-be-produced word is already right in front of the speaker and merely needs to be transformed from letters to sounds, naming pictures starts at conceptualization and encompasses all stages of language production. More processes are therefore involved in processing pictures than reading words. The Dual Route Cascaded Model (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001) demonstrates a theory of a two-way route of reading words out loud: a route from the phonological lexicon to the orthographic lexicon and a grapheme-to-phoneme route (Figure 2). The semantic system is not required for this latter route in order to access the phoneme system and is therefore the easiest accessible route. However, the semantic system is required for naming pictures. This model confirms that less effort is required for reading words out loud rather than for naming pictures. It is clear that more production processes are involved in processing pictures than in reading words.

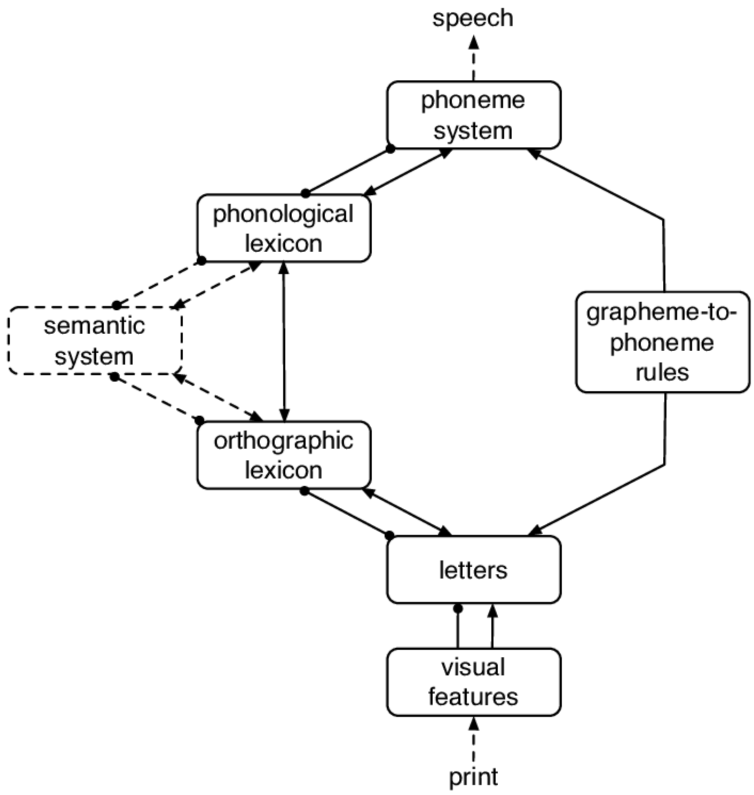

Figure 2: A dual route cascaded model of reading aloud by Coltheart, Rastle, Perry, Langdon and Ziegler (2001)

To summarize, the production effect is calculated by subtracting the memory performance of reading words silently from the memory performance of reading words out loud. The effects found in the study by Rommers et al. (2015) were small, which might be due to participants not using their entire production system when reading words out loud. In the present study we will show pictures in order to engage the full production system from conceptualization to articulation.

We will investigate whether production is like generation by using predictability, images and memory. Two studies come close to this design, but both addressed slightly different research questions. The first study by Rommers et al. (2015) has been discussed earlier. The second study, by Ghatala (1983), analysed the generation effect with predictability, using highly and weakly constraining sentences. She presented participants with sentences with the last word missing and asked the participants to complete these sentences. For highly constraining sentences, the last word was expected (i.e. "The pioneers traveled in covered …"). For weakly constraining sentences, the last word was plausible but unexpected (i.e. "The pioneers discovered that they were without …"). Ghatala (1983) found a stronger generation effect for weakly constraining sentences than for highly constraining sentences, which is consistent with our expectations and hypotheses. However, as Ghatala herself acknowledged, it is possible that weakly constraining sentences allowed for more idiosyncratic responses that were therefore easier to remember. In the present study, we presented pictures in order to control participants spoken responses even in weakly constraining sentences.

1.3 The present study

The aim of the present study is to explore the relationship between prediction and the generation of words. Not only have we seen in previous research that when participants suggest words themselves, these are remembered better than words that are presented to them, but we have also seen that participants remember words better when they say them out loud rather than read them silently. We will examine this generation effect and, with the promising results of previous studies, we hope to confirm that prediction is similar to generation.

To summarize, the production effect is the difference between reading words out loud and reading them silently, when words are simply presented to the participant. Alternatively, when using pictures, participants retrieved the words themselves and therefore used the entire production system. In this study we will analyse the extent to which predictability affects the generation effect.

We expect to find a generation effect in this study as we expect a better memory for words that are self-generated from a picture rather than heard. In addition, predictability may also affect memory. Most importantly, we expect that if processing predictable words involves generation, then the memory improvement from generating the words should be smaller for predictable words than for unpredictable words.

2. Methods

2.1 Participants

2.1.1 Pre-test

For the pre-test, a survey was made to test the expectations of the ending of each sentence. Two groups of 22 native speakers of Dutch were asked to participate. A total of 15 men and 29 women participated; the mean age was 22.36. They all signed an informed-consent form in which they agreed to participate voluntarily and gave permission to use their data anonymously. Participants of the pre-test were excluded from the main experiment.

2.1.2 Main study

32 native speakers of Dutch participated in the main study. All participants were between 18 and 35 years old, had normal hearing, had corrected-to-normal vision and did not suffer from dyslexia or other language impairments. A total of 10 men and 21 women participated; the mean age was 23.54. They were all asked to sign an informed-consent form and were paid for their participation. One participant's data was incomplete and consequently this was excluded from the analyses. Hence, 31 participants remained. The local ethics committee at the Donders Institute in Nijmegen approved the experiment.

2.2 Materials

2.2.1 Pre-test: sentence selection

For a total of 140 words, one predictable and one unpredictable sentence was formed both ending in the same word. All sentences were in Dutch. In the predictable sentence, the last word was guessed correctly by the majority of the participants (>63%). In the unpredictable sentence, the last word could have been anything. For example, a predictable sentence with the critical word "broom" would be "He swept the floor with a …". An unpredictable sentence ending with "broom" would be "In the department store she asked for a …". The average length of predictable and unpredictable sentences was 7.9 and 7.83 words, respectively.

During the experiment, participants were asked to complete the sentences. Using the results of the experiment, constraint values were calculated to determine whether the sentences were either highly constraining or weakly constraining. The sentences with the final word missing were presented to participants with the survey software Qualtrics (2014). The 44 participants were divided into two groups and each group was given a different version of the survey. All participants were shown a random selection of 22 sentences (each sentence having 22 responses). The participants were asked to complete the sentence in a meaningful way and to avoid using the same word twice.

Using the survey's data, constraint values were calculated as the cloze probability of the most frequently provided completion, and these were used to select a subset of all sentences. The average constraint values are shown in Table 1. The cloze probability, operationalised in a study by Rommers and Federmeier (2018) as the number of participants giving the critical word as the completion of the sentence, was calculated for both strongly constraining sentences and weakly constraining sentences. A sentence was assumed to be strongly constraining if the critical word was guessed more than 60% of the time. These sentences were removed if the critical word was guessed by fewer than 14 participants (63%). The data showed that strongly constraining sentences had a cloze probability value of 0.86 (range 0.63-1) and weakly constraining sentences had a cloze probability value of 0.01 (range 0-0.05). Furthermore, the sentence constraint was operationalised as the cloze probability of the most frequently provided completion of the sentence. For weakly constraining sentences, the sentence constraint was 0.18 (range 0-0.27), while for strongly constraining sentences the constraint was 0.86, as the most frequently provided completion was in all cases identical to the matching critical word.

2.2.2 Main experiment: design and materials

For the main experiment, 120 pairs of sentences were used, all ending with a critical word which had been replaced by a picture. These pictures were taken from the database of a previous study on language (Piai, Roelofs, Rommers & Maris, 2015). Each pair contained one strongly constraining and one weakly constraining sentence and ended with the same picture, which represented either a predictable or an unpredictable ending of the sentence. Each sentence was presented in one of five conditions. The first manipulation, *predictability*, is shown in Table 1. In the second manipulation, *generation*, the participant would either hear a recording of the word and see the picture simultaneously (heard) or would only see the picture and produce the word themselves (generated). Consequently, four conditions were created: predictable/heard, predictable/generated, unpredictable/heard and unpredictable/generated, all containing 24 words. The remaining 24 words were used in the memory test as a fifth condition: new words.

Table 1: Examples of sentences

Predictability	Sentence	Picture	Word Cloze Probability	Sentence Constraint
High	He swept the floor with a …		0.86	0.86
Low	In the department store she asked for a …		0.01	0.16

Multiple recordings were made of the words, using Audacity software (2000), and these were edited with Praat software (2016). For every word generated, the participant's utterance was recorded to ensure that the participant had named the picture correctly. For every word heard, the participant was also recorded to ensure that they had followed the instructions correctly and had not produced the word themselves. If they had not produced the word themselves or if there was a disfluency, the word was removed from the data analysis.

The 120 sentences were divided into five lists, each being one of the five conditions. A Mix program (Van Casteren & Davis, 2006) was used to create five new lists, which contained 24 words for each condition, creating five different versions of the experiment. Participants would only see one of the five lists. During the first task of the experiment, participants would see four conditions of one of these lists and all five conditions during the memory test. This Mix program ensured that all participants saw the same words and that each list had been thoroughly randomised. Thus, a different list was created for each participant. A total of 192 words were presented in the memory test and contained 24 words in each condition (Table 2). To create a 50/50 ratio for old and new words, 72 new words (fillers) were added to the memory test. All fillers had a concreteness level above 4.5 (out of 5), according to Brysbaert, Stevens, De Deyne, Voorspoels, & Storms (2014). Matlab software (2010) was used to make the presentation and present the instructions to the participants.

The participants were informed that they would be reading sentences for comprehension and that a short comprehension test would follow, but no details about the test were given. They were asked to read the sentences as they would read a magazine.

Table 2: Conditions during memory test

Number	Predictability	Generation
24	Predictable	Generated
24	Predictable	Heard
24	Unpredictable	Generated
24	Unpredictable	Heard
24	New words	-
72	Fillers	-

2.3 Procedure

The experiment consisted of three parts. In the first part, sentences were presented in a word-by-word presentation (Figure 3). Depending on the condition, the participants would either hear or generate the last word of the sentence while being shown the picture. The sentence was presented in either red or blue, indicating whether the participant was required to listen to or to produce the word. To counterbalance the influence of the colours of the words, the colour was randomly switched between participants. All participants were given a total of 96 sentences (48 generated and 48 heard). At the beginning of each sentence, a central fixation cross was shown for 0.5 seconds, followed by a blank screen for 0.3 seconds. Each word was then shown for 0.3 seconds, with a 0.3 second inter-stimulus interval, and the picture was displayed on the screen for 1 second. After the picture, a one-second blank screen followed. Each block contained 12 sentences and participants were allowed a break after each block if required.

Figure 3: A word-by-word presentation ending in a picture

The second part of the experiment was a short distraction task, in order to clear short-term memory. Participants were asked to solve as many simple arithmetic problems as possible in one minute (i.e. addition and subtraction).

The final part of the experiment was the memory test, in which the participants judged words on whether or not they had encountered them before (regardless of whether they had heard them or generated them themselves). From left to right, the options “Sure New” (“Zeker Niet Voorgekomen”), “Maybe New” (“Misschien Niet Voorgekomen”), “Maybe Old” (“Misschien Voorgekomen”) and “Sure Old” (“Zeker Voorgekomen”) were displayed on the screen (Figure 4). Participants used the keys *z, x, n,* and *m* on the keyboard to indicate how well they remembered the words (*z* indicating “Sure New” and *m* indicating “Sure Old”). They could only press one key at a time and the words remained on the screen until one of the keys had been pressed. Once a key had been pressed, the word on the screen would flash in white. A blank screen was then shown for 0.3 seconds before the next word was shown on the screen. The words were presented in green, as shown in the example in Figure 4.

Figure 4: Old/new judgements

2.4 Data analysis

To find out if generation and predictability affect the participant’s memory, the correct and incorrect participants’ responses were compared in each condition.

A frequently used method for measuring participants’ sensitivity independently from response biases is the statistic d-prime. To calculate this d-prime value for all participants per condition, hit rates and false alarm rates were calculated. A hit-rate of 1 means that the participant correctly recognized all the words that had been encountered before. The false alarm rate indicates how often participants incorrectly indicated that a new word had been encountered before. Using the Excel formula Normsinv, the inverse of the standard normal cumulative distribution (Z-score) was calculated. By subtracting the Z-score of the False alarm rate from the Z-score of the Hit rate, the d-prime was calculated. For the hit rates and false alarm rates, a probability corresponding to the normal distribution with a value between 0 and 1 is required, as 0 and 1 themselves yield z-scores of –infinity and +infinity. The rates of “0” were therefore replaced by 0.02 (0.5 divided by the number of items, 24) and the rates of “1” were replaced by 0.98, following Stanislaw and Todorov (1999).

As the experiment contained categorical independent variables with two factors (predictability and generation) with two levels each, and all participants encountered all conditions, we used a 2x2 Repeated Measures Anova in SPSS to compare means.

3. Results

3.1 Hit rates and false alarms rates

Table 3 shows the average mean and maximums of the false alarm rates and the hit rates. We calculated these values for each condition (predicted generated (PG), predicted heard (PH), unpredicted generated (UG) and unpredicted heard (UH)). The false alarm rate of the new words had an average of 0.14, which means that in 14% of these trials participants incorrectly indicated a word had been encountered before. For the other conditions, we calculated the correctly recognized words. The hit rates for the generated condition were 0.95, which means that people correctly recognized these words 95% of the time. However, for the heard condition, lower hit rates between 0.71 and 0.73 were calculated.

Table 3: Averages and maximum rates

	False alarm rates (new)	Hit rates (PG)	Hit rates (PH)	Hit rates (UG)	Hit rates (UH)
Average	.14	.95	.73	.95	.71
Maximum	.33	1	1	1	.92
Participants with 100% rate	0	14	2	12	0

3.2 Comparing d-prime means

Figure 5 shows d-prime as a function of generation (generated, heard) for predictable and unpredictable words. Predictable words generated (*M* = 2.99, *SD* = .68) were better remembered than predictable words heard (*M* = 2.02, *SD* = .72). Additionally, unpredictable words generated (*M* = 2.98, *SD* = .66) were better remembered than unpredictable words heard (*M* = 1.87, *SD* = .64). A Repeated Measures ANOVA confirmed this main effect of generation: participants were significantly better at remembering words they had generated than words they had only heard ($F(1.30) = 175.818$; $p < .001$). No main effect of predictability was found ($F(1.30) = 1.895$; $p = .179$). Most importantly, there was no statistically significant interaction between the effects of generation and predictability on participants' memory ($F(1.30) = 2.367$; $p = .134$).

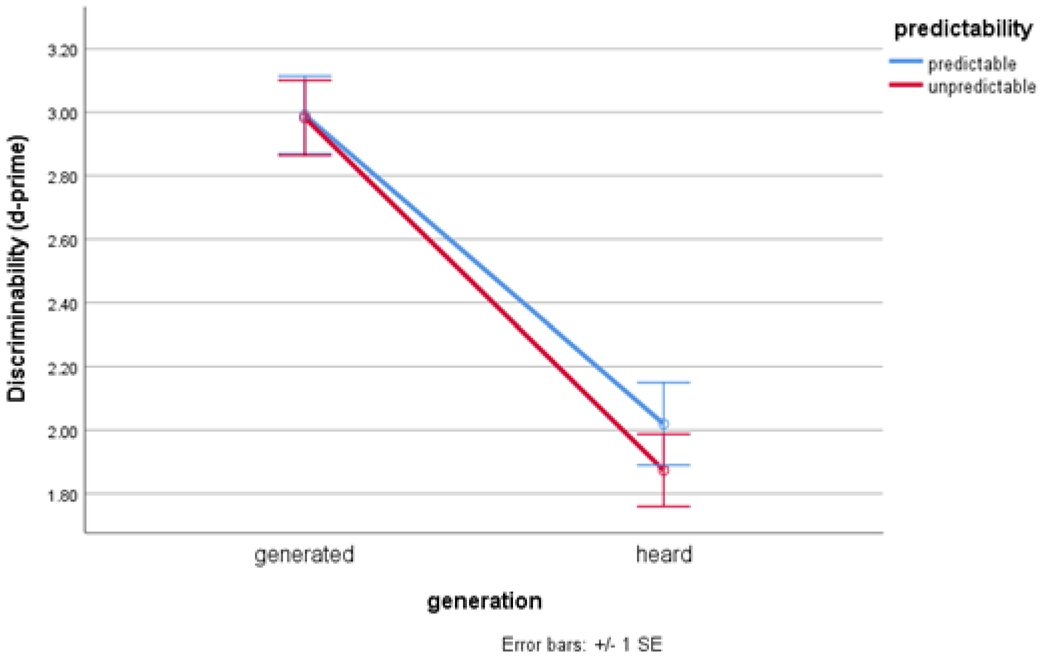

Figure 5: d-prime as a function of predictability and generation.

4. Discussion

The results clearly showed a major generation effect. Participants remembered words significantly better if they had generated the words themselves rather than if they had simply heard them. This confirms that the generation effect is robust and can be found not just for generated antonyms or synonyms (Slamecka & Graf, 1978), but also for named pictures relative to pictures accompanied by words (see also Zormpa et al., 2019).

More importantly, the interaction effect between predictability and generation was not significant. One possible reason for the lacking interaction is that predictable words are not generated using the production system. This would contrast with earlier research (Dell & Chang, 2014; Federmeier & Kutas, 1999; Federmeier et al., 2010; Hintz et al., 2017; Martin et al., 2018; Pickering & Garrod, 2007; Rommers et al., 2015; Wlotko & Federmeier, 2007). However, a small noticeable reduction of the production effect was visible in Figure 5, and one could argue that a ceiling effect was present.

Further research should be conducted on this topic to establish whether word predictions are generated by the production system, while avoiding ceiling effects. Firstly, a suggested improvement to this experiment would be to use more items. Three out of four conditions were guessed completely correctly by at least one participant. Participants clearly remembered too many words to observe an interaction. A follow-up study could begin by presenting more items and then extending the distraction task to prevent this ceiling effect from occurring. Participants could be asked to resume the second part of the experiment, the memory test, for example the next day, possibly preventing the observed ceiling effect. Secondly, in addition to modifying the design and the duration of the distraction test in between the experiment and the memory test, effect sizes could be calculated to obtain a better indication of the number of participants needed for statistical power.

5. Conclusion

The results clearly showed a generation effect for naming pictures. In contrast, we could not find a significant effect for predictability. Most importantly, if processing predictable words is like generation, then the memory improvement from generating the words should be smaller for predictable words than for unpredictable words. However, this interaction effect was not found, perhaps because of a ceiling effect. Suggestions for improvement were discussed.

Although the current study is based on a small sample of participants and could not confirm the expected interaction effect, the results did point in the expected direction. Further research should be conducted to establish whether prediction is like generation, by adding more items to the test and extending the distraction task.

References

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer software]. *Version 6.0.55*. Retrieved from http://www.praat.org.

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.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, *108*(1), 204.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature neuroscience*, *8*(8), 1117.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120394.

Fawcett, J. M., & Ozubko, J. D. (2016). Familiarity, but not recollection, supports the between-subject production effect in recognition memory. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*(2), 99.

Federmeier, K. D., & Kutas, M. (1999). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*(3), 373-392.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and language*, *115*(3), 149-161.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of psycholinguistic research*, *30*(1), 3-20.

Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of memory and language*, *25*(3), 348-368.Ghatala, E. S. (1983). When does internal generation facilitate memory for sentences?. *The American Journal of Psychology*, 75-83.

Ghatala, E. S. (1983). When does internal generation facilitate memory for sentences? *The American Journal of Psychology*, 75-83.

Graves, W. W., Grabowski, T. J., Mehta, S., & Gordon, J. K. (2007). A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of cognitive neuroscience*, *19*(4), 617-631.

Hintz, F., Meyer, A. S., & Huettig, F. (2017). Predictors of verb-mediated anticipatory eye movements in the visual world. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1352.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(1), 1-38.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of experimental psychology: Learning, memory, and cognition*, *36*(3), 671.

Martin, C. D., Branzi, F. M., & Bar, M. (2018). Prediction is production: The missing link between language production and comprehension. *Scientific reports*, *8*(1), 1079.

MATLAB. (2010). *version 7.10.0 (R2010a)*. Natick, Massachusetts: The MathWorks Inc.

Mazzoni, D., & Dannenberg, R. (2000). Audacity [software]. *The Audacity Team, Pittsburg, PA, USA*.

Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. *Human brain mapping*, *36*(7), 2767-2780.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension?. *Trends in cognitive sciences*, *11*(3), 105-110.

Qualtrics, L. L. C. (2014). Qualtrics [software]. *Utah, USA: Qualtrics*.

Rommers, J., Dell, G. S., & Benjamin, A. S. (2018). Predictable words leave production-like traces in memory. *Poster presented at the International Workshop on Language Production (IWLP), July 2-4, Nijmegen, The Netherlands.*

Rommers, J., & Federmeier, K. D. (2018). Predictability's aftermath: Downstream consequences of word predictability as revealed by repetition effects. *Cortex*, *101*, 16-30.

Rommers, J., Meyer, A. S., & Huettig, F. (2015). Verbal and nonverbal predictors of language-mediated anticipatory eye movements. *Attention, Perception, & Psychophysics*, *77*(3), 720-730.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of experimental Psychology: Human learning and*

Memory, *4*(6), 592.
Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers*, *31*(1), 137-149.
Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443.
Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior research methods*, *38*(4), 584-589.
Weldon, M. S., & Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition*, *15*(4), 269-280.
Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience letters*, *346*(3), 165-168.
Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of cognitive neuroscience*, *16*(7), 1272-1288.
Wlotko, E. W., & Federmeier, K. D. (2007). Finding the right word: Hemispheric asymmetries in the use of sentence context information. *Neuropsychologia*, *45*(13), 3001-3014.
Zormpa, E., Brehm, L. E., Hoedemaker, R. S., & Meyer, A. S. (2019). The production effect and the generation effect improve memory in picture naming. *Memory*, *27*(3), 340-352.