Articulation rate as a means of distributing information and its effect on the N400-component

Christoffer Forbes Schieche

Department of Linguistics
Bachelor's Thesis 15 ECTS credits
Linguistics – Bachelor's Course, LIN600
Bachelor's Programme in Linguistics 180 ECTS credits
Spring semester 2021
Supervisors: Robert Östling, Petter Kallioinen
Swedish title: Distribution av information med hjälp av artikulationshastighet och dess effekt på N400-komponenten

Stockholm University

Articulation rate as a means of distributing information and its effect on the N400-component

Christoffer Forbes Schieche

Abstract

Information theoretical approaches to language state that the most efficient communication occurs when the amount of information transmitted is distributed as uniformly as possible over time. Previous research has shown that speakers tend to adhere to strategies for distributing information efficiently, using mechanisms at multiple linguistic levels. This study aims to investigate whether articulation rate (AR) is used in continuous speech to achieve a more uniform distribution of information within sentences, quantified as surprisal estimated by the state-of-the-art language model GPT-2, and if this has an effect on the amplitude of the N400 brain response in listeners. In neurolinguistics, surprisal has been observed to be a good predictor of the N400, which is related to processing of semantics and meaning in general. The results showed a significant, though small, effect of surprisal on AR, indicating that AR may have some role in achieving more uniform distribution of information on the word level. In line with previous research, surprisal showed an effect on the N400 where higher surprisal led to larger amplitudes. Results regarding AR and distributional effects on the N400 were inconclusive, although some independent effects of AR were found that could be further explored in more controlled experimental settings.

Keywords

Information theory, surprisal, articulation rate, EEG, ERP, N400

Distribution av information med hjälp av artikulationshastighet och dess effekt på N400-komponenten

Christoffer Forbes Schieche

Sammanfattning

Informationsteoretiska perspektiv på språk säger att den mest effektiva kommunikationen sker när information sänds ut så jämnt fördelat som möjligt över tid. Tidigare studier har visat att talare tenderar att följa vissa strategier för att distribuera information jämnt, vilket de gör på flera språkliga nivåer. Denna studie ämnar att undersöka om artikulationshastighet (eng. *articulation rate* (AR)) används i kontinuerligt tal för att uppnå en mer jämn distribution av information inom meningar, kvantifierat som informationsteoretisk *surprisal* med hjälp av språkmodellen GPT-2, samt om detta ger effekt på hjärnresponsen N400:s amplitud hos lyssnare. Inom neurolingvistik har surprisal visats kunna predicera N400, som är kopplad till bearbetning av semantik och meningsfullhet generellt. Resultaten visade en signifikant, om än liten, effekt av surprisal på AR, en indikator på att AR kan ha en roll i att uppnå mer jämn distribution av information på ordnivå. I linje med tidigare forskning så hade surprisal en inverkan på N400, där högre surprisal gav större amplituder. Resultaten utifrån AR och distribution av information var inte entydiga, däremot observerades vissa självständiga effekter av AR på amplituden av N400 och dessa skulle kunna vidare undersökas i mer kontrollerade experiment.

Nyckelord

Informationsteori, surprisal, artikulationshastighet, EEG, ERP, N400

Contents

1 Introduction

Surprisal is an information theoretic concept quantifying the unexpectedness of an upcoming item given previous contextual information. Applied to language, this can be measured on multiple levels, e.g., phonemes, syllables, and words. Accounts in linguistics following information theory assume that language users aim to be economic and efficient in that the most optimal communication would strive to output an even amount of information over time (or segment), also known as information density (Fenk-Oczlon 2001; Jaeger 2006).

One such account is the Uniform Information Density hypothesis, or UID (a.o. Jaeger 2010; Jaeger 2006). UID and similar approaches have successfully shown or predicted behaviors among speakers that support the assumption that they aim for a more uniform rate of information over time. How speakers manage information density have been observed on several levels of language, such as syllabic or word duration, omission or reduction of segments and word choice (Aylett and Turk 2004; Jaeger 2010; Mahowald et al. 2013).

Articulation rate (AR), commonly measured as syllables per second have been shown on the utterance level to change as an effect of surprisal, where more informative utterances are pronounced at a lower average rate (Sjons, Hörberg, et al. 2017). On the word level though, where individual words inside of an utterance could vary in their respective AR, its relationship to surprisal has not been explicitly studied, even though syllabic and word duration can be considered closely related (Aylett and Turk 2004; Bell et al. 2009); AR could possibly be another level or means of achieving more uniform information density on individual units inside of an utterance.

Additionally, surprisal and information density is not only applicable for study on the side of speakers, but can be studied in how it affects the processing of incoming linguistic material for a listener. Since its identification in the 1980s, the EEG/ERP-component N400 has been shown to be sensitive to semantic incongruities and failed predictions of upcoming linguistic material (Kutas and Federmeier 2011; Kutas and Hillyard 1980). More recent studies have found that surprisal can be a good predictor of the amplitude of the N400 (e.g. Frank, Otten, et al. 2015); the more unexpected an item is, the larger the negative amplitude, taken as indicator of more processing or cognitive load required.

This study aims to investigate surprisal and AR, both from a speaker's as well as a listener's perspective. In the case of speakers, whether the AR of individual words are affected by surprisal and thus how speakers may utilize it to achieve a more uniform information density. For listeners the aim is to investigate how surprisal and AR may affect the N400 amplitude. The state-of-the-art language model GPT-2 (Radford et al. 2019) will be used to estimate surprisal of individual words in Hemingway's novel *The Old Man and the Sea*. The study will use audio and EEG-data from a previously performed EEG-recording where participants listened to excerpts of the novel (Broderick et al. 2018; Di Liberto et al. 2015).

2 Background

2.1 Measuring information

2.1.1 Information theory

The field of information theory, in large part defined by Shannon (1948), concerns the quantification of information, how it is communicated or transmitted over a channel, and the distribution of information in a transmitted signal. The transmission of information in this theoretical sense should not be thought of as actual informative content of a message, such as the semantic meaning of a word, but rather how much new information a particular outcome brings within the context in which it appears. Though information theory as proposed by Shannon (1948) started out as "a mathematical theory of communication" (the title of the 1948 paper) closely linked to telecommunications (Hale 2016), it has been applied in numerous fields, among them linguistics (such approaches will be specifically described in section 2.1.3–4).

The quantification of information in regards to information theory goes by several names, e.g., Shannon information (Frank and Jaeger 2008), self-information, or surprisal (the term used in this thesis) (Frank, Otten, et al. 2015), and is a measure of the unexpectedness of a particular outcome given its probability in that context. In regards to language, the information-carrying units of interest could pertain to several linguistic levels, e.g., phonemes, part of speech, and words (for this study, words will be the main focus). Regardless of terminology, the definition of this measure is the negative, or inverse, logarithm of an outcome's probability. Shannon (1948) used base-2 for the logarithm and referred to these units of information as bits (from *binary digits*), but other bases can be used (e.g., Frank, Otten, et al. 2015, used base-e). Surprisal of a word measured in bits can thus be defined as:

$$surprisal(word_i) = -\log_2 P(word_i|context)$$

The relation between an outcome's probability and its surprisal entails that a highly probable outcome will have low surprisal, or carry little new information as its occurrence is more expected, whereas an outcome with low probability will have high surprisal. For a simple, non-linguistic, example (using base-2): a regular six-sided die will have equal probability (1/6) and surprisal (2.58 bits) for each side. But, rolling a die where the faces four and five have been replaced by sixes, rolling a six will be more probable (1/2) and less surprising and will accordingly be thought as to carrying less information (1 bit).

In regards to the channel in information theory, it is, in short, the medium which a signal, and thus information, is transmitted with from a source to a receiver (Shannon 1948) [1]. This could be a cable, or a Wi-Fi signal; for language it is the linguistic symbols at our disposal, whether spoken, written or signed. The amount of information passing through a channel over time can be referred to as information rate or information density. If two outcomes carry the same amount of information but one is shorter in duration, the information density of the shorter one will be higher. The channel, whatever medium or type of information transmitted, will have a certain limit or maximum capacity in regards to the amount of information that can be passed through it per time unit; exceeding this limit would lead to erroneous transmission. The most optimal, effective, and economic use of a channel would be to let the information density in any message lay as close as possible to the limit without exceeding it; the most amount of information is continuously transmitted without it risking to be erroneously decoded on the receiving end.

[1] While Shannon made a a distinction between noiseless and noisy channels, this distinction will not be further discussed in this study. The channel referred to in this study is the noisy channel.

From an information theoretic perspective, language can be seen as transmitting information over a noisy channel with a limited capacity (or bandwidth) (Shannon 1948). Effective language use would thus be, as with any transmission, to keep the information density in the signal at an optimal level close to the channel's maximum capacity without exceeding it. In turn, providing too little information over time, having too low information density, would be ineffective and uneconomical (Fenk-Oczlon 2001), or not "socially useful and acceptable" (Pellegrino et al. 2011, p. 540). Imagine someone presenting a lecture on quantum physics at high speed (and the listener being a novice in the field), vs. someone reading a children's story at a very slow pace. The first scenario could result in a lot of the content being lost on the listener, whereas the second would simply be, at least for an adult, nothing but frustrating.

That language may adhere to these kinds of principles of optimality and efficiency was already present with Shannon (1948). The following sections will describe some past observations regarding how probabilities and surprisal seem to influence linguistic form as well as how speakers may be affected in the way they structure their utterances.

2.1.2 An example regarding relative frequency, context, and surprisal

For language, surprisal is a measure of how unexpected a linguistic unit is given a specified context. If no context is considered, a word's surprisal would simply be a reflection of its relative frequency in a given language. Consider the following sentences:

$$cats_1\ love\ dogs_1 \qquad cheetahs\ love\ dogs_2 \qquad dogs_3\ love\ cats_2 \qquad cats_3\ love\ cheetahs$$

In this language, the word type *cats* appears as many times as *dogs* (three) and their relative frequencies are 0.25 each ($3/12$). But, when including an immediately preceding word as context, for instance *love*, the resulting conditional probability of the next word being *dogs* is 0.5, since *love* is followed by it in half the cases that it appears. Expressed in bits, the difference between *dogs* and *cats* with context *love* are:

$$surprisal(cats_2) = -\log_2 P(cats_2|love) = 2\ bits$$
$$surprisal(dogs_2) = -\log_2 P(dogs_2|love) = 1\ bit$$

For this thesis, surprisal and bits will be used when referring to measures where the context is greater than zero, else (relative) frequency will be used.

2.1.3 Frequency and surprisal effects on word form and duration

It has long been observed that the frequency of a word has an inverse relation to its length. Zipf (1935) concluded that the more frequent a word is the shorter it will be, in respect to orthographic form, the number of syllables, or phonemes. This principle is well established, the so called Zipf's law of abbreviation, and may also account for how linguistic forms change in accordance with changes in their usage. But, relative frequency as sole predictor of linguistic form may be considered a slight simplification in that it does not consider context as a factor.

As previously exemplified, two distinct words may have the same relative frequency, but when introducing context as a factor, their probabilities will be conditioned on where and how they appear. With this in mind, Piantadosi et al. (2011) investigated the correlation between frequency and form both when including context and solely looking at a word's relative frequency (Zipf 1935). Including context (one, two or three preceding words) the correlation with orthographic word length was significantly higher than if only using a word's relative frequency and

this correlation was present over several languages. So while Zipf's law of abbreviation is not faulty in its observation, Piantadosi et al. (2011) propose a slight revision. Rather, it is the more predictable words that are shortened, which are not necessarily the most frequent ones (although they do not rule out that frequency has some partial role as well). These observations, both Zipf's "simpler" relative frequency and Piantadosi et al.'s (2011) inclusion of context, fall in line with information theoretic views where more even information density seems to be preferred, in this case that less expected words are orthographically longer.

When considering such correlations in spoken language, quite intuitively similar patterns of frequency on words can be found where more frequent words will be shorter in duration. While orthographic form may give a rough estimate of duration (at least in relation to other words), Gahl (2008) showed that in continuous speech, homophone pairs differed in duration, regardless if they also happened to be homographs, when their respective relative frequencies differed. For example, even though they phonologically consist of the same linguistic segments, *time* had shorter duration than *thyme*, *time* being the more frequent version.

Further evidence for that context, not just frequency, influences the duration of individual words have been shown by Bell et al. (2009), who investigated the effects of frequency and predictability on word duration in spontaneous speech of American English speakers, as well as if these effects differed between function and lexical words. For lexical words, there were significant effects of frequency and conditional probability, though the latter was significant in regards to the probability given a following word, rather than a preceding one-word context. In both cases, a more frequent or probable word had shorter duration. Function words on the other hand showed no significant effects of frequency but did reach significance with the preceding word as context, at least for highly frequent function words. While the results for lexical words differ in regards to what words, a previous or following one, have an effect as compared to, for instance, Piantadosi et al.'s (2011) results regarding orthographic length, they still suggest that probability derived from context influences the duration of words. A comparison of the results for function and lexical words indicate that these two general classes are differently affected, with content words generally more sensitive to these effects.

While the previously described studies observed effects on length and duration of full words, effects have also been observed on the duration of individual syllables, thus affecting the overall duration of a word. Using a large corpus of spontaneous speech between Glaswegian English speakers, Aylett and Turk (2004) observed an inverse relationship between the probability or expectedness of a syllable and its duration. This relationship was seen both when including the previous two syllables as context for the probability of an upcoming one, as well as the absolute frequency of the word that the syllable was part of [2]. The more expected a syllable was, either based on conditional probability or frequency, the shorter its duration, leading to an overall change in the duration of the whole word.

Furthermore, it's not necessary that frequency and surprisal only affect the duration of specific word forms and syllables, but can also affect what word form is chosen if there are more than one available. Using word pairs such as *chimp* and *chimpanzee*, one being a shortening of the other, Mahowald et al. (2013) showed through a corpus study (using the same material with a three word context window as Piantadosi et al. 2011) that the longer forms tended to have higher surprisal, resulting in a spreading out of the information density. The same study included a behavioral test where participants were asked to finish sentences that were either neutral as to what word could finish the sentences, or where the word was more predictable from the context. The results showed a preference for the shorter form in the predictable conditions, assumed as an effect of it being more efficient and economical. Even though the behavioral study did not

[2] Both of these measures, probability and frequency, were log-transformed and referred to as *redundancy*.

quantify surprisal, the results seem to follow information theoretic accounts where the longer form in a predictable context would have resulted in information density lower than the optimal rate, thus uneconomical (Fenk-Oczlon 2001).

2.1.4 Effects of surprisal beyond single words

The previous section described how frequency and surprisal affect language on word level, with duration and reduction effects. The influence of surprisal on language is not bound to this level and effect only, but may have larger implications in how language use is organized.

Jaeger (2006) proposed the Uniform Information Density hypothesis (or UID), which follows an information theoretic approach for language production. UID states that, assuming the presence of an optimal information density, speakers would attempt to formulate their utterances in order to achieve this optimal information density, seen as the most efficient communication. This is argued to take place not only on the level of words, but any linguistic level. According to UID, speakers would wish to avoid great variations in information density over time (Jaeger 2010) to keep the information density as constant as possible at the optimal rate, which can influence the formulation of utterances. In case there are multiple ways of formulating an utterance, speakers would prefer the one where information density is the most uniformly distributed (as seen on word level in Mahowald et al. 2013).

A case where UID have predicted differences in regards to how speakers form their utterances in spontaneous speech can be found in the inclusion or omission of the optional complement clause marker *that*, as in the sentence *My boss confirmed* [(*that*) *we were absolutely crazy*], where the square brackets indicate the complement clause. Jaeger (2010) found significant effects of information density on whether or not *that* was included, where if the onset of the complement clause (*we* in the above example) had high surprisal or information density, speakers were more likely to include *that*. This inclusion results in a spreading out of information (over two words, e.g., *that* and *we*) thus lowering and evening the information density. In case the onset word would already have low surprisal, the inclusion of the marker would be considered less economical, as the spreading of information would put the information density at a lower than optimal rate ("an uneconomical expenditure of signs, time, and energy", Fenk-Oczlon 2001, p. 436).

For this study, aspects of UID and the spreading of information density will be tested, both in relation to how a speaker adapts their utterances as an effect of surprisal and informativity, but also how these modulations may affect listeners' processing of words.

2.1.5 Estimation of surprisal using language models

In order to make estimations of surprisal of words, language models (or LMs) are utilized. The task of any such language model, of which there are several types, is to predict the probability of a linguistic unit given a context of units with which it appears; this probability can in turn be log-transformed into surprisal.

N-gram models assign probabilities of a word or sequence given a previous context of size n [3]. These types of models have proven fruitful in order to show correlations between surprisal and word forms, such as in previously mentioned Piantadosi et al. (2011) and Mahowald et al. (2013). There are limits to the functionality of n-gram models though: when increasing the context window n, the number of available units and sequences quickly rises to unmanageable

[3] As previously described, a context window of zero, i.e., unigram, simply reflects a word's relative frequency, void of context.

sizes, as does the number of sequences of size *n* that the model has not explicitly encountered that it may at some point attempt to estimate the probability of (Bengio et al. 2003).

In order to overcome some of n-gram models' limitations, neural networks have been used in language modeling for some time; Bengio et al. (2003) implemented one of the first more successful ones, insofar that it outperformed n-gram models. When training such models, it will attempt to predict outcomes and in effect change internal parameters in order to increase its precision. This also allows such models to better handle cases that it hasn't seen before. Further improvements on neural network models have been Recurrent Neural Networks and Long-Short-Term-Memory networks (Sundermeyer et al. 2012), which can better increase context windows and selectively "remember" information over wider distances.

A recent contribution to language modeling is the GPT-2 model (Radford et al. 2019), the model to be used in order to estimate surprisal in this study. GPT-2 is a deep neural model, utilizing up to 1.5 billion parameters, based on self-attention (Vaswani et al. 2017). It is trained on a data set called WebText, consisting of scraped text material from websites linked from Reddit [4]. Furthermore, it uses a compression technique called Byte Pair Encoding (BPE) (Sennrich et al. 2016); instead of training on every word in the data as is, they are broken down into smaller sub-word units. These are strings of characters who frequently occur together, e.g., *lower* may be split into *low* and *er* [5]. This technique has improved the capability of language models to estimate probabilities and surprisal of words that they have not encountered before, since they can rather estimate the sub-words separately. Estimating the surprisal of an actual word can be done by adding together the surprisal of its individual sub-words; since surprisal is log-transformed probability, this is equal to multiplying the sub-words' probabilities, resulting in the joint probability of these sub-words occurring together.

2.2 EEG, ERPs and the N400

Electroencephalography, or EEG, is a non-invasive technique that allows the measuring of electrical activity in the brain. By placing electrodes on the scalp, the electrical activity of neurons can be recorded at a high temporal resolution, down to milliseconds (though with a spatially poor resolution, the inverse of techniques such as fMRI). EEG can thus be used to observe neuronal activity in subjects in a temporally fine-grained manner while presenting them with stimuli from a range of modalities (e.g., auditive, visual).

By averaging the EEG signal's response time-locked to certain stimuli, noise in the signal is mitigated and event related responses (ERP) can be obtained. In ERPs, several components have been found to occur with stable latencies and polarities as response to certain kinds of tasks, stimuli, or changes in the stimuli-environment (see Luck 2014; Kemmerer 2015, for more in-depth descriptions of EEG-technique and ERPs).

2.2.1 The N400 component

The N400 is a negative ERP component that peaks at approximately 400 ms after onset of meaningful stimuli. The N400 was first identified by Kutas and Hillyard (1980), in a study where sentences were visually presented word by word, with the final word either semantically congruent with the preceding context, semantically related to a congruent word, or semantically incongruent (though syntactically correct). Previous studies had shown the occurrence of a positive

[4] https://www.reddit.com

[5] For this particular example (taken from Sennrich et al. 2016), this corresponds to the word's morphemic boundaries—this is not always the case.

component called P300 when participants were presented with low probability non-linguistic stimuli among repetitions of another stimulus (Tueting 1978). Thus, it was expected that this effect would occur when using linguistic, meaningful, material as well under the assumption that preceding sentence context in some way or other builds expectation of what linguistic items may follow.

The incongruent sentence endings (analogous with unexpected items) did not, as had been expected, elicit a P300 response. Rather, semantic incongruity resulted in the aforementioned negative response peaking roughly at 400 ms, with the semantically related target also eliciting a negative response, albeit not as strong in amplitude.

Since its discovery, the N400 has become a highly studied component, especially in research of cognitive processes tied to linguistic comprehension (see Kutas and Federmeier 2011, for a review). The N400 has been shown to appear in experiments using written, spoken as well as signed language but have also been elicited by non-linguistic material. Thus, rather than being a component reflecting linguistic processing exclusively, the range of modalities with which it can appear rather suggests that it—in some way—reflects the processing of any meaningful stimuli; that it serves as a response to stimuli that in one way or other clashes with or doesn't fulfill expected outcomes. This can be on the level of semantic incongruity of words in isolated sentences (e.g., Kutas and Hillyard 1980) but also on higher discourse levels and in relation to general world knowledge (Hagoort et al. 2004).

2.2.2 N400 and surprisal

Psycholinguistic studies using eyetracking and reading time have shown correlations with surprisal (Smith and Levy 2013), where the longer duration of reading words with higher surprisal (i.e., carrying more information) suggests more cognitive processing needed. While reading time studies may give hints as to underlying cognitive processes, the measures of EEG/ERP may shed more light on the neural underpinnings of cognition.

As described in the previous section, the N400 seems to reflect some underlying cognitive processes occurring when a mismatch between expected outcomes and actual outcomes presents itself, based on expectation or prediction from the previous context (see Van Petten and Luka 2012, for a discussion whether linguistic comprehension, and thus the N400, involves prediction of specific units or general semantic expectations). Similarly, surprisal is a (computationally driven) measure of unexpectedness given the outcome of a certain context.

Though somewhat nascent, some studies investigating the relationship between surprisal and the N400 have been carried out in recent years. In a study where participants read English sentences, Frank, Otten, et al. (2015) found a correlation between surprisal and the size of the N400 component with higher surprisal of a word resulting in larger negative amplitudes. Using several language models in the estimation of surprisal, n-gram models did perform best overall, though a Recurrent Neural Network model performed slightly better when only looking at lexical words. Similar results where found in Aurnhammer and Frank (2019) when using the same EEG-data but using a Long-Short-Term-Memory model.

Similarly, Yan and Jaeger (2020) used excerpts from existing novels and again found correlations between surprisal and the amplitude of the N400. Furthermore, they found that surprisal, which is log-transformed probability, was a better linear predictor than non-transformed probability, indicating that surprisal can be an important tool in future work of predicting the N400.

2.3 The speed of speech

During speech production, a speaker may modulate several prosodic features of the signal, such as pitch, stress, duration and tempo. These modulations may serve different functions; rising in pitch at the end of an utterance may indicate a question (even when word choice and/or order could already have indicated the utterance as such), whereas lowering one's pitch could indicate that the speaker is finishing up. Duration has in part previously been described in section 2.1.3, but a related feature is tempo: the speed or rate at which speech is produced.

To measure this rate, specified linguistic units are used, such as words or syllables and then calculated how the rate of production of these varies over time. Two common such measures are speech rate (SR) and articulation rate (AR). The main difference between them is that SR includes pauses between linguistic units whereas AR does not, or at least has a specified limit of how long a silence can be to consider two segments part of the same unit or not (e.g., Sjons and Hörberg 2016, used an upper limit of 300 ms). This study will look at articulation rate on the word level and use the measure syllables per second.

Speakers' AR is not at a constant rate over time, and these changes may occur for different reasons. Some may of course belong to idiosyncrasies of particular speakers, but some may be more systematic. The context of an utterance, or the style in which it is spoken, has been shown to affect the rate of speakers' production. Jacewicz et al. (2009) compared read sentences and spontaneous speech in two dialects of North American English; the average AR was lower for read sentences (3.4 syllables/second) than in spontaneous speech (5.12 syllables/second). The study also found some significant results related to the particular dialect, age and gender of the speaker, indicating that certain social aspects may also influence AR. The age of the person one speaks to have also been shown to affect AR; Sjons and Hörberg (2016) showed that in child directed speech, AR increases as a function of age. Sjons, Hörberg, et al. (2017) also found a difference between child directed speech and adult directed speech, where AR is higher when speaking to other adults. The same study also found a negative correlation between AR and surprisal on utterance level, suggesting that speakers modify the rate of speaking after the amount of information in an utterance.

While measuring the speed of speech may more commonly be made over utterances, some work has been done approaching AR on the word level specifically [6]. The duration of a syllable segment in a word have been shown to be affected by the amount of syllables preceding and succeeding it, resulting in that the overall syllable count of a word affects the rate in which it is spoken (Lindblom and Rapp 1973). Specifically, when increasing the number of syllables in a word it will be spoken at a higher rate. Lindblom and Rapp (1973) also accounted for, though not by this term, the known phenomenon of final lengthening: the final word of a phrase or sentence will be spoken at a slower rate than if its position had been earlier in the sentence. A group of studies by Dankovičová (1999), including both Czech and English speakers, showed similarly to Lindblom and Rapp (1973) that AR varied on (phonological) word level depending on the number of syllables in it. Furthermore, it was also observed that a word's position in a phrase (not only final position as in final lengthening) had an effect on AR, as well as part of speech to some degree. In regards to surprisal and AR on the word level, individual segments, i.e., syllables, (Aylett and Turk 2004) and words (Bell et al. 2009) have been shown to vary in duration due to surprisal-like measurements and this would entail differences in AR since words', and their segments', surprisal vary based on context.

Thus far AR has been described only in relation to the speaker, but how does a listener perceive changes in AR? As previously described, Sjons and Hörberg (2016) and Sjons, Hörberg,

[6] Note that when looking at word level, SR and AR can be argued to be the same measurement.

et al. (2017) showed that speakers changed their AR depending on who they spoke to, in regards to age group, and the amount of surprisal in their utterances. Both of these changes could serve the purpose of aiding the listener to better understand, or decode, what is being conveyed by managing information density; though it may be that these changes do not have to be directly perceivable to be helpful for a listener. To put the size of perceivable change in AR in perspective, Eefting and Rietveld (1989) concluded that the Just Noticeable Difference (JND) of AR was approximately 4.43 % on the sentence level. This would mean that a sentence produced at 5 syllables/second need a change of ±0.215 syllables/second to be deemed faster or slower by a listener in comparison to a baseline [7].

2.3.1 AR and N400

Studies of potential effects of AR on the N400 are hard to find, but a possible example for slight comparison involving changes in speed could be Wlotko and Federmeier (2015). Similar to Kutas and Hillyard (1980), the study presented sentences visually where final target words were semantically congruent, semantically related to a congruent word, or incongruent to the preceding context, but this study also altered the presentation rate of the words.

At a lower presentation rate (two words visually presented per second) the N400 amplitude was low when the final word was semantically congruent and high when incongruent, with responses to semantically related words somewhere in-between. When increasing the presentation rate (four words per second), the related words' amplitudes increased to be similar to levels of the incongruent ones. The results were interpreted as that the change in speed may influence the amount of semantic prediction allowed to take place, insofar that the higher speed limits the range only to explicitly predictable words based on context, not including words that may be semantically related.

While the study by Wlotko and Federmeier (2015) may not be entirely analogous for a study into AR on the word level, it shows the potential that changes in presentation or production rates may influence the processing of linguistic information as measured with the N400 component.

2.4 Aim and research questions

Following information theoretic approaches to language production, such as UID, speakers aim for producing utterances that uniformly distribute information (surprisal) with several linguistic levels at their disposal to achieve this. This study aims to investigate how AR on the word level manages distribution of information density. Since the N400 has previously been shown to correlate with surprisal, but its relationship to AR has not been well-researched, this study additionally aims to investigate the role of AR, alongside surprisal, in how they may affect the N400 component's amplitude in a listener.

Thus follow two research questions:

1. Does AR function as a means of distributing information uniformly in continuous speech?

2. Does variation in AR and the resulting distribution of information affect the amplitude of the N400?

[7] The study does note that it cannot completely rule out that other acoustic effects due to transformation of stimulus contributed to the perception of difference.

3 Method

3.1 EEG-data

No novel EEG-data was recorded for this study. Rather, data collected for Di Liberto et al. (2015) and Broderick et al. (2018) (which is publicly available [8]) was used. This data consists of EEG-recordings from 19 native English speakers (age range 19–38, six women) listening to excerpts from Hemingway's novel *The Old Man and the Sea*, read by a male American English speaker. Each excerpt, 20 in total, were approximately three minutes long. The recordings were made with a Biosemi system using 128 channels with two mastoid channels for online referencing at a rate of 512 Hz. Before made publicly available, the data was downsampled to 128 Hz.

The raw data was further pre-processed using the Python package MNE (Gramfort et al. 2013) to obtain ERP-waves for each participant and electrode per token [9]. One participant was excluded due to faulty data. In order to filter out some of the noise and potential artifacts in the signal, a high-pass filter of 0.1 Hz and a low-pass filter of 40 Hz were applied; these limits fall in the range of standard recommendations (Luck 2014, Chapter 5). The window –100–0 ms was used as baseline for the ERPs and the estimated amplitude of the N400 was calculated as the average of all measured amplitudes in the window 300–500 ms. While the latency of the N400 is fairly stable with a peak around 400 ms after stimulus onset, there is some typical variation that a window range of ±100 ms would cover (Duncan et al. 2009). If any amplitude value inside this window was larger than ±100 μV, the whole window was rejected for that electrode.

Though the full range of 128 channels were available, only the centro-parietal Cz-electrode was used in this study, due to that N400 effects have been shown to be more visible at this and neighboring centro-parietal electrodes (Kutas and Federmeier 2011).

3.2 Language model

To acquire surprisal for each word, the GPT-2 language model (Radford et al. 2019) was used through the the Python package lm_scorer [10] [11]. A context window of up to 1,000 tokens were used for each target token. Though possible to estimate surprisal with different sized context windows, only this window was used for this study since it was considered suitable for a novel where themes may reappear after several sentences. To avoid breaking up sentences, the window was rounded up to always include full sentences. The model was given each token and its corresponding context and returned its estimated surprisal; since GPT-2 uses BPE-tokens, a word in the text can be represented by more than one token by the model. In such cases where a word in the text was represented by several BPE-tokens, these and their respective estimated surprisals were added together to match the word in the original text, resulting in the joint surprisal of its individual parts.

3.3 Hemingway's *The Old Man and the Sea*

The total number of tokens, inter-punctuation excluded, in the excerpts from *The Old Man and the Sea* were 11,304. Due to possible effects of final lengthening, all words preceding a ".",

[8] https://datadryad.org/stash/dataset/doi:10.5061/dryad.070jc

[9] All pre-processing was done by thesis supervisors Robert Östling and Petter Kallioinen prior to work on this thesis started.

[10] https://github.com/simonepri/lm-scorer

[11] Acquiring surprisal for the text was done by Robert Östling prior to work on this thesis started.

Table 1: Absolute frequencies of tokens and types in the excerpts of *The Old Man and the Sea*, excluding sentence final words. Note that due to homographic overlaps, the sums for types do not add up.

Part of Speech	Tokens	Types
All words	10495	1484
Function words	5955	220
Lexical words	4540	1328
Nouns	1661	558
Verbs	1469	481
Adjectives	667	217
Adverbs	743	146

"!", or "?" were removed from any analysis. This amounted to 809 tokens being removed, approximately 7.2 % of the total token amount, resulting in a final data set of 10,495 tokens.

The text was part of speech-tagged with the Python package Stanza (Qi et al. 2020) using the Universal Dependencies tag set [12]. This allowed separating function words and lexical words into two separate groups. Previous studies have shown different duration effects of surprisal between the two groups (Bell et al. 2009), and the N400 is most commonly associated with meaningful linguistic material, where function words may be considered to have less semantic weight. For this study, nouns, adjectives, verbs, and adverbs were included as lexical words, all other parts of speech were considered function words. The frequency of types and tokens in the material separated by function and lexical words, as well as for the four lexical parts of speech are shown in Table 1.

The duration of each token in the audio material had previously been measured for the studies by Di Liberto et al. (2015) and Broderick et al. (2018) and was included in the public data. To calculate AR as measured in syllables per second as well as information density as bits per syllable, all tokens' syllable counts were approximated using the Carnegie Mellon University Pronouncing Dictionary, CMUdict [13], through its Python module in the nltk package (Bird et al. 2009). The dictionary returns a word's separate phonemes, where phonemes that could potentially carry stress are marked with a digit. The number of digits returned was taken as a proxy for the number of syllables in a word. 43 tokens weren't included in the dictionary, for these syllable count was calculated as the number of vowels in a word except where vowels followed each other.

3.4 Statistical analysis

To investigate the stated research questions simple and multiple linear models and a mixed effects model were implemented. All statistical analysis was performed in the RStudio software (v1.3.1073, RStudio Team 2020). For the linear models, the built-in `lm` function was used. The `lme4` (Bates et al. 2015) package was used for the mixed effects models, with `lmerTest` (Kuznetsova et al. 2017) added in order to obtain p-values.

[12] Tagging was done by Robert Östling prior to work on this thesis started.
[13] http://svn.code.sf.net/p/cmusphinx/code/trunk/cmudict/

4 Results

4.1 AR and surprisal

Means, medians, standard deviations, and the range for AR and surprisal for all words as well as function and lexical words separated are presented in Table 2. The highest mean AR can be found with function words which also has the lowest mean surprisal. The lowest mean AR and highest mean surprisal can be found with lexical words. Figure 1a–d illustrates the distribution of AR and surprisal divided for function and lexical words, where function words have a bigger range of values for AR than for surprisal and the opposite relationship can be seen for lexical words.

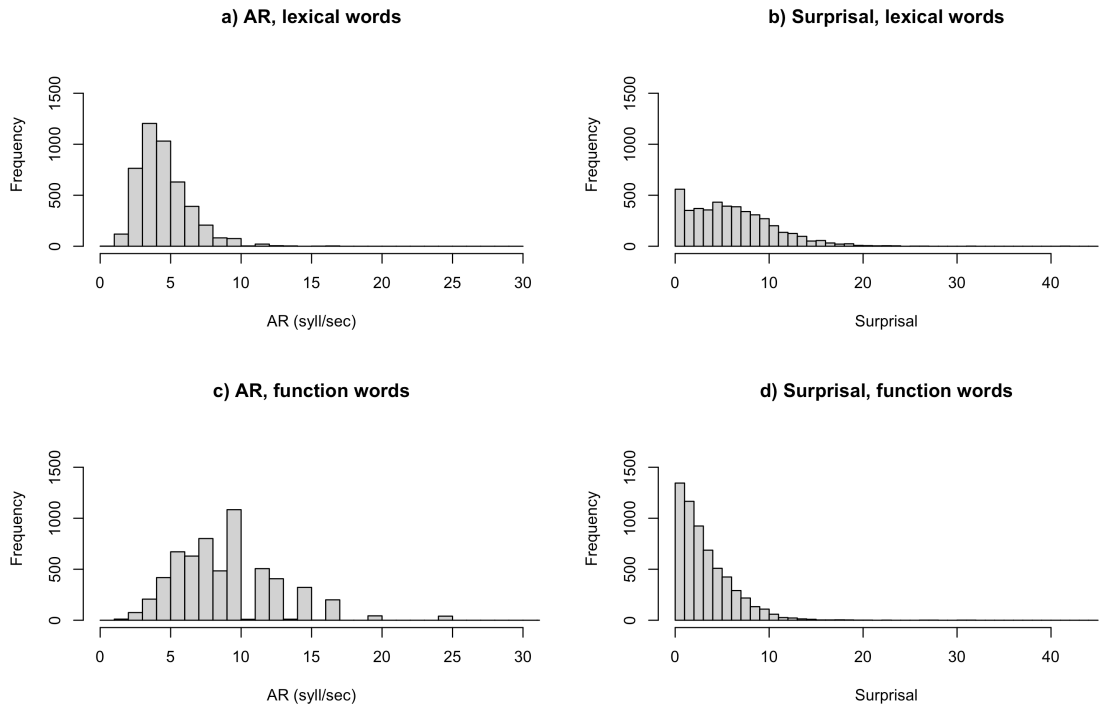

Figure 1: The range of values for AR and surprisal, divided by lexical and function words.

Table 2: Mean, median, standard deviation, and range of AR and surprisal in the data.

	AR				Surprisal			
Part of Speech	*M*	*Mdn*	*SD*	Range	*M*	*Mdn*	*SD*	Range
All words	7.0	6.3	3.9	[0.34–33.33]	4.4	3.5	3.8	[0.001–42]
Lexical words	4.5	4.2	1.8	[0.34–16.67]	6.0	5.5	4.3	[0.001–42]
Function words	8.9	8.3	4.1	[0.53–33.33]	3.2	2.5	2.8	[0.001–31.18]

4.1.1 Regression models

Four regression models (one of them a mixed effects model) were implemented in order to investigate the relationship between surprisal and AR. The first three models can be defined as:

$$AR = \alpha + \beta \cdot \textit{surprisal}$$

with the difference between them the constitution of the data set: all words; only function words; only lexical words.

When including all words, surprisal accounted for approximately 8.2 % of the variance, R^2 = .082, $F(1,10493) = 937$, $p < .001$, and showed an effect of surprisal on AR of –0.29 (95 % CI [-0.31, -0.27]). Thus, for an increase in surprisal of one standard deviation (3.83 bits), AR would decrease by 0.29 standard deviations (1.14 syllables per second).

Regression models separating function and lexical words showed that the effect seen when modeling over all words may stem from function words, as the coefficient for these were similar to the results when including all words, –0.28 (95 % CI [-0.32, -0.25]). In this model, roughly half of the variance as the model using all words was accounted for, R^2 = .04, $F(1,5953) = 246.1$, $p < .001$. The model with only lexical words had a very small, though positive, effect, 0.0035, but did not show to be significantly different from 0 (95 % CI [–0.008, 0.015]), $p = .57$.

The mixed effects model included only lexical words and used syllable count as random effect (intercept only). Table 3 shows the frequency of words with different number of syllables as well as their mean AR and information density in bits/syllable (surprisal/number of syllables). This is also illustrated in Figure 2, showing a trend for AR as the more syllables in a word, the higher the mean AR. For information density the inverse trend can be observed; fewer syllables lead to higher information density.

Words with five or six syllables were excluded from the model, due to such words only making up 7 and 2 tokens respectively in the data. The resulting effect of surprisal on AR when controlling for syllable count was –0.056 (95 % CI [–0.067, –0.044]) and was highly significant $(p < .001)$. This would mean that for every one bit increase in surprisal, a word's AR would decrease by approximately 0.06 syllables/second.

4.2 N400, AR and surprisal

To investigate the possible effects of surprisal and AR on the amplitude of the N400 both simple and multiple linear models were used. Only lexical words were used in these models and in

Table 3: Frequencies of the amount of syllables in lexical words, as well as mean, median, standard deviation, and range for AR and information density measured in bits/syllable.

		AR				Information density			
Syllables	Freq.	*M*	*Mdn*	*SD*	Range	*M*	*Mdn*	*SD*	Range
One	3271	4.0	3.7	1.6	[0.34–16.67]	5.3	4.8	3.8	[0.001–22.74]
Two	1022	5.6	5.3	1.7	[2.25–15.39]	3.8	3.5	2.4	[0.02–20.99]
Three	208	6.1	6.3	1.2	[2.75–12]	3.2	3.1	1.7	[0.03–8.76]
Four	30	6.3	6.1	1.2	[4.26–9.09]	3.2	3.2	1.2	[0.81–5.6]
Five	7	7.2	6.8	1.3	[5.56–9.26]	2.7	2.7	1.0	[1.06–4.46]
Six	2	7.1	7.1	0.4	[6.82–7.32]	2.0	2.0	0.8	[1.46–2.57]

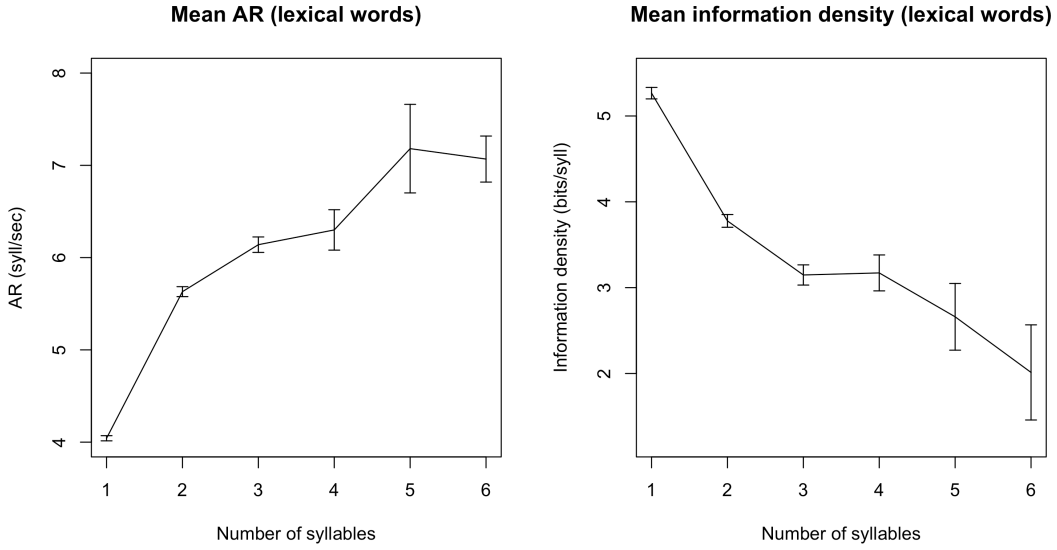

Figure 2: Means of AR and information density (bits/syllable) grouped by number of syllables in a word. Bars indicate standard error.

total there were 75,734 data points divided over 18 subjects ($4540\,tokens * 18\,participants - 5986\,rejected\,datapoints$).

4.2.1 Regression models

The estimated effect of surprisal in lexical words on the N400 was –0.038 (95 % CI [–0.055, –0.022]); more unexpected words were predicted to result in larger, more negative, N400 amplitudes, i.e., an increase in amplitude of 0.038 µV per increase of 1 bit [14]. Even with this observed effect, the model only accounted for approximately 0.03 % of the observed variation in the data, R^2 = .0003, $F(1,75732) = 20.97$, $p < .001$. When predicting the N400 only with AR, the effect was much smaller, –0.01 (95 % CI [–0.05, 0.03]), though it was not significant ($p = .64$).

The lack of effect from AR of the present token on the N400 could stem from that the evoked ERP response would arise from only a few 100 ms of exposure to that token, and such a window may be too short for AR to actually affect the N400. It was thus decided to explore if AR of preceding words could elicit effects, as this might be a more relevant measure in relation to processing load of the present token.

The N400 was modeled after the AR of the immediately preceding word as well as two words back, regardless if preceding words belonged to a lexical class or not. Both showed to be significant predictors of the N400: the immediately preceding word predicted an increase in N400 amplitude, –0.037 (95 % CI [–0.053, –0.02]), and two words back a decrease, 0.028 (95 % CI [0.01, 0.045]). The immediately preceding word was better to account for variation, R^2 = .00024, $F(1,75732) = 18.38$, $p < .001$, than two words back, R^2 = .00012, $F(1,75732) = 9.4$, $p = 0.002$.

The change of effect direction two words back as compared to one word back may possibly be related to an autocorrelation in the N400 amplitudes that was found during preprocessing (using Pz, a neighboring centro-parietal electrode to the Cz). This showed a negative autocorrelation of -.26 at lag 2, indicating a shifting polarity in the N400 every other word. An

[14] Since the N400 is a *negative* component, a negative regression coefficient entails an increase in its amplitude.

Table 4: Results from multiple linear regression models, using AR and information density in bits per second as predictors for the amplitude of the N400.

Model	Predictor	*B*	95% CI
1. One word back	AR[a]	–0.033***	[–0.05, –0.015]
	Information density[b]	–0.002	[–0.006, 0.002]
	R^2 (adjusted)	0.00023	
	F	9.683	
2. Two words back	AR[a]	0.022*	[0.0024, 0.04]
	Information density[b]	0.003	[–0.0008, 0.007]
	R^2 (adjusted)	0.00013	
	F	5.893	

Note. Both models $N = 75{,}734$. CI = confidence interval.
[a] = syllables/second. [b] = bits/second.
*$p < .05$. ***$p < .001$.

autocorrelation for surprisal, also at lag 2, was instead weakly positive, .04, indicating that the amplitude of the N400 and surprisal may have independent patterns.

Information density defined as bits/second was also used as a predictor, combining both surprisal and a measure of the rate at which it is produced. Modeling these over previous words as predicting the N400 showed significant effects: information density for the previous word predicted increased amplitudes, –0.004 (95 % CI [–0.008, –0.001]) while two words back predicted them to decrease, 0.005 (95 % CI [0.001, 0.008]). The models accounted for similar amounts of variation: for information density one word back, $R^2 = 8.15 \times 10^{-5}$, $F(1,75732) = 6.169$, $p = .013$, and two words back, $R^2 = 9.14 \times 10^{-5}$, $F(1,75732) = 6.924$, $p = .009$.

Multiple linear models were used to predict N400 with AR and information density (bits/second) for previous words included simultaneously. In the model using AR and information density of the immediately preceding word (Table 4, Model 1), AR was still highly significant in predicting an increased amplitude of 0.033 μV for every one increase in AR, while information density did not reach significance. Modeling the same parameters two words back (Model 2), AR remained significant with a decrease in amplitude of 0.022 μV per every one increase in AR; similarly to the first model information density did not reach significance.

5 Discussion

5.1 AR and surprisal

As was shown in Table 1 in section 3.3, the frequency distribution of function words and lexical words were slightly skewed in the material. Thus, the fact that the results of the regression model with all words as well as the model only using function words were significant is not entirely surprising. Not only were function words in a slight majority in the data, their distribution in regards to articulation rate is much broader than for lexical words while they have a smaller range for surprisal.

That the number of syllables, or segments, in a word affects its overall AR as shown by Lindblom and Rapp (1973) and Dankovičová (1999) could clearly be observed in the present data. The tendency of such word lengthening effects was large enough for any effects of surprisal on AR to become invisible when not controlling for syllable count in lexical words. Also, this probably accounts for the strength of the effect when including function words, since they are usually shorter in length and thus have fewer syllables in general. By observing the averages of AR and information density (Table 3 and Figure 2, only lexical words), this may hint at a relationship between AR and information density that points towards a uniform distribution of information over time. Words with fewer syllables will have their duration prolonged, while at the same time these same syllables will have higher information density. Though the figures may seem to show that this relation holds when the syllable count is increased, they are not reliable as there are few words with a high number of syllables.

The effect seen when controlling for syllable count was small, with only a change of approximately 0.06 syllable/second per one change in bit. Even so, this would still follow predictions from information theoretic accounts, such as UID. AR could thus be another level affected by surprisal and something that speakers may strategically modulate to achieve more uniform information density. The fact that the effect is of such a small size does not have to be entirely unexpected either, since previous work has shown that relative frequency (Zipf 1935) and surprisal-like measures (Piantadosi et al. 2011) already have an effect on the length of words. The inherent length or duration of a given word may already account for much of how information density spreads over that word; AR may instead be an even more locally situated tool to slightly alter information density, if the context makes need for it.

In order to put the size of the effect in perspective, the JND (Just Noticeble Difference) of articulation rate on sentence level, according to Eefting and Rietveld (1989), was a change of 4.43 %. If applying this to words in the present data, the average AR for a one syllable word was 4 syllables/second, thus a difference of 4.43 % entails that a word would have to change by approximately ±0.18 syllables/second to be noticeably different. According to the model, a change of approximately three bits would be needed for a change in AR to be noticeable by a listener ($0.06 * 3 = 0.18$).

When Jaeger (2010) describes that a speaker may chose between different, fairly synonymous, versions of an utterance in order to achieve the most optimal information density, this would not require an active, conscious, decision on the part of the speaker. Inversely, if comprehension and integration of meaning for a listener is affected by information density, how the density is managed would probably not be required to be all that apparent. And, whether or not a facilitation of AR due to surprisal must be noticeable to be useful, the effect of changing surprisal by one bit indicates a doubling (or half) the probability of an outcome, which may not be all that different in terms of expectedness. Rather, a change in 10 bits, a much bigger shift in probability ($p/2^{10}$ or $p * 2^{10}$) would according to the model imply a change of ±0.6 sylla-

bles/second for an average one syllable word, clearly beyond the threshold of JND for AR as proposed by Eefting and Rietveld (1989).

5.2 N400 as predicted by surprisal and AR

In line with previous research (Frank, Otten, et al. 2015; Yan and Jaeger 2020), surprisal could predict some of the amplitude of the N400. In comparison, the effect shown in Frank, Otten, et al. (2015) was 0.17–0.22 µV/bit (depending on language model), whereas the observed effect in this study was 0.038 µV/bit. This difference of effect size may have multiple reasons, among them that Frank, Otten, et al. (2015) used a visual paradigm (as did Yan and Jaeger 2020) rather than an auditive one as well as potential idiosyncrasies in the present EEG-data.

In regards to effects of AR (as well as information density) on the N400, the results are less conclusive. The fact that AR of the present word did not show significant effects on the N400 may not be completely surprising; the time window where the AR may affect the amplitude might be too small for it to be a significant causer of more cognitive load. The effect seen when modeling with AR of the previous word, in that a higher AR increases the N400's (negative) amplitude, could taken at face value mean that the rate of the previous word somehow induces more cognitive processing required of the next word. But, this may be an erroneous conclusion, as there may possibly be factors (e.g., syntactic) as to how lower and higher surprisal words appear in the text, which the effect of AR would rather be a reflection of. Additionally, the effect of previous words' AR does not seem to stem from a higher information density. When modeled separately, higher information density resulted in larger N400 amplitudes, but when including AR as well, information density did not reach significance. In summary, AR of the previous word does seem to have an effect on the N400; it is possible that this is not a result of changes in the distribution of information in that word but rather an independent effect.

That the N400 in itself showed a quite large negative autocorrelation two words away makes the interpretation of the results even more difficult, especially when surprisal doesn't seem to share similar autocorrelation, neither by size or direction. The effects seen when modeling the N400 with AR and information density from words two steps back are thus hard to evaluate. Whether or not the autocorrelation of the N400 is an effect of meaningful semantic processing that carries over several words or reflects some other non-related process can not be determined here. The former may be less likely since autocorrelation seems to be a general phenomenon (Guthrie and Buchwald 1991), and Yan and Jaeger (2020) utilized models that attempted to correct for autocorrelation in their ERP data.

5.3 Method and material discussion

5.3.1 Articulation Rate

The current study took into account two effects known to influence articulation rate on word level: syllable count and final lengthening (Dankovičová 1999; Lindblom and Rapp 1973). The inclusion of syllable count showed to be important in order to see what kind of effect surprisal had on AR and in return how AR may be utilized in order to distribute information in lexical words. While syllable count was explicitly controlled for, in the case of final lengthening sentence final words were simply omitted. It is of course possible that even inside of sentences, there are phrases which in themselves include final lengthening. When filtering the data, further steps could have been taken to automatically exclude more tokens, such as those followed by any other inter-punctuation, than the ones already defined as sentence final. A completely man-

ual approach would have been to listen to the recordings for what would be considered phrase endings, though this would probably have been more time consuming.

Furthermore, Dankovičová (1999) also showed that beyond a word being in final position or not, its position inside a phrase as a whole is also a potential factor for variation in AR that may not stem from surprisal. If word position was to be included though, this would potentially necessitate further parameters since the length of a whole utterance may have an effect on overall AR (Sjons and Hörberg 2016), so word position itself would need to stand in relation to the utterance as a whole.

5.3.2 Surprisal and language models

Though available, context windows of other sizes than 1,000 were not included when using surprisal as a predictor of either AR nor the N400, as this study was not intended as an evaluation of language models. It could of course be valuable in further work to utilize these different windows, to see if surprisal effects on AR (and the N400) would appear to be more strongly based on local or wider context. An argument for a bigger window could be made as suitable for the domain of the text material, that of a narrative story, where certain themes represented by specific words may come back but having larger distances than a sentence or two. A model with a larger window would be able to "remember" such themes in bigger spans. For example, one instance of the word *fish* had a range of surprisal of 0.54–14.19, depending on the size of the context window (where the lower value had a window of 1,000 tokens).

GPT-2 was trained on vast amounts of internet texts, with an intent to be applicable over many domains and genres. It could potentially be an issue that texts scraped from the internet may not be as applicable for narrative fiction texts. While state-of-the-art, it could be interesting to compare GPT-2 with models trained specifically on novels if using similar stimuli-material.

5.3.3 EEG and stimuli material

In regards to the EEG-data, there are some steps that could have been taken to reduce noise. First of all, there was no Independent Component Analysis applied, which would have eliminated artifacts in the signal. Secondly, there may be overlap effects that could have been accounted for, though this of course would have required more work in processing the data.

The material itself, taken from Broderick et al. (2018) and Di Liberto et al. (2015), consisted of readings of a narrative text. While this is closer to spontaneous speech than what may be used in more controlled experiment environments where unrelated words or sentences may be played in isolation (e.g., McCallum et al. 1984), it doesn't necessarily follow the same patterns as spontaneous speech proper. In order to further assess how AR may be used by speakers to uniformly distribute information and how this may affect listeners, it could be fruitful to do similar studies on further read material as well as actual spontaneous speech.

5.3.4 Further research

In further investigations in AR's role as distributor of information, it might be of interest to look at different types of read text material as well as styles of speech to see if AR is used similarly as has been observed in this study. Narrative texts, as the one used here, may differ in a number of ways from non-narrative texts, e.g., compositional choices and the existence of subtext; whether or not this leads to different speaker behavior in regards to AR would be interesting to explore.

Furthermore, another path of interest would be to similarly analyze actual spontaneous speech, i.e., dialogue; both since read and spontaneous speech have been shown to differ in average AR

(with spontaneous speech spoken at a higher rate, Jacewicz et al. 2009) as well as to further settle UID and similar accounts in more spontaneous communication. The fact that dialogue partners have the possibility to give feedback to one another, verbal and non-verbal, may also be informative in regards to how speakers adapt their speaking strategies in order to achieve more uniform distribution of information given direct feedback.

The results of this study regarding AR and resulting information density effects on the N400 were inconclusive, but potentially indicate that the N400 is not solely reflective of semantic processing load. In order to more closely investigate and parse out how other factors may influence the N400, more controlled experiments could be performed, where AR, surprisal and information density can be specifically targeted and manipulated. The use of more naturalistic stimuli, as has been done here, have given hints of possible effects; controlled experiments would allow to isolate these parameters for closer investigation. One concrete such experiment would be to adapt Wlotko and Federmeier's (2015) visual presentation rate experiment to auditive stimuli, where AR of words preceding a specific target word is manipulated in place of overall presentation rate. Additionally, in such controlled experiments it may be of value to expand the scope of ERPs investigated, not only the N400, but a further range of components, such as the P600, another language related ERP component.

6 Conclusions

1. Does AR function as a means of distributing information uniformly on the word level in continuous speech?

The results indicate that yes, AR is potentially a means to achieve more uniform information density of words within sentences, in line with UID accounts that speakers may achieve this uniformity through several linguistic levels (Jaeger 2006). However, according to a regression model the effects of surprisal on AR are small; possibly other factors, such as form and inherent duration of words, may already account for much of the distribution. AR could thus be interpreted as a tool to take care of some irregularities in the distribution arising from contextual needs. Further research could utilize different text genres and types of speech, such as actual spontaneous speech, in order to assess if AR is a consistent mean to distribute information, albeit possibly to different extents in different speaking situations.

2. Does variation in AR and the resulting distribution of information affect the amplitude of the N400?

Surprisal does, in line with previous studies (Frank, Otten, et al. 2015), seem to predict some of the N400 amplitude, where higher surprisal elicits a more negative response. However, AR of a word did not seem to give any significant effects on the N400. AR and information density of an immediately preceding word as well as two words back separately showed significant effects; in the former both higher AR and information density resulted in larger N400 amplitudes and the latter the inverse effect was observed. However, these effects appear to be independent, so the increased amplitude of the N400 due to higher AR in a previous word does not seem to stem from differences in the distribution of information. This could indicate that the N400 may be a reflection of more than semantic processing load. Further research could perform similar analyses in more controlled experimental settings to be able to separately manipulate parameters of AR, information density, and surprisal. Explorations in this direction could work towards a better understanding as to what factors affect the N400, as well as other neural responses.

References

Aurnhammer, Christoph and Stefan L. Frank (2019). Evaluating information-theoretic measures of word prediction in naturalistic sentence reading. In: *Neuropsychologia* 134, p. 107198. DOI: 10.1016/j.neuropsychologia.2019.107198.

Aylett, Matthew and Alice Turk (2004). The Smooth Signal Redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. In: *Language and Speech* 47, pp. 31–56. DOI: 10.1177/00238309040470010201.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker (2015). Fitting linear mixed-effects models using lme4. In: *Journal of Statistical Software* 67.1, pp. 1–48. DOI: 10.18637/jss.v067.i01.

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky (2009). Predictability effects on durations of content and function words in conversational English. In: *Journal of Memory and Language* 60.1, pp. 92–111. DOI: 10.1016/j.jml.2008.06.003.

Bengio, Yoshua, Réjean Ducharme, Pascal Vincent, and Christian Janvin (2003). A neural probabilistic language model. In: *The Journal of Machine Learning Research* 3, pp. 1137–1155. URL: https://dl.acm.org/doi/10.5555/944919.944966.

Bird, Steven, Ewan Klein, and Edward Loper (2009). *Natural language processing with Python*. 1st ed. Beijing: O'Reilly Media Inc.

Broderick, Michael, Andrew Anderson, Giovanni Di Liberto, Mick Crosse, and Edmund Lalor (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. In: *Current Biology* 28.5, pp. 803–809. DOI: 10.1016/j.cub.2018.01.080.

Dankovičová, Jana (1999). Articulation rate variation within the intonation phrase in Czech and English. In: *Proceedings of the XIVth International Congress of Phonetic Sciences* (San Francisco, CA, United States). Ed. by John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville, and Ashlee C. Bailey. International Phonetic Association, pp. 269–272. URL: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_0269.pdf.

Di Liberto, Giovanni M., James A. O'Sullivan, and Edmund C. Lalor (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. In: *Current Biology* 25.19, pp. 2457–2465. DOI: 10.1016/j.cub.2015.08.030.

Duncan, Connie C., Robert J. Barry, John F. Connolly, Catherine Fischer, Patricia T. Michie, Risto Näätänen, John Polich, Ivar Reinvang, and Cyma Van Petten (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. In: *Clinical Neurophysiology* 120.11, pp. 1883–1908. DOI: 10.1016/j.clinph.2009.07.045.

Eefting, Wieke and A.C.M. Rietveld (1989). Just noticeable differences of articulation rate at sentence level. In: *Speech Communication* 8.4, pp. 355–361. DOI: 10.1016/0167-6393(89)90017-4.

Fenk-Oczlon, Gertraud (2001). Familiarity, information flow, and linguistic form. In: *Frequency and the emergence of linguistic structure* 45, pp. 431–448. DOI: 10.1075/tsl.45.22fen.

Frank, Austin F. and T. Florian Jaeger (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In: *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (Washington, DC, United States). Ed. by B. C. Love, K. McRae, and V. M. Sloutsky. Cognitive Science Society, pp. 939–944. URL: https://escholarship.org/content/qt7d08h6j4/qt7d08h6j4.pdf.

Frank, Stefan, Leun Otten, Giulia Galli, and Gabriella Vigliocco (2015). The ERP response to the amount of information conveyed by words in sentences. In: *Brain and Language* 140, pp. 1–11. DOI: 10.1016/j.bandl.2014.10.006.

Gahl, Susanne (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. In: *Language* 84.3, pp. 474–496. DOI: 10.1353/lan.0.0035.

Gramfort, Alexandre, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen (2013). MEG and EEG data analysis with MNE-Python. In: *Frontiers in Neuroscience* 7.267, pp. 1–13. DOI: 10.3389/fnins.2013.00267.

Guthrie, Donald and Jennifer S Buchwald (1991). Significance testing of difference potentials. In: *Psychophysiology* 28.2, pp. 240–244. DOI: 10.1111/j.1469-8986.1991.tb00417.x.

Hagoort, Peter, Lea Hald, Marcel Bastiaansen, and Karl Magnus Petersson (2004). Integration of word meaning and world knowledge in language comprehension. In: *Science* 304.5669, pp. 438–441. DOI: 10.1126/science.1095455.

Hale, John (2016). Information-theoretical complexity metrics. In: *Language and Linguistics Compass* 10.9, pp. 397–412. DOI: 10.1111/lnc3.12196.

Jacewicz, Ewa, Robert A. Fox, Caitlin O'Neill, and Joseph Salmons (2009). Articulation rate across dialect, age, and gender. In: *Language Variation and Change* 21.2, pp. 233–256. DOI: 10.1017/S0954394509990093.

Jaeger, T. Florian (2010). Redundancy and reduction: Speakers manage syntactic information density. In: *Cognitive Psychology* 61.1, pp. 23–62. DOI: 10.1016/j.cogpsych.2010.02.002.

Jaeger, T. Florian (2006). Redundancy and syntactic reduction in spontaneous speech. PhD thesis. Stanford University.

Kemmerer, David L. (2015). *Cognitive neuroscience of language*. New York, NY: Psychology Press.

Kutas, Marta and Kara D. Federmeier (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). In: *Annual Review of Psychology* 62.1, pp. 621–647. DOI: 10.1146/annurev.psych.093008.131123.

Kutas, Marta and Steven A. Hillyard (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. In: *Science* 207.4427, pp. 203–205. DOI: 10.1126/science.7350657.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen (2017). lmerTest package: Tests in linear mixed effects models. In: *Journal of Statistical Software* 82.13, pp. 1–26. DOI: 10.18637/jss.v082.i13.

Lindblom, Björn EF and Karin Rapp (1973). Some temporal regularities of spoken Swedish. In: *Papers from the Instititue of Linguistics at the University of Stockholm* 21, pp. 1–59.

Luck, Steven J. (2014). *An introduction to the event-related potential technique*. 2nd edition. Cambridge, Massachusetts: The MIT Press.

Mahowald, Kyle, Evelina Fedorenko, Steven T Piantadosi, and Edward Gibson (2013). Info/information theory: Speakers choose shorter words in predictive contexts. In: *Cognition* 126.2, pp. 313–318. DOI: 10.1016/j.cognition.2012.09.010.

McCallum, W.C, S.F Farmer, and P.V Pocock (1984). The effects of physical and semantic incongruites on auditory event-related potentials. In: *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 59.6, pp. 477–488. DOI: doi.org/10.1016/0168-5597(84)90006-6.

Pellegrino, François, Christophe Coupé, and Egidio Marsico (2011). A cross-language perspective on speech information rate. In: *Language* 87.3, pp. 539–558. DOI: 10.2307/23011654.

Piantadosi, Steven, Harry Tily, and Edward Gibson (2011). Word lengths are optimized for efficient communication. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.9, pp. 3526–3529. DOI: 10.1073/pnas.1012551108.

Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning (2020). Stanza: A Python natural language processing toolkit for many human languages. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (Online). Ed. by Asli Celikyilmaz and Tsung-Hsien Wen. Association for Computational Linguistics. DOI: 10.18653/v1/2020.acl-demos.14.

Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever (2019). Language models are unsupervised multitask learners. In: *OpenAI blog* 1.8. URL: http://www.persagen.com/files/misc/radford2019language.pdf.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA. URL: http://www.rstudio.com/.

Sennrich, Rico, Barry Haddow, and Alexandra Birch (2016). Neural machine translation of rare words with subword units. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Berlin, Germany). Ed. by Katrin Erk and Noah A. Smith. Association for Computational Linguistics, pp. 1715–1725. DOI: 10.18653/v1/P16-1162.

Shannon, Claude E. (1948). A mathematical theory of communication. In: *Bell Systems Technical Journal* 27.3, pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Sjons, Johan and Thomas Hörberg (2016). Articulation rate in child-directed speech increases as a function of child age. In: *Fonetik 2016* (Stockholm, Sweden). URL: https://www.diva-portal.org/smash/get/diva2:945646/FULLTEXT01.pdf.

Sjons, Johan, Thomas Hörberg, Johannes Bjerva, and Robert Östling (2017). Articulation rate in Swedish child-directed speech increases as a function of the age of the child even when surprisal is controlled for. In: *Proceedings of Interspeech 2017* (Stockholm, Sweden). Ed. by Francisco Lacerda, David House, Mattias Heldner, Joakim Gustafsson, Sofia Strömbergsson, and Marcin Włodarczak. The International Speech Communication Association (ISCA), pp. 1794–1798. DOI: 10.21437/Interspeech.2017-1052.

Smith, Nathaniel J and Roger Levy (2013). The effect of word predictability on reading time is logarithmic. In: *Cognition* 128.3, pp. 302–319. DOI: 10.1016/j.cognition.2013.02.013.

Sundermeyer, Martin, Ralf Schlüter, and Hermann Ney (2012). LSTM neural networks for language modeling. In: *Interspeech 2012* (Portland, OR, United States). The International Speech Communication Association (ISCA), pp. 194–197. URL: https://www.isca-speech.org/archive/archive_papers/interspeech_2012/i12_0194.pdf.

Tueting, P. (1978). Event-related potentials, cognitive events and information processing: A summary of issues and discussion. In: *Multidisciplinary perspectives in event-related brain potential research*, pp. 159–169.

Van Petten, Cyma and Barbara Luka (2012). Prediction during language comprehension: Benefits, costs, and ERP components. In: *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology* 83.2, pp. 176–190. DOI: 10.1016/j.ijpsycho.2011.09.015.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin (2017). *Attention Is All You Need*. arXiv: 1706.03762 [cs.CL].

Wlotko, Edward W. and Kara D. Federmeier (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. In: *Cortex* 68, pp. 20–32. DOI: 10.1016/j.cortex.2015.03.014.

Yan, Shaorong and T. Florian Jaeger (2020). (Early) context effects on event-related potentials over natural inputs. In: *Language, cognition and neuroscience* 35.5, pp. 658–679. DOI: 10.1080/23273798.2019.1597979.

Zipf, George K. (1935). *The psycho-biology of language*. Mifflin: Houghton.

Stockholm University
SE-106 91 Stockholm, Sweden
Telephone +46 (0)8 16 20 00
https://www.su.se/