Saarland University
Department of Language Science and Technology
Faculty of Humanities

Master Thesis

The Predictive Power of Language Model Surprisal for N400 & P600 Effects

Benedict Schneider

23. November 2023

Supervisors:
Prof. Dr. Matthew W. Crocker
Dr. Francesca Delogu

Advisors:
Christoph Aurnhammer
Dr. Harm Brouwer

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Ich versichere, dass die gedruckte und die elektronische Version der Masterarbeit inhaltlich übereinstimmen.

Statutory Declaration

I hereby declare that the thesis presented here is my own work and that no other sources or aids, other than those listed, have been used. I affirm that the electronic version is identical in content to the printed version of the Master's thesis.

Ort, Datum / *Place, date*: Saarbrücken, 23.11.2023

Unterschrift / *Signature*:

Acknowledgments

I would like to thank the members of the A1 research project of the IDeal SFB 1102, that is, Christoph Aurnhammer, Harm Brouwer, Matthew Crocker and Francesca Delogu for their excellent guidance, professional expertise and infinite patience throughout the course of this project. Whether it was through our regular group meetings or the conversations I had with each individual member, I always felt enriched and encouraged afterwards to continue the work on the thesis with a new perspective. I feel honored to be part of the team and hope that I can make a meaningful contribution with this thesis and with further work in the future.

Especially, I would like to express my gratitude towards Christoph Aurnhammer, for not only sharing his rERP-code, which I could adapt for my own analyses, but also for his endless patience and invaluable advice in the numerous conversations we had about basically all matters relevant to writing a thesis.

Moreover, I would like to thank Julius Steuer and Marius Mosbach as members of the SFB's B4 project, for granting me access to their Language Model Toolkit as well as their computational resources, which made it possible for me to train the LSTM that was used in this work. In addition, Julius Steuer provided helpful technical tips for the training procedure and introduced me to a more efficient method of collecting surprisal estimates, for which I am thankful. In this respect, I would also like to thank Stefan Schweter for granting me access to the GPT-2 training data which was fundamental for the language model training.

Additionally I would like to thank all of my colleagues for creating such a pleasant working atmosphere and especially Heiner Drenhaus and Torsten Jachmann for patiently listening to my occasional questions and considerations about statistical matters. I would also like to mention two dear friends, Alina Leippert and Siyu Tao, who accompanied me throughout the master's program and always kept an open ear for my personal concerns.

Finally, I would like to thank my entire family, also Doris and Patrick Laub, and especially my fiancée Vanessa Krieger for their loving support and always being concerned about my well-being.

Abstract

Expectancies about upcoming words in online language comprehension are driven not only by syntactic and semantic constraints, but also other factors such as pragmatic reasoning and world knowledge. Words that are less expected may induce a higher cognitive processing effort when encountered in the input language stream. *Surprisal* is an information-theoretic linking-theory that has been introduced to formalize this notion on a computational level. Its operationalization through probabilistic neural language models (*LM surprisal*) has been widely used in psycholinguistic research of recent years to explain empirically observed measures such as reading times, eye movements and ERPs.

The present work examines the surprisal estimates of two language models of different architecture (RNN and transformer-based) with regard to their ability to predict N400 and P600 effects found by four recent ERP-studies (Aurnhammer et al., 2021, 2023; Delogu et al., 2019, 2021). In addition to expectancy, the authors manipulated the distinct but confounded property of semantic association, defined as the extent to which a word's meaning is primed by its prior context. The studies have shown the N400 to be sensitive to manipulations of both semantic association and expectancy, while the P600 was shown to be sensitive to expectancy but *not* semantic association. Given that the theoretical definition of surprisal reflects expectancy, it was anticipated that LM surprisal would capture the expectancy-driven N400 and P600 effects, but not the association-driven N400 effects.

However, the results of an rERP-analysis and the comparison with human-based judgments of association, expectancy and plausibility indicate that this was not the case. Instead, LM surprisal overall showed a stronger pattern with semantic association and was better able to capture the association-driven N400 effects, while only the transformer model captured the expectancy-driven N400 and P600 effects to a certain extent. Hence, the results suggest that it is necessary to carefully re-evaluate the interpretation of surprisal estimates derived from contemporary language models as reflecting expectancy in human language comprehension, since it appears that they actually may reflect contextual semantic similarity to a greater extent than expectancy.

Contents

Chapter 1.

Introduction

Within the field of computational psycholinguistics, increasingly capable language models have been deployed to predict the amplitude of the N400, an ERP-component that has amongst other factors been shown to be sensitive to word expectancy (Kutas & Federmeier, 2011). While successful in this endeavor, expectancy itself is a very general concept, which may be influenced by a number of constituting factors. Some of these factors may be observable from a purely statistical view on language, such as word frequencies and contextual co-occurrences, while others may be encoded on a structural level or follow more implicitly from pragmatic principles of communication and profound knowledge about the world. Consider the following example sentence with four possible completions:

(1) Anna ordered a pizza with ...
   a. pepperoni.
   b. artichokes.
   c. dough.
   d. bicycles.

Without taking any additional constraining context into account, **a** may intuitively offer the most expected completion out of the four, with the other completions being gradually less expected relative to **a**, crucially though for different reasons. Although **b** is just as plausible as **a**, *artichokes* may co-occur way less frequently alongside the context noun *pizza* compared to *pepperoni*, which may lead to a weaker semantic association between *artichokes* and *pizza* compared to *pepperoni* and *pizza* within a listener. On the other hand, *dough* and *pizza* are presumably strongly associated with each other, but taking extra-linguistic knowledge into account, it seems implausible to top a pizza with dough, rendering **c** more unexpected than both **a** and **b**. Finally, **d** is drastically more implausible than the other continuations by violating the contextual constraint of the argument having the property of being edible. Moreover, *bicycles* presumably is far less associated with *pizza* than the other completions.

Hence, how well a specific operationalization of expectancy can capture specific ERP-effects should depend on how sensitive this operationalization internally reacts to manipulations of the constituting factors that led to the occurrence of the effects. Moreover, if the operationalization is very sensitive to one of the factors, while at the same time being insensitive to another, it may fail in predicting human behavior in certain scenarios.

Two important findings from recent studies indicate, that the N400 is sensitive to both manipulations of expectancy and semantic association, while the P600, another prominent ERP-

component, is sensitive to manipulations of expectancy but *not* association (Aurnhammer et al., 2021, 2023; Delogu et al., 2019, 2021).

While a number of studies predicted effects in the N400 window by the means of neural language models, the P600 has not been well researched in this regard yet. Thus, the thesis aims to evaluate on the four above-mentioned ERP-studies not only to which extent language model surprisal can predict N400 effects that have been elicited by manipulations of association, but also if is able to capture the aforementioned P600 effects as well.

The following section presents the theoretical concepts of expectancy and surprisal, experimental and model-derived operationalizations of these concepts as well as their application in neurophysiological research.

1.1. Surprisal Theory and other Expectancy-related Notions

Human utterance comprehension is driven by incremental expectations about upcoming words. In order to formalize this notion on a computational level (Marr, 2010), the concept of *surprisal* (1.1), originating from information theory (Shannon, 1948), has been introduced. According to surprisal theory, the cognitive effort required to process a word w in a sentence at position $t+1$ is proportional to its surprisal, which is defined as its negative logarithmic probability given the preceding context words ranging from w_1 until w_t (Hale, 2001; Levy, 2008).

$$Surprisal(w_{t+1}) = -log_2 p(w_{t+1}|w_{1...t}) \tag{1.1}$$

Importantly, this formalization remains agnostic towards the exact origins and mechanisms that can lead to processing difficulty, providing only a general measure of expectancy. Thus, due to its nature as a linking hypothesis, surprisal by itself does not rely on assumptions about the algorithmic or implementational level (Marr, 2010).

While the amount of information a specific word conveys in a specific context can be *computed* by probabilistic models, the amount of cognitive effort required to process this word may be *observed* through (neuro)behavioural methods (paraphrasing Frank et al., 2015). Specifically, a positive correlation between reading times and word surprisal has been established in the past (Fernandez Monsalve et al., 2012; Fossum & Levy, 2012; Mitchell et al., 2010; Roark et al., 2009; Smith & Levy, 2008). While reading times provide an overall index of word-by-word processing effort, neurophysiological methods may be able to reveal the mechanisms that are underlying and driving this effort.

Event-related potentials (ERPs) provide a multidimensional window into language comprehension with an excellent temporal resolution and have been widely used to study neural activity during sentence comprehension. The N400 and P600 components have been identified as indices of processing effort during language comprehension and, importantly, they have been shown to be differentially sensitive to experimental manipulations of *plausibility* and *semantic association*, which are themselves related to the general concept of expectancy.

Although the precise nature of the underlying mental representations of semantic memory is subject to ongoing research, one common theory is to conceptualize it in terms of a network (Hutchison, 2003). Under a holistic approach, in such a network meaning representations or concepts can be thought of as nodes that are connected, or *associated* to each other, not only

directly but also indirectly via shared properties. Encountering a word then activates not only the semantic meaning of that word itself, but activation is thought to spread out to nodes that are associated, pre-activating them. Therefore, the concept of the *semantic association* between a word and a context can be seen as indicating the extent to which the meaning of the word is primed by this context (Rabs et al., 2022). *Strength* of association may then be reflected in distributional properties of language, that is, how often words or groups of words tend to co-occur. Naturally there exists a relationship to the notion of expectancy, in the sense, that words that are strongly associated with each other are also more likely to complete the same contexts (Ettinger et al., 2016). Importantly though, the concept of semantic association does not include any assumption about syntactic well-formedness or contextual plausibility. Thus, the association of a given continuation to the preceding context can but does not necessarily entail expectedness, i.e. a strongly associated continuation may still be an unexpected one (see example 1 c).

Extending beyond linguistic information, the concept of *plausibility* reflects to which extent an utterance aligns with our knowledge of the world (Haeuser & Kray, 2022). This definition of course is very general in nature, and there are a multitude of ways in how plausibility can be diminished, specifically in experimental settings. For example, the plausibility of a certain sentence continuation can be reduced on a discourse level by an atypical sequential ordering of events (Delogu et al., 2019, 2021), or by violating the selectional restrictions of a preceding verb (Aurnhammer et al., 2021, 2023; Haeuser & Kray, 2022). Detecting such violations requires the listener to be able to reason about the state of the world and compositional meaning, that is, they can't be explicitly observed on the surface level of language. The relation of plausibility to expectancy may be to some extent clearer compared to association. Generally, under rational theories of communication (Grice, 1967), we may expect implausible continuations to be more unlikely than plausible ones. Crucially though, the relationship is not symmetric, that is, rather unexpected continuations may still be perfectly plausible (see example 1, b)

For all of these concepts, that is, semantic association, plausibility and expectancy, there exist a number of possible operationlizations which may be either derived from (language) models or based on human judgements. Typically, these operationalizations capture only certain aspects of these highly complex underlying notions, and thus all of them may have their own strengths and weaknesses in capturing certain phenomena. Although well known in the areas of psycholinguistics and computational linguistics, the ones that are highly relevant for the subject of this thesis will be introduced in the following subsections, divided into human-based and model-derived operationalizations.

1.1.1. Human-based Operationalizations

Cloze probability (oftentimes abbreviated as *cloze*) has been originally introduced by Taylor (1953) as result of a normative sentence-completion study. According to this study, the cloze probability of a word equals the percentage of people who continued a sentence fragment with it. As such, cloze can be viewed as a measure of word probability gathered from human judgements rather than from probabilistic models (Frank et al., 2015). While providing a good estimate about the more likely range of completions that humans may expect in a given context, one known weakness is a lack of differentiation in the lower probability spectrum of possible

continuations. That is, reasonable but less likely continuations may very often result in a zero cloze probability in the same way as impossible continuations. Moreover, it has to be noted that cloze probability can also be logarithmically transformed to cloze surprisal, but this requires a smoothing technique, such as Laplace smoothing, since zero probabilities would naturally result in infinite surprisal (De Varda et al., 2023). A further method to operationalize expectancy is to collect predictability ratings, that is to ask participants to rate the expectedness of a target word given the preceding context on a Likert-scale (De Varda et al., 2023). However, the thesis will focus on cloze probabilities, as those have been collected in the studies that are evaluated here.

Analogous to expectancy, semantic association is usually operationalized either by a production or a rating task. In norming studies, it may be estimated by participants rating the strength of relatedness between a target and usually a content word (or group of words) from the preceding context (Aurnhammer et al., 2021; Delogu et al., 2019, 2021) or by participants producing related words under time constraints (Battig & Montague, 1969; Keppel & Postman, 1970; Kutas, 1993). The former variant will be relevant in the studies evaluated in this thesis.

Finally, plausibility is usually operationalized by a rating task (rather than a production task). That is, participants are presented with an experimental stimulus in its entirety, or just until including the target word and asked to rate it on a Likert-scale with respect to how plausible it is (Aurnhammer et al., 2023; Delogu et al., 2019, 2021; Michaelov et al., 2023).

A challenging problem arises from the correlation between these metrics and their relation to the more generalistic notion of expectancy. While cloze probability can be viewed as a human-based operationalization of expectancy (that stands conceptually near to surprisal), the concept itself is confounded with both semantic association and plausibility, as described above. Hence, when comparing metrics gained from operationalization of these notions (association, plausiblity, expectancy) it is important to note that these are usually correlated.

1.1.2. Model-derived Operationalizations

The objective of a statistical language model is to generate a probability distribution over the next possible words known to it given a context of preceding words. Since surprisal can be formalized in terms of logarithmic word probabilities, language models have been extensively deployed to estimate these probabilites[1] in the past (see Table 1.1).

Importantly, the surprisal estimates from the studies listed in Table 1.1 stem from a range of different model-architectures. Originally, Hale (2001) used a *probabilistic Early parser*, relying on a *phrase structure grammar (PSG)*. But, while having the advantage of capturing hierarchical structure in language, the limited scalability of grammar-based approaches has led to research turning to architectures that rely solely on distributional information, which is easily available on a large scale in form of corpora.

As such type of model, *n-grams* are contextually restricted to the *n-1* preceding words when estimating word probability and respectively surprisal. While these estimates rely straightforwardly on collocation frequencies gathered from corpora, due to data sparseness, the order of *n*

[1]Note that the unit in question does not need to be restricted to words. For example, surprisal can and has also been computed for POS-tags (Frank et al., 2015) and CCG-tags (Arehalli et al., 2022) to operationalize a notion of syntactic versus lexical surprisal.

Grammar-based	Frank et al. (2015) Parviz et al. (2012) Mitchell et al. (2010) Hale (2001)	
n-gram	Degaetano-Ortlieb and Teich (2022) Goodkind and Bicknell (2018) Frank (2017) Frank and Willems (2017) Frank et al. (2015)	Mitchell et al. (2010)
RNN	Slaats and Martin (2023) Michaelov et al. (2022) Merkx and Frank (2021) Michaelov et al. (2021) Michaelov and Bergen (2020)	Aurnhammer and Frank (2019a) Frank and Hoeks (2019) Aurnhammer and Frank (2019b) Goodkind and Bicknell (2018) Frank et al. (2015)
Transformer	Michaelov et al. (2023) Michaelov et al. (2022) Oh and Schuler (2022) Merkx and Frank (2021) Michaelov et al. (2021)	

Table 1.1.: A selection of studies that collected surprisal estimates from different language model architectures (sorted by model type and publication date).

is usually ceiling at 5, even after including smoothing techniques, meaning that only the 4 preceding words are taken into account for the estimate (Jurafsky & Martin, n.d.). Choosing an n greater than this critically exacerbates the sparse data problem n-gram models inherently face, which is, that the number of unseen word combinations massively grows. Though their limited context size renders n-gram models cognitively implausible, they have shown to be remarkably accurate in certain cases (Frank et al., 2015). Due to the design of the evaluated studies though, n-gram models won't be considered in this thesis, as will be explained in chapter 2.

As computing power has increased and the field of machine learning has advanced rapidly, *neural networks* have become widely used for language modeling. A crucial difference to n-gram models lies in the requirement of neural networks for a numerical representation of the input language units[2]. Since these representations may play a crucial role in shaping the underlying probability distribution that neural language model generate and are also commonly used in operationalizing semantic relatedness, they will be introduced here. Although a straightforward localist word representation, such as a one-hot encoding over the vocabulary of the model may be feasible to implement, it comes with the disadvantage of a complete lack of comparability between words. Therefore, more sophisticated approaches leading to high-dimensional vector based representations, also referred to as *embeddings*, have been developed.

The most well known methods can be divided into two distinctive approaches, based on either *counts* or *predictions* (Mandera et al., 2017). For the former, two prominent examples are the *Latent Semantic Analysis* (LSA; Landauer and Dumais, 1997) and *Global Vectors for Word Representation* (GloVe; Pennington et al., 2014), whereas the latter are most prominently rep-

[2] Often but not necessarily words.

resented by *Word2Vec* (Mikolov et al., 2013) and its more recent instantiation of *fastText* (Bojanowski et al., 2017). In favor of prediction based methods, arguments have been made for their stronger psychological plausibility (Mandera et al., 2017) and overall better performance (Baroni et al., 2014; Nieuwland et al., 2020) compared to count based methods.

To serve as input for a neural network, these representations can either be externally pretrained, for example with one of the methods described above, or be learned jointly during the neural network's own training process. Importantly, they are static, global representations that are semantically comparable but don't capture polysemy (Ethayarajh, 2019).

The words of an input sequence are preprocessed by a tokenizer and mapped to their respective embeddings and passed as an input to the network. Hence, in neural language models the dimension of the input layer usually follows from the pre-determined embedding dimensionality. Importantly though, in order to arrive at a more parsimonious, memory effective organization of the model internal vocabulary, subword-tokenization algorithms such as *Byte-Pair-Encoding* (BPE ; Sennrich et al., 2016) are widely applied. This leads to the model internal language units, and hence the input units, not being words but subwords, derived from splits that are not necessarily linguistically motivated. The potentially problematic implications of such a representation will be discussed in chapter 4.

The embedding values from the input nodes are mapped through one or multiple hidden layers of pre-determined size, applying a function of nonlinearity on each layer, onto an output layer. For the goal of language modeling, the network is trained on the task of next word prediction, given a sequence of preceding words. Thus, the output layer dimension equals the size of the vocabulary and holds unnormalized scores, also referred to as logits, which reflect the models "raw" predictions for the next word. To make these scores interpretable, commonly a softmax function is applied to the output layer, resulting in a probability distribution over next words.

The training procedure is carried out by applying an algorithm such as backpropagation (Rumelhart et al., 1986) that leads model to reduce its error on the training data by successively adjusting its weights according to a gradient.

In terms of context size and cognitive plausibility, the *Recurrent Neural Network* (RNN; Elman, 1990) offers a considerable improvement over n-gram models. As a subtype of this model class, the *Simple Recurrent Network* (SRN; Rumelhart et al., 1986) has been succesfully used within psycholinguistic research to model aspects of incremental language comprehension (see Frank et al., 2015 as an example).

Recurrent Neural Networks are able to handle input sequences of variable length, making them capable to operate on natural language. This is achieved by using a recurrence mechanism, that allows to record and pass on contextual information in hidden states across time steps, computing the current hidden state activations by combining the current input with the previous hidden state's output. Crucially, the weights and biases are shared between time steps, removing the constraint of pre-specifying the input sequence length the model can encounter. For the training procedure, a variant of the backpropagation algorithm, *backpropagation through time* is required (Elman, 1990;Werbos, 1988).

Although this in theory allows training and evaluation on arbitrarily long sequences, it was found that the ability of RNNs to store and access long-range information is limited (Hochreiter, 1991; Bengio et al., 1994), due to the *exploding/vanishing gradient problem* (Hochreiter, 1991;

Hochreiter, 1998). This problem increasingly occurs for longer sequences and deeper networks, when the network's weights are being updated and the repeated application of the chain rule during backpropagation leads to multiplying gradients that are very large or small. The former can manifest itself in an oscillating loss due to overstepping minima, while the latter can result in a slow or even halting training process.

As a solution to this problem Hochreiter and Schmidhuber (1997) introduced the *Long-Short-Term-Memory* (LSTM) network[3]. In this variant of the RNN, the hidden unit, comprised of a combination of the previous hidden state, the current input and a nonlinear activation function, is replaced by a more complex unit, with the idea to use two pathways to distinctively model long-term and short-term information flow. Long-term information is passed along in the so-called *cell state*, that is being updated without having trainable weights and biases, therefore making it unresponsive to gradients. Short-term information is stored in the *hidden state* that is connected to weights involved in the unit's gating mechanism. The *gates* control which information is to which extent kept or forgotten. There are three gates: the *forget gate*, the *input gate* and the *output gate*. As the sequentially first block in the unit, the forget gate determines a proportion of the cell state that is forgotten, by multiplying the previous cell state with the output of the gate's sigmoid activation (ranging between 0 and 1). Next, the input gate determines how to update the cell state. This is achieved by multiplying the output of a tanh-layer (ranging from -1 to 1), determining which information to add, with the output of another sigmoid-layer, determining how much of this information to then add to the cell state. Lastly, the output gate produces a new hidden state, which is at the same time the output of the LSTM unit. For this purpose a tanh-function is applied to the cell state to determine the new hidden state information. This output is then again multiplied with the output of a sigmoid-layer, determining a proportion of information that represents the new hidden state and can be passed alongside with the cell state to the next LSTM unit. Hence, the cell state augments the short-term hidden state with a proportion of long-term information.

The LSTM has to this date proven to be one of the most successful RNN-variants on various NLP tasks. Prominently, it was used in the field of machine translation within an *Encoder-Decoder* architecture (Cho, van Merrienboer, Gulcehre, et al., 2014; Sutskever et al., 2014). In this type of model a number of multi-layered LSTMs are combined to form respectively an Encoder and a Decoder block, mapping an input to an output sequence, to solve for example a translation problem. Essentially, the encoder compresses the entire input sequence into a context vector representation of fixed dimensionality, which is then passed as input to the decoder, which uses it to generate output tokens until it produces a special *end-of-sequence* token. However, this compression was found to be a bottleneck leading to a deteriorating performance for longer sequences, especially for sentences longer than those the model observed during training (Cho, van Merrienboer, Bahdanau, & Bengio, 2014).

To address this problem, *attention* was introduced as a mechanism allowing the Decoder to focus on relevant parts of the input (Bahdanau et al., 2014; Luong et al., 2015). This is achieved by computing a similarity score, e.g. a scaled version of the dot product, between the current hidden state of the decoder relative to each of the hidden state representations of the encoder, respectively associated with an encoder input token. A softmax function is then applied to the

[3] Another alternative solution was proposed in form of the *Gated Recurrent Unit* (GRU) by Bahdanau et al. (2014).

set of similarity scores and the resulting softmax values are used to scale the encoder's hidden state representations. The scaled hidden representations are then summed to form a vector of attention values that is concatenated with the decoder's current hidden state. Running this concatenated vector through a feedforward neural network then generates the next output token. Although adding an attention mechanism strongly improved the results of encoder-decoder seq2seq models, a limitation with respect to training efficiency was still to be found in the recurrent nature of these models, since parallelization was in general not possible.

To address this shortcoming, Vaswani et al. (2017) introduced the *transformer* architecture, with the key idea to completely discard the principle of recurrence and solely use the attention mechanism, which is implemented slightly differently in the encoder and the decoder blocks and also across the blocks. Since the thesis deploys a *decoder-only* model, the following description will only consider this transformer subtype that is sketched in figure 1.1.

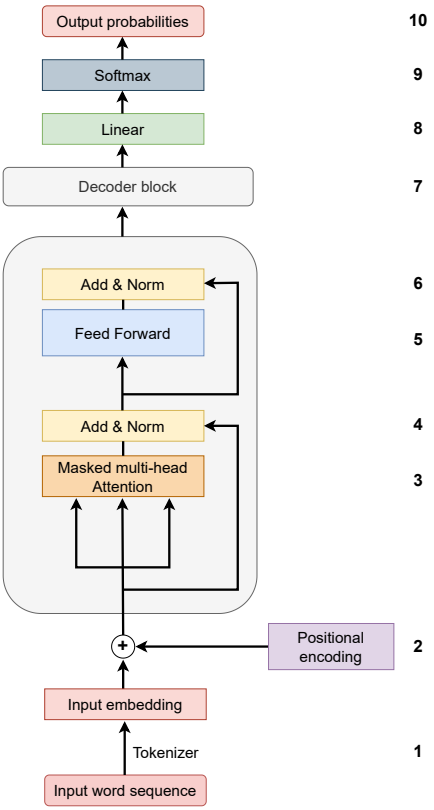

Figure 1.1.: The decoder-only transformer architecture derived from Vaswani et al. (2017).

As described above, an input sequence is tokenized into (sub)word units which are then mapped to their respective input embeddings (fig. 1.1, **1**). Due to the renunciation of recurrence, a positional encoding vector is added onto the input embeddings (fig. 1.1, **2**), for example by applying sine or cosine functions of varying frequencies to keep track of the token position within the sequence (Vaswani et al., 2017).

What follows is a layer of *masked self-attention*, which allows the model to monitor the

relationship among tokens (fig. 1.1, **3**). The key idea is to compute a similarity score between the current token and all the preceding tokens including itself[4] and to use this information to decode the next token. This is done by generating a *query* vector for the current token (with the same dimensionality as the input & positional embedding) and *key* vectors for the current and all preceding tokens.

A similiarity score is computed between the query vector and each of the key vectors, for example using a scaled version of the dot product (Vaswani et al., 2017). Then, a softmax function is applied to these similarity scores, which can conceptually be seen as determining a percentage for each (preceding and current) token of how much influence it should have on the encoding of the next token. Another vector, referred to as *value* vector, is generated for each token, and this vector is scaled by the respective similarity score associated with it after applying the softmax function. Note, that the attention function can be computed for all queries in parallel and therefore queries, keys and values can be stored in matrices Q, K and V. The original attention function proposed by Vaswani et al. (2017) can therefore be expressed as

$$\text{Attention}(Q, V, K) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{1.2}$$

where d_k represents the dimensionality of both queries and key vectors. Note, that this is the attention function that is used within the transformer model used in this thesis, described in chapter 2.

Lastly, all of the scaled value vectors for the current token are summed, and the resulting vector constitutes the self-attention embedding of what is referred to as an *attention head*. The procedure of generating self-attention representations is usually performed multiple times in parallel, leading to the term of *multi-head* attention. All the outputs that have been computed in the different attention heads are then to be concatenated and projected with a separate weight matrix onto the model's embedding size.

In the next step, the positional encodings are again added to the output of the multi-head attention in what is referred to as *residual connections*, which allows the model to neglect positional information during the self-attention process and in combination with layer normalization prevents gradients from vanishing during training with backpropagation (fig. 1.1, **4**).

Lastly, the output of the residual layer is fed as input into a fully connected feed forward neural network (fig. 1.1, **5**). After applying another residual layer, the resulting representation then constitutes the output of the decoder block (fig. 1.1, **6**). This output can either serve as input to another decoder block[5](fig. 1.1, **7**) or map to an output layer that has the dimensionality of the model's vocabulary (fig. 1.1, **8**).

After applying a softmax function to the output layer (fig. 1.1, **9**) it reflects a probability distribution over the vocabulary for what the model predicts to be the next token (fig. 1.1, **10**). Note, that this last step of obtaining next word probability is functionally the same for both RNN and transformer models. Applying the negative of a logarithmic function to the probabil-

[4]Note that this differs when the self-attention is not masked, i.e. in an encoder block, where a score is computed between the token and *all* other tokens.

[5]usually several blocks are stacked

ity estimate for an observed word in the output layer of the preceding sequence position then yields surprisal. Henceforth, surprisal estimates derived by any language model architecture will be abbreviated as **LM surprisal**.

Although immensely popular and successful when it comes to NLP tasks, the cognitive plausibility of transformer models is more or less debatable depending on their specific architecture. In contrast to all types of RNNs, which process words in a strictly incremental manner, the transformers' self-attention mechanism allows them to selectively attend to previous (or all) parts of the input. While this mechanism seems cognitively implausible at first, Merkx and Frank (2021) discuss how it could relate to cue-based retrieval theories rather than recurrent ones, where processing happens in a strictly incremental manner.

Though at least some of the model architectures clearly lack cognitive plausibility, an important observation is that there seems to be a tendency that more sophisticated models perform closer to human data (Goodkind & Bicknell, 2018; Merkx & Frank, 2021; Michaelov et al., 2021). Nevertheless, this conclusion does not universally hold (Frank et al., 2015) and specifically for reading times, raising model complexity within transformer-based models appears to have a reverse effect (Oh & Schuler, 2022).

While the operationalization of surprisal through language models can be viewed to be conceptually related to the metric of cloze probability[6], there have also been a number of different methods for a model-derived quantification of semantic association in psycholinguistically motivated studies (see Table 1.2).

As stated above, when describing the input representations of neural language models, these approaches can overall be divided into prediction-based and count-based methods. Regardless of the underlying approach, all methods derive high-dimensional vector representations of words, based on their co-occurence patterns in large text corpora. This form of representation brings along the useful property of comparability on a mathematical level. That is, the semantic relatedness between two words is quantified by the distance between their vector representation, usually measured by cosine similarity. When comparing the relatedness between a single word and its preceding context, the context is usually represented by the sum of the vector representations that it is spanning over. Usually, this sum representation is averaged, though Michaelov et al. (2023) point out that with respect to cosine similarity the magnitude is irrelevant and hence including this step should yield the same result as taking into account only the sum (Frank & Willems, 2017). In some cases, only content words are taken into account as context (Frank, 2017; Frank & Willems, 2017).

Since plausibility builds upon world knowledge outside of the linguistic domain, models that are solely trained on language do not naturally capture this notion. While surprisal in human listeners is characterized by expectations from both world knowledge and linguistic experience, surprisal in language models originates only from the latter source, that is, the linguistic input they have been trained on. This distinction has lead to the definition of *comprehension-centric surprisal* (*cc-surprisal*; Venhuizen et al., 2019). In their model of comprehension, world knowledge is introduced into the model within the framework of *Distributional Formal Semantics* (*DFS*; Venhuizen et al., 2022).

[6]Crucially, with the important difference that language model surprisal solely reflects distributional information of language, while human judgements may also be modulated by numerous other factors.

LSA	Smolka and Eulitz (2018) Frank (2017) Van Petten (2014) Parviz et al. (2012) Mitchell et al. (2010)	Pynte et al. (2008) Chwilla and Kolk (2002)
Word2Vec/fastText	Michaelov et al. (2023) Nieuwland et al. (2020) Frank and Willems (2017) Frank (2017) Ettinger et al. (2016)	
GloVe	Michaelov et al. (2023)	
Other	Van Petten (2014)	

Table 1.2.: A selection of studies that collected semantic association estimates from different distributional methods (sorted by method and publication date).

This section has reviewed different ways to operationalize the concepts of expectancy (in its high level formalization as surprisal), semantic relatedness and plausibility of a word in context with either human or model based methods. With respect to general expectancy, a commonly applied measure in psycholinguistics is to collect the cloze probability for a word by presenting participants with a sentence completion task. Language models on the other hand can generate a probability distribution over next words by either straightforwardly counting frequencies or minimizing their error during training the task of next word prediction.

Semantic relatedness can in the setting of psycholinguistic experiments be estimated by asking partcipants to rate the strength of relationship between a prime and a target word on a scale. On the model side, a common way is to derive corpus based high dimensional word representations that factor in distributional information, i.e. the co-occurrences of words in same contexts. These global word representations also serve as input to neural language models, or can be trained jointly with the model.

World knowledge driven plausibility can be estimated in a similar rating task as semantic association, by presenting participants with sentences and asking them to judge their plausibility on a scale. On the model side the notion of plausibility is more challenging if currently not even possible to operationalize. While world knowledge has been introduced in neurocognitive models of comprehension (Venhuizen et al., 2019), there is no direct correlate of this notion to be found in large scale language models.

That being said, plausibility should be reflected to a certain extent within general expectancy, that is, under a rational view of communication implausible continuations should be for the most part less expected than plausible ones. Crucially, the same assumption can not be clearly made for semantic association.

1.2. ERP-components

While behavioural measures such as reading times provide an estimate of overall processing effort and have therefore been linked to surprisal, neurophysiological research may be better able to reveal the sub-processes underlying comprehension and to answer how those processes interactively unfold in real-time. In this line of research, two components in the ERP signal

	DBC19[1]		DBC21[2]		ADSBC21[3]		ADBC23[4]	
	N4	P6	N4	P6	N4	P6	N4	P6
Association	✓	X	✓	X	✓	X	–	–
Plausibility	X	?	X	✓	–	–	X	✓
Cloze	X	?	X	✓	✓	✓	X	✓

[1] Delogu et al. (2019)
[2] Delogu et al. (2021)
[3] Aurnhammer et al. (2021)
[4] Aurnhammer et al. (2023)
– not manipulated
✓ effect found
X no effect found
? conflicting results

Table 1.3.: Overview of the N400 & P600 modulation pattern found in the four RI-studies.

have taken a prominent role: the *N400* and the *P600*.

The N400 is a negative voltage deflection peaking around 400 ms after stimulus onset and has for many years been taken to reflect the process of semantic integration, while the P600, a positive deflection emerging from around 500-600 ms, has been linked to syntactic integration. However, several findings have challenged this view, leading to more complex multi-stream accounts, which in turn have struggled to broadly account for the experimental data (see Brouwer et al., 2012 for a review). Utilizing their neurocomputational model, Brouwer et al. (2017) demonstrate a single-stream account of *retrieval* and *integration*, that is able to account for all the previous findings. According to this account, the N400 indexes the retrieval of word meaning from semantic memory, while the P600 reflects the integration of this meaning into an interpretation of the unfolding utterance.

In a series of following studies (which will from now on be referred to as **RI-studies**), experimental manipulations of expectancy, association and plausibility have led to results that provide further evidence for the *Retrieval-Integration* (*RI*) account (see Table 1.3). It has been observed that manipulations of expectancy may result in a biphasic *N400* and *P600*, although these effects may sometimes be obscured by component overlap (Van Petten and Luka, 2012; see Brouwer et al., 2017 for a discussion). Consistent with this finding, Aurnhammer et al. (2021), henceforth **ADSBC21**, have shown both the N400 and P600 to be sensitive to manipulations of cloze probability. In their context manipulation design, an intervening adverbial clause contained additional material that was either associated or unassociated to the target word, while the target that followed the intervening clause was either expected or unexpected[7] given the selectional restrictions of the main clause. While expectancy modulated both the N400 and P600, association only influenced the N400. As pointed out by the authors, the influence of expectancy on both components appears valid under the retrieval-integration account, since a higher expectancy of a word may facilitate its retrieval from long-term memory (N400) as well as ease the effort of updating the utterance meaning representation (P600), whereas differences

[7] operationalized by cloze probability

in association should only affect retrieval effort but not integration effort.

Delogu et al. (2019), henceforth **DBC19**, also found an N400-effect of association and additionally manipulated plausibility, expecting (in line with the retrieval-integration account) to elicit a P600 effect for implausible versus plausible target words, since the former would be more difficult to integrate into the unfolding utterance representation. While this was indeed the case for the *event-related & implausible*[8] condition, no P600 effect was observed in the *event-unrelated & implausible* condition (both relative to the plausible baseline). As hypothesized and confirmed by the authors in a follow-up study, the apparent absence of a P600 effect was due to spatiotemporal-overlap, i.e. the N400 and P600 overlapping both in space in time due to the additive nature of the waveform-based component structure (see Brouwer and Crocker, 2017 for a discussion and Brouwer et al., 2021; Delogu et al., 2021 for empirical evidence). Interestingly, cloze probability didn't pattern with the N400 in this study. Rather, it only patterned with the P600, in that average cloze in the two implausible conditions (event-related and event-unrelated) was significantly lower than in the baseline condition.

Delogu et al. (2021), henceforth **DBC21**, replicated these results, providing further evidence for a strong link between semantic association and the N400. Most importantly, the authors found that the P600 effect resurfaces when comparing the event-unrelated implausible condition to an event-unrelated baseline, i.e. a baseline that elicits a similar N400 amplitude. Crucially, applying the technique of *regression-based ERP* (rERP) estimation (Brouwer et al., 2021; Smith & Kutas, 2015), the authors revealed an increase in P600 amplitude in their event-unrelated implausible condition, that had been attenuated by a preceding increase in N400 amplitude. Moreover, the rERP analysis showed, that using *association & plausibility* versus *cloze & plausibility* as continuous predictors resulted in a closer fit to the EEG-data, rendering association a stronger predictor of N400 amplitude than cloze.

Further, Aurnhammer et al. (2023), henceforth **ADBC23**, established that the P600 continuously indexes integration effort: compared to a plausible baseline, a less plausible and an implausible condition led to increasingly positive P600 amplitudes. Since lexical association was high across conditions, achieved by lexical repetition of the target word, the difference in plausibility did not elicit any N400 modulations. While the mean target cloze probability didn't explicitly enter the analyses, it was considerably lower in the implausible conditions while being highly correlated with plausibility at the same time.

In sum, these studies provide evidence that both N400 and P600 are modulated by expectancy since it respectively facilitates retrieval and integration. For the N400 it was found that the manipulated association between target and context could either elicit or inhibit an effect independently of the overall expectancy as measured by cloze, which itself may have been influenced by other factors such as plausibility.

1.3. Research questions

The overarching goal of the thesis is to gain further insights into the power of LM surprisal in predicting the ERP profile of human language comprehension. A number of studies have found it to be a good predictor of the N400 amplitude (Frank & Willems, 2017; Frank et al.,

[8] Event-related being equivalent to associated here

Stimulus	A	E	Con
Yesterday sharpened the lumberjack, before he the wood stacked, the *axe*...	+	+	A
Yesterday sharpened the lumberjack, before he the movie watched, the *axe*...	-	+	B
Yesterday ate the lumberjack, before he the wood stacked, the *axe*...	+	-	C
Yesterday ate the lumberjack, before he the movie watched, the *axe*...	-	-	D

Table 1.4.: **RQ1 (a)**: showing the manipulations of Association (A) and Expectancy (E) in the conditions (Con) of ADSBC21.

2015; Merkx & Frank, 2021; Michaelov & Bergen, 2020; Michaelov et al., 2022, 2023; Parviz et al., 2012). In contrast, to our knowledge only two recent studies have utilized it to predict the P600 (De Varda et al., 2023; Li & Futrell, 2023).

One strength of the four RI-studies is that they dissociate the concepts of association, plausibility and expectancy and show how manipulations of respective operationalizations leads to either isolated or bi-phasic N400 and P600 effects. More precisely, an isolated N400 effect of association was observed in DBC21 and ADSBC21, whereas an isolated P600 effect of plausibility was elicited in DBC19, DBC21 and ADBC23. Moreover, DBC19 and DBC21 demonstrated how a simultaneous manipulation of association and plausibility can lead to a bi-phasic ERP profile, where the association-driven N400 effect is spatiotemporally concealing the plausibility-driven P600.

The results of the RI-studies indicate that the N400 is not modulated alone by overall expectancy. In fact, semantic association may be able to overwrite its influence. It is unclear, which pattern LM surprisal will follow in these cases and this leads to the first research question of the thesis:

1. Can language model surprisal capture N400 effects that have been elicited by a manipulation of association but *not* overall expectancy?

As cloze probability and association are usually correlated, an important task is to assess language model surprisal on pairs of experimental conditions that fall under one of two cases:

(a) Keeping expectancy constant, a difference in association elicited an N400 effect.

(b) Keeping association constant, a difference in expectancy did *not* elicit an N400 effect.

For (a), table 1.4 shows the relevant conditions from ADSBC21. For (b), table 1.5 shows the relevant conditions from DBC19 and DBC21. Under the assumption that LM surprisal reflects a measure of general expectancy that encapsulates a number of constituting factors such as semantic, syntactic and pragmatic constraints and world knowledge driven plausibility, the expected outcome is that it overall strongly patterns with cloze (as a human derived operationalization of expectancy) while not as strongly with semantic association which can make different predictions than expectancy. Thus, LM surprisal should falsely predict no N400 effect between the conditions in (a) and it should falsely predict an N400 effect between the conditions in (b).

	Stimulus	A	E	Con
DBC19	John entered the restaurant. Before long, he opened the *menu*...	+	+	A
	John left the restaurant. Before long, he opened the *menu*...	+	-	B
DBC21	John left the restaurant. Before long, he opened the *umbrella*...	-	+	B
	John entered the restaurant. Before long, he opened the *umbrella*...	-	-	C

Table 1.5.: **RQ1 (b)**: showing the manipulations of Association (A) and Expectancy (E) in the conditions (Con) of DBC19 and DBC21.

Another finding from ADSBC21 is, that the P600 is sensitive to expectancy but crucially not to association. Thus, this leads to the second research question:

2. Does language model surprisal provide a strong predictor of the P600 amplitude?

Again, under the assumption that LM surprisal reflects the general notion of expectancy which may itself be subject to various influences, the anticipated outcome here is that LM surprisal should be able to capture cases where decreased expectancy due to world knowledge violations led to P600 effects. It has to be noted that in all four RI-studies the target word that elicited the P600 was a less plausible or implausible continuation given the context. While it is possible, that a rather unexpected continuation may be a plausible one (see example 1 b), the reverse case seems unlikely from a communicative point of view. That is, implausible continuations should simultaneously be less expected. Considering that the language model which is generating the surprisal estimates was trained on a naturalistic corpus (rather than carefully manipulated experimental stimuli), the number of implausible continuations it was trained on is most likely small. Therefore, the expected outcome with respect to the second research question is an overall strong correlation.

Another point raised by Michaelov et al. (2023) is to consider how different language model architectures may yield a better or worse fit to the N400 and P600 window respectively. The authors cited a related study (Michaelov & Bergen, 2020), which found a better correlation of RNN-based estimates with post-N400 positivities than with the N400. Therefore, surprisal estimates of both a recurrent based architecture (LSTM) and a transformer based architecture (GPT-2) will be considered for both research questions.

Chapter 2.

Methodology

With the research questions laid out, this section offers a closer look into the foundational experimental data as well as the (computational) methods that are used to answer them. The thesis seeks to specifically evaluate the four RI-studies, that have all been conducted in German. Depending on which factors were manipulated, a combination of cloze, association and plausibility ratings for the target words have been collected during pre-studies (see Table 2.1). Having the ratings available alongside the stimuli, the first step was to collect surprisal estimates from the two language models. To incorporate two prominent architectures that have been used by other studies, both an LSTM and a Decoder-only transformer model were used. While the transformer model used was pre-trained, the LSTM had first to be trained as an additional intermediate step.

The next step was to evaluate how the surprisal estimates are distributed generally across conditions and how human ratings and model estimates correlate. In particular, a correlation between LM surprisal and plausibility was expected to be a first indicator that LM surprisal might be a good predictor of the P600 amplitude (RQ2). A correlation with cloze might also reflect a good fit between human derived and model derived measures of expectancy. Alongside assessing overall correlations, initial judgements were made for specific sets of conditions. Under the assumption that LM surprisal is strongly reflecting overall expectancy and only to a lesser extent semantic association, it was expected to observe the following with respect to RQ1:

1. There is no systematic difference for LM surprisal estimates comparing conditions A versus B and C versus D in ADSBC21.

2. Lm surprisal estimates will be systematically higher in condition B versus A in DBC19 and C versus B in DBC21.

Patterns in mean surprisal between conditions were however not be assessed inferentially (Sassenhagen & Alday, 2016), but rather reported and used descriptively. To assess more thoroughly how well the surprisal estimates can predict the ERP-effects observed in the RI-studies an rERP analysis was conducted, respectively for each model architecture and each study. In the remainder of this chapter, first the stimulus material of the four RI-studies will be introduced. Next, the language models that were used will be presented and their architectural differences will be highlighted. Finally, the rERP-analysis procedure that was applied will be described.

	DBC19	DBC21	ADSBC21	ADBC23
Association	✓	✓	✓	–
Plausibility	✓	✓	–	✓
Cloze	✓	✓	✓	✓
#Items	90	90	120	60
#Conditions	3	3	4	3
#Participants	26	23	40	30

Table 2.1.: Overview of the stimuli and ratings of the four RI-studies.

2.1. Experimental Stimuli and Ratings

While the motivation behind the RI-studies and the distinct patterns of N400 and P600 effects they found is described in section 1.2, this section provides a closer view on the stimulus material that was used, including the different experimental conditions and the respective human judgements of cloze probability, semantic association and plausibility. All studies featured an ERP-experiment that is relevant for this thesis. Additionally, ADSBC21 and ADBC23 each also conducted a reading time study which will not be included in the analysis. With respect to the ERP-experiments, all four RI-studies followed the same setup, deploying 26 active scalp electrodes according to the 10-20 system. The context sentence or paragraph was presented in its entirety until a button was pressed. Then, a fixation-cross appeared in the center of the screen for 750 ms. After that, using rapid serial visual presentation (RSVP) the final sentence containing the target word was shown word by word. All studies implemented a design that manipulated the context rather than the target word between conditions.

DBC19 features 90 items with 3 conditions, leading to a total of 270 stimuli. Each stimulus consists of a single context sentence that precedes the target sentence including a target noun. Run as separate pre-studies, metrics for association, plausibility and expectancy were collected. The English translation of an example item and the respective ratings per condition are provided in Table 2.2.

For association, the semantic relatedness between the prime (locations or activities mentioned in the context) and the target noun was rated on a scale ranging from 1 (not related) to 7 (strongly related) by a total of 20 participants. For conditions A and B the prime and target were equal, sharing a mean rating of 6.32 (SD=0.53) while for condition C the mean rating was 1.56 (SD=0.46).

With respect to plausibility, 30 participants rated the entire stimulus on a 1-7 scale up until including the target, excluding the rest of the target sentence. This resulted in a mean rating of 6.28 (SD=0.53) for condition A, 2.42 (SD=0.80) for condition B and 1.93 (SD=0.82) for condition C.

Lastly, cloze probability was estimated by presenting the stimulus up until including the determiner preceding the target noun to participants, collecting their continuations. For condition A cloze probability was 0.38 (SD=0.33), for condition B 0.13 (SD=0.19) and for C 0.008 (SD=0.4).

To summarize, the baseline condition features targets that relative to the other conditions

Con	A	P	C	Stimulus
A	6.32	6.28	0.38	John entered the *restaurant*$_{p}$. Before long, he opened the *menu*...
B	6.32	2.42	0.13	John left the restaurant. Before long, he opened the *menu*...
C	1.56	1.93	0.008	John entered the apartment. Before long, he opened the *menu*...

Table 2.2.: **DBC19**: mean ratings in each condition (Con) for semantic association (A), plausibility (P) and cloze probability (C) alongside an example stimulus.

rank high in association, plausibility and cloze. It was referred to as "Baseline". For condition B, association stays the same as the prime noun/activity remains unchanged, but plausibility lowers due to the context featuring a different main verb that leads to a violation of the typical sequential ordering of sub-events. Overall expectancy in terms of cloze probability also lowers in this condition. The condition was referred to as "Event related violation". In condition C then association decreases relative to A/B due to a different prime. The overall context is also changed so that it renders the activity carried out in the target sentence rather implausible. Cloze probability is also lower than in conditions A and B. Condition C was referred to as "Event unrelated violation".

Similarly to DBC19, **DBC21** also features 90 items with 3 conditions and a total of 270 stimuli. The stimuli also consist of a single context sentence followed by the target sentence containing a target noun. Judgements for association, plausibility and cloze probability were collected in the same way as in DBC19. The English translation of an example item and the respective mean ratings per condition are provided in Table 2.3.

For association, the mean rating was 6.66 (SD=0.47) for condition A. Conditions B and C share the same prime and target combination, therefore the mean rating was 1.46 (SD=0.56) for both conditions.

For plausibility, condition A had a mean rating of 6.65 (SD=0.45), condition B 5.45 (SD=0.79) and C 1.61 (SD=0.49)

Cloze probability was 0.61 (SD=0.27) for condition A, 0.02 (SD=0.04) for condition B and 0.004 (SD=0.03) for condition C.

To summarize, condition A features targets that relatively to the other conditions exhibit high ratings in association and plausibility as well as a high cloze probability. It was referred to as "Related-Plausible". For condition B the context sentence changes so that the target sentence is slightly less but still plausible while at the same time being only very weakly associated to the context. Cloze probability is also lower in this condition that was termed "Unrelated-Plausible". Condition C contains the same prime-target combination as B, therefore the association stays weak relative to A. Applying the same type of plausibility violation as DBC19, changing the sequential ordering of sub-events, plausibility is substantially lower compared to both A and B as well as cloze probability. This condition was referred to as "Unrelated-Implausible".

ADSBC21 features 4 conditions with 120 items per condition, resulting in a total of 480 stimuli. The English translation of an example item and the respective ratings per condition are provided in Table 2.4. The stimuli consist of a single sentence. To manipulate association and expectancy independently, an either associated or unassociated adverbial clause was inserted into the main clause before the target word, while the high vs. low expectancy manipulation

Con	A	P	C	Stimulus
A	6.66	6.65	0.61	John went out in the rain. Before long, he opened the *umbrella*...
B	1.46	5.45	0.02	John left the restaurant. Before long, he opened the *umbrella*...
C	1.46	1.61	0.004	John entered the restaurant. Before long, he opened the *umbrella*....

Table 2.3.: **DBC21**: mean ratings in each condition (Con) for semantic association (A), plausibility (P) and cloze probability (C) alongside an example stimulus.

Con	A	C	Stimulus
A	6.29	0.67	Yesterday sharpened the lumberjack, before he the wood stacked, the *axe*...
B	2.09	0.64	Yesterday sharpened the lumberjack, before he the movie watched, the *axe*...
C	6.29	0.008	Yesterday ate the lumberjack, before he the wood stacked, the *axe*...
D	2.09	0.008	Yesterday ate the lumberjack, before he the movie watched, the *axe*...

Table 2.4.: **ADSBC21**: mean ratings in each condition (Con) for semantic association between the target and the noun of the adverbial clause (A) and cloze probability (C) alongside an example stimulus.

was determined by the verb of the main clause preceding the adverbial clause.

For association norming, 60 participants rated the relatedness between each content word of the adverbial clause and the target word on a 1-7 scale (where 7 represented a strong association). Since the difference between associated versus un-associated was greater for Noun-target than Verb-target, only Noun-target association is considered in the thesis and thus reported in Table 2.4. Conditions A and B share the same main clause, crossed with B and C sharing the same adverbial clause. This leads to conditions A and C sharing the same higher association estimate of 6.29 (SD=0.82) and conditions B and C sharing the lower association estimate of 2.09 (SD=1.01).

Expectancy was operationalized through cloze probability, which was collected in a pre-study where 48 participants were asked to complete the stimuli up to but excluding the determiner preceding the target word. For condition A, cloze probability was 0.67 (SD=0.23), for condition B 0.64 (SD=0.23), for condition C 0.008 (SD=0.025) and for condition D 0.008 (SD=0.028).

Ratings for plausibility were not collected though it can be strongly assumed that this would have resulted in very high ratings for conditions A and B and very low ratings for C and D, since the expectancy manipulation strongly violates the argument restrictions of the main clause verb.

To summarize, conditions A and B both feature the same verb in the main clause that leads to a high cloze probability of the target word. Semantic association is controlled by the inserted adverbial clause and is high for A but low for B. Conditions C and D on the other side feature the same adverbial clauses as respectively A and B, with high association for C and low association for D. Cloze probability is low in both C and D as the conditions share the same main clause verb that renders the target word to be unlikely (and implausible).

ADBC23 features 60 items with 3 conditions resulting in a total of 180 stimuli. The English translation (in German word order) of an example item and the respective ratings per condi-

Context			
A tourist wanted to take his huge suitcase onto the airplane. The suitcase was however so heavy that the woman at the check-in decided to charge the tourist an extra fee. After that, the tourist opened his suitcase and threw several things out. Now, the suitcase of the ingenious tourist weighed less than the maximum of 30 kilograms.			
Con	**P**	**C**	**Continuation**
A	5.84	0.8	Then dismissed the lady the *tourist*...
B	3.69	0.09	Then weighed the lady the *tourist*...
C	1.42	0.02	Then signed the lady the *tourist*...

Table 2.5.: **ADBC23**: mean ratings in each condition (Con) for plausibility (P) and cloze probability (C) alongside an example stimulus.

tion are provided in Table 2.5. In order to prevent N400 effects at the target word, the same context paragraph precedes the target sentence in each condition, mentioning the target and a distractor word several times to maximally prime the target word's meaning. Cloze probability and plausibility of the target are then manipulated by using a different main verb in the target sentence, rendering the target continuation more or less expected and plausible.

Cloze probability was collected by presenting 90 participants with the entire context paragraph and the target sentence up until but excluding the determiner of the target word. The determiner was excluded to not constrain the set of possible continuations. This led to a cloze probability of 0.8 (SD=0.2) for condition A, 0.09 (SD=0.11) for condition B and 0.02 (SD=0.04) for condition C.

For plausibility norming, 60 participants were presented with the entire context plus the target sentence up until including the target word and asked to rate it on a scale ranging from 1 (not plausible) to 7 (very plausible). For condition A, this resulted in a mean rating of 5.84 (SD=0.93), 3.69 (SD=1.33) for condition B and 1.42 (SD=0.33) for condition C.

To summarize, in condition A the target word provides a highly plausible and expected argument to the preceding verb. In condition B the verb changes so that the target becomes less plausible (while at the same time the distractor becomes more attractive). In condition C the plausibility violation intensifies as the verb here requires an argument with properties that the target doesn't typically fulfill, i.e. "signing" is typically not an activity performed on humans. The drop in plausibility is also reflected in substantially lower cloze probabilities in both conditions B and C.

2.2. Language Model Architectures

A goal of the thesis is to compare model archtitechtures that substantially differ in their internal mechanisms to arrive at probability estimates for the next word in a sequence. As reviewed in chapter 1, three of the most prominent model types that have been used to predict the ERP are n-gram models, RNNs and transformer-based models. N-gram models provide some appealing features compared to the other two types: they compute their probability estimates by straightforwardly counting word frequencies on a training corpus, normalizing them over the vocabulary size and adding a smoothing mechanism to handle out-of-vocabulary words.

One of their greatest limitations though is their restricted access to context.

The four RI-studies all implement a context manipulation design, keeping the target word constant across conditions. The manipulation of the context though, that leads to differences between conditions, happens outside a window that is feasible for an n-gram with $n < 6$, except for a range of items in ADBC23. For this reason, n-gram models were not considered here. Instead, the aim was to compare the decoder-only transformer architecture, using attention without recurrence, to an LSTM, a recurrent network without attention.

As an instance of the decoder-only transformer architecture, the aim was to use a pre-trained German GPT-2 model (Schweter, 2020). Since it was found that due to a bug in the tokenizer, leading to a mapping of one vocabulary id to more than one token, the original model's performance was not optimal, a re-trained version of the model was kindly shared by its author, and this version has been used in this thesis (Schweter, n.d.-b).

Alongside the re-trained model, the training data was also shared. The overall size amounts to approximately 16 gigabytes of data and comprises several smaller sub-corpora (see table 2.6 for an overview). The sub-corpora include a Wikipedia dump, NewsCrawl, ParaCrawl, EU bookshop corpus and Open Subtitles. The authors of original pre-trained GPT-2 model (Schweter, 2020) reference to the training data that was used for a German BERT model (Schweter, n.d.-a). It is to note that while the file size and the number of tokens approximately matches between the training set used in this thesis and the set referenced by Schweter (n.d.-a), the latter also reference CommonCrawl as a sub-corpus, which doesn't appear to be part of the shared training data.

Although model comparisons between different architectures are difficult to draw, the goal was to train an LSTM that approximates the GPT-2 model in as many aspects as possible. Therefore, the same training dataset and the same tokenizer were used, and since no further preprocessing was conducted on the data this results in the same vocabulary for both models. Moreover, the embedding dimensionality, optimizer and dropout rate of the LSTM were chosen to be the same as for the pre-trained GPT-2 model. For an overview of the hyperparameters of both models see table 2.7.

In order to train the LSTM the LM toolkit software (Mosbach et al., 2023) was used. To assess both models' general capabilities, perplexity was computed on a small test set of 25 German publicly available texts from different genres. The genres encompass fairy tales, newspaper articles, short stories, Wikipedia articles and abstracts of scientific studies. Each genre is represented by 5 texts and all texts amount to a word count of 6542. A more detailed overview of the text material is given in Appendix C. Perplexity was computed with the following formula

$$PPL = 2^{-\frac{1}{N}\sum_{i=1}^{N} log_2 p(x_i)} \tag{2.1}$$

where N is the number of tokens in the test set and $p(x_i)$ is the conditional probability the model assigned to token x_i. The individual texts were splitted into sentences using spaCy (Honnibal et al., 2020) and sentences served as input to the model, that is, the internal state of the model was resetted after each sentence.

	# Tokens	# Lines	File size
Wikpedia	743,960,730	17,290,362	5.1G
News	610,679,770	38,647,678	4.1G
Paracrawl	450,749,095	42,385,503	3.1G
EU Bookshop	302,833,155	18,203,612	2.2G
Open subtitles	226,426,451	41,612,280	1.3G
Total	2,334,649,201	158,139,435	15.8G

Table 2.6.: The training data used for both language model architectures: GPT-2 and LSTM.

	GPT-2	LSTM
Parameters	124M	58M
Vocab size	50,257	50,257
Context size	1024	128
Embedding dim	768	768
Decoder/Hidden layers	12	4
Attention heads	12	-
Training epochs	20	3,5
Optimizer	Adam	Adam
Dropout	0.1	0.1
Test perplexity	134.50	301.38

Table 2.7.: Parameter overview for the two deployed language model architectures: GPT-2 and LSTM.

2.3. Data Preprocessing and Surprisal Retrieval

The stimulus material and human ratings of the 4 RI studies are publicly available and were kindly provided by the authors file format. To serve as input to the language models, the stimuli of all studies were preprocessed in the same way by firstly applying a regular expression. This expression ensured that a single whitespace is present each time a non-alphanumeric character is followed or preceded by an alphanumeric one. As an example, the string *"Johann betrat das Restaurant."* was modified to *"Johann betrat das Restaurant ."*. This entails that the tokenizer, shared by both language models, treats non-alphanumeric characters as unique tokens instead of belonging to a preceding or subsequent word. This step is relevant since the pre-trained GPT-2 model (and hence the LSTM) rely on a tokenizer that implements Byte-Pair Encoding (Sennrich et al., 2016). Initially developed as a text-compression algorithm, BPE is used in many transformer models to efficiently reduce vocabulary size. The core idea is to arrive at a compact vocabulary that is based on the most frequent subunit combinations. Roughly described this is achieved by firstly gathering the unique set of words of a training corpus. Secondly, the characters that constitute these words are recorded alongside with their frequencies. Then, the character combinations that co-occur most frequently are merged into a subword unit and added to the vocabulary. This last step is then applied recursively to the subword units that have been found in a previous iteration. Although this method is very efficient from an engineering perspective, it introduces a potential problem when being used for neurocognitive

modeling that will be discussed in chapter 4.

For all RI studies the target word was not the final word of the target sentence. Hence, as a next preprocessing step the stimuli were cut off after the target word. The preprocessed stimuli were then presented stimulus by stimulus to the tokenizer and models, including all the prior context up until including the target word. As an example, the original stimulus *"Claudia verließ den Blumenladen. Schnell fragte sie nach einer* ***Rose*** *für eine Freundin."* would be presented to the tokenizer as *"Claudia verließ den Blumenladen . Schnell fragte sie nach einer* ***Rose****"*, the target being marked in bold here for illustrational purposes. The tokenizer then splitted the stimulus into BPE-tokens and encoded them into a sequence of ids which served as input to the model. Additionally, the id of a special beginning of sequence (*bos*) token was prepended to the input.

The amount of context that precedes the target word differs between the RI-studies with **ADBC23** having the greatest context length by a large margin. Even when considering the larger amount of tokens induced by the BPE subword splitting, the context length in **ADBC23** was not a concern for the GPT-2 model's context window of 1024 tokens. However, three items from this study minimally exceeded the LSTM's context window of 128 tokens, which resulted in the context being chunked, so that for the target in these items the context wasn't considered to its full extent by the LSTM.

After retrieving the input word embeddings for the respective token ids, the model executed a forward pass, leading to the output layer containing a vector of logits for each input token, including the beginning-of-sequence token. Each output vector had the dimensionality of the vocabulary size (50,257), and after applying a softmax function the value on each dimension represented the probability that the model assigned to the token associated with the respective id to be the next token. That is, the probability the model assigned to the vocabulary id number 17 being the token at sequence position t is the softmax value of dimension 17 at position $t-1$. Therefore, the probability for each token was collected by looking back into the output vector at the previous sequence position, extracting the value at the dimension of the token's vocabulary id. Surprisal was then computed by applying the $-log_2$ function to the probability estimate.

The BPE-tokens were then combined back into the words of the original stimulus by using explicit whitespace information that the tokenizer is keeping track of. That is, the tokenizer encoded a word differently, depending on whether it is preceded by a whitespace or not. For example, *"Laufband"* would be encoded with the ids *13412* and *4662* while *"Laufband"* (containing whitespace) would be encoded as *2879* and *4662*, where *4662* would be decoded as *"band"*, *13412* as *"Lauf"* and *2879* as *"ĠLauf"* [1]. Since each word in the input sequence was separated by a single whitespace (also the first one after prepending the bos-token), this information could be used to concatenate a token to the previous one when it didn't contain trailing whitespace after decoding. Analogously, the surprisal values for the concatenated tokens were added to collect an estimate for complete words. Although not unproblematic from a theoretical perspective, this method has been applied in other work (De Varda et al., 2023;Oh and Schuler, 2022).

Due to the preprocessing step of cutting of the stimulus after the target, the surprisal for the last word was equal to the target word surprisal.

[1] The special character *Ġ* is indicating the whitespace.

2.4. Data Analysis

After retrieving surprisal estimates from both the GPT-2 and the LSTM models for all target words, the first step in analyzing the data was to compute descriptive statistics and visually assess the surprisal distributions for each condition within each of the four RI studies. This assessment was to provide a first estimate about how well LM surprisal overall patterns with effects found between conditions in the experiments, irrespectively of how well it predicted specific effects in the N400 or P600 window.

Moreover, the correlations between LM surprisal and the human association, plausibility and cloze ratings were computed. If LM surprisal reflects the notion of general expectancy, it may be anticipated to observe an overall strong correlation with cloze probability and to a lesser extent with the semantic association and plausibility ratings. To assess the relationship between LM surprisal and the human ratings further, a linear model will be fit, using the ratings as predictors for LM surprisal. That is, a model with the specification

$$Y = \beta_0 + \beta_1 \text{Association} + \beta_2 \text{Plausibility} + \beta_3 \text{Cloze} + \varepsilon \tag{2.2}$$

was fit for DBC19 and DBC21. For ADSBC23 and ADBC23 the specification will miss plausibility or association as predictor respectively, since ratings were not collected.

An essential part of the methods of DBC21, ADSBC21 and ADBC23 is the application of the rERP framework (Brouwer et al., 2021; Smith & Kutas, 2015). This technique of analysis allows revealing ERP-effects that may be hidden due to spatio-temporal overlap (Brouwer et al., 2021; Delogu et al., 2021). More specifically, it is able to isolate the individual contributions of experimentally manipulated factors to the observed ERP profile. In its essence, "the core idea is to replace each individual voltage measurement in the ERP data—each observed voltage scalar—with a voltage estimate from a linear regression model that optimally combines the manipulated variables to explain the variance in the signal." (Brouwer et al., 2021, p. 976). In the re-estimated signal the contributions of single variables can become visible by keeping the other variables constant. With the ERP-data available, the goal for this thesis was to re-estimate the amplitudes from the RI-studies, using the surprisal estimates from the previously trained language models respectively as a single predictor. Crucially, the aim was to observe the explanatory power of LM surprisal in isolation, leading to the model specification

$$Y = \beta_0 + \beta_1 \text{LM surprisal} + \varepsilon \tag{2.3}$$

With this equation a separate linear regression model was fitted for each subject at each electrode and each time sample. Then, the forward solutions were averaged across subjects, analogous to the traditional ERP-analysis procedure in which condition averages are computed from individual subject means (as described in ADBC23). Importantly, the estimates of the models were only informed by the properties of the stimulus, that is, models didn't have access to condition-coding information and estimates were only averaged per condition retrospectively. The evaluation included comparing the re-estimated ERP with the observed ERP, inspecting the trajectory of the surprisal coefficient and the residual error across time and electrodes. For the precise ranges that were defined in the RI-studies for the N400 and P600 time

windows it was evaluated if LM surprisal was a significant predictor. For this purpose, following ADBC23, the same models were computed across subjects, resulting in a single t and p-value for each electrode and time sample. To account for the multiple comparisons problem arising from the multitude of time samples and electrodes, p-values were corrected, following the method proposed by Benjamini and Hochberg (1995), that ADBC23 implemented.

Chapter 3.

Results

In this chapter the results from obtaining the LM surprisal estimates for the stimulus material of the four RI-studies and using them in the rERP analysis will be presented. Figure 3.1 displays the densities for the surprisal estimates from both language models within the RI-studies across conditions. Mean values, standard deviations and ranges for the surprisal distributions from both language models are displayed for each of the RI-studies in Table 3.1.

As described in chapter 2, both language models are implemented with a sub-word tokenizer that utilizes byte-pair-encoding (Sennrich et al., 2016). As will be discussed in chapter 4, while not easily avoided, this may have an impact on the surprisal estimates of words that are split during tokenization. Therefore, the proportion of instances where the target word was subject to being split into sub-words by the tokenizer was recorded for all studies. Splits occurred for 43 % of the target words in DBC19, 33 % in DBC21, 15 % in ADSBC21 and 45 % in ADBC23.

Overall, mean LSTM surprisal is higher than mean GPT-2 surprisal. This was to be expected due to the higher perplexity of the LSTM on the test set, reflecting its overall more limited performance as language model. For the GPT-2 model, mean surprisal is highest for DBC19, followed by DBC21, ADSBC21 and ADBC23 in descending order. For the LSTM, the order of ADSBC21 and ADBC23 is reversed and mean surprisal for DBC21 and ADBC23 is very similar.

The next subsections follow the same structure in presenting specific results for respectively one of the RI-studies. First, surprisal statistics and densities within conditions will be provided and used to state an expectation about which of the N400 and P600 effects may potentially be captured. Then, correlations between human ratings (association, plausibility and cloze) and LM surprisal will be reported. Since the ratings are generally not normally distributed, Kendall correlation was used. For an approximate breakdown of correlation strength, cutoff values from Botsch (2011) are used. Further, linear models featuring the available human ratings as predictors of LM surprisal will be presented. Predictors were z-transformed for all linear models.

Finally, results of the subsequent rERP analysis will be reported respectively for GPT-2 and LSTM surprisal, focusing on electrode Pz. First, the experimentally observed voltages per condition will be presented together with the surprisal coefficient plotted over time, to enable a direct visual assessment of where the coefficient deviates from the intercept and which of the observed voltage deflections it may capture. Moreover, the models' forward solutions and residuals relative to the observed values, averaged across conditions, will be presented and used to gauge which of the observed effects the surprisal predictor can capture to which extent. Lastly, to assess significance of the surprisal predictor, t-values will be plotted for nine central electrodes alongside bars representing time samples where t-values were significant

	GPT-2			LSTM		
	Mean	**SD**	**Range**	**Mean**	**SD**	**Range**
DBC19	11.49	5.49	0.81-38.18	13.93	7.31	3.33-61.09
DBC21	10.07	4.42	0.79-21.18	11.52	4.48	3.06-24.33
ADSBC21	6.79	3.51	0.18-19.71	9.83	4.03	2.13-36.53
ADBC23	4.96	4.24	0.07-19.42	11.47	4.66	2.33-28.10

Table 3.1.: Mean values, standard deviations and ranges for the distributions of surprisal estimates from both the GPT-2 and LSTM model for the four RI-studies.

after multi-comparison correction.

For the computation of descriptive statistics and correlations, Python's Pandas library (The pandas development team, 2020) was used. Linear models of LM surprisal, using human ratings as predictors, were fitted with R (R Core Team, 2023). The authors of ADBC23 provided their implementation of the rERP analysis procedure, implemented in Julia (Bezanson et al., 2017) and R, which was adapted and modified for this work.

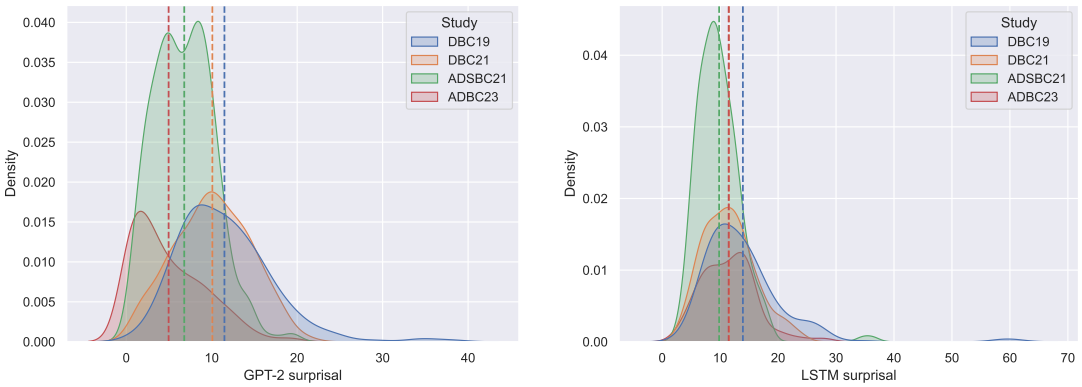

Figure 3.1.: Densities for the surprisal estimates from both the GPT-2 and LSTM models within the RI-studies across conditions. Vertical lines indicate the condition mean.

3.1. DBC19

The mean LM surprisal for both models per condition is displayed in Table 3.2. Densities are plotted in figure 3.2. Item 44, featuring the target word *Weihnachtskiosks* (English: christmas kiosk), stands out with a very high surprisal for both models across conditions[1]. One possible reason could be that this word is split by the tokenizer into 4 subwords: *Weihnachts*, *ki*, *os*, *ks*. Adding the surprisals for several subwords that may each respectively had a low probability may have resulted in this very high compound surprisal. Though this item constitutes a potential outlier with respect to LM surprisal, all items are required to conduct the rERP analysis,

[1] A: 35.17, B: 34.06 and C: 38.17 for GPT-2; A: 58.60, B: 59.34 and C: 61.09 for the LSTM.

therefore no items have been excluded.

Overall, mean surprisal of both models patterns with the mean association ratings within conditions (see Table 2.2). Conditions A and B share the same prime and target combination with a strong association rating, while C features a different prime with a weak rating. Analogously, mean LM surprisals appear more similar to each other (though not identical) and relatively lower in conditions A and B, compared to a higher and more distinct mean surprisal in C. This holds for both models. Plausibility ratings and cloze probability on the other hand show a different pattern, with gradually lower ratings for B and C compared to A, which LM mean surprisal does not seem to strongly follow. It is to note though, that mean surprisal in B appears slightly higher compared to A. This difference is more pronounced for the GPT-2 model, while hardly observable in the LSTM.

	GPT-2			**LSTM**		
	Mean	**SD**	**Range**	**Mean**	**SD**	**Range**
A	9.83	5.15	0.85-35.17	13.20	7.23	3.33-58.6
B	10.64	5.23	0.81-34.06	13.45	7.39	4.03-59.34
C	14.00	5.22	2.76-38.17	15.13	7.24	5.68-61.09

Table 3.2.: **DBC19** mean, standard deviations and ranges for the distributions of surprisal estimates for theGPT-2 and LSTM model per condition.

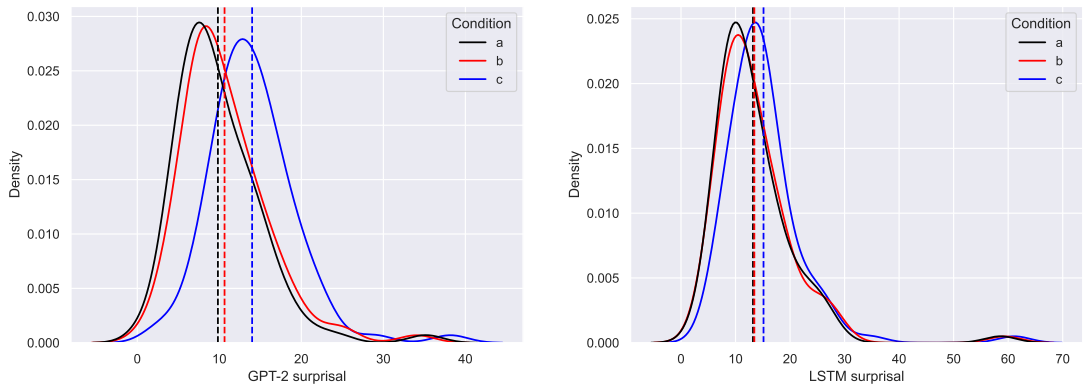

Figure 3.2.: **DBC19**: densities for the surprisal estimates from both the GPT-2 and LSTM models within conditions. Vertical lines indicate condition mean.

Correlations are reported for respectively each model in Table 3.3. There is a strong positive correlation (0.58) between the surprisal estimates of the language models. For GPT-2, there is a moderate negative correlation with association (-0.26) and cloze probability (-0.24), and a weak negative correlation with plausibility (-0.18). For the LSTM, there is a weak negative correlation with association (-0.13) and cloze probability (-0.10) and only a very weak negative correlation with plausibility (-0.07). Relatively, both models show the strongest correlation with association, followed by cloze probability and the weakest correlation with plausibility.

Overall, correlation strengths are weaker for the LSTM than for the GPT-2 model.

	GPT-2	LSTM	Association	Plausibility	Cloze
GPT-2	1	0.58	-0.26	-0.18	-0.24
LSTM	0.58	1	-0.13	-0.07	-0.10
Association	-0.26	-0.13	1	0.35	0.43
Plausibility	-0.18	-0.07	0.35	1	0.43
Cloze	-0.24	-0.10	0.43	0.43	1

Table 3.3.: **DBC19** Kendall correlation between human ratings and LM surprisal.

A summary of the linear regression models predicting LM surprisal with association, plausibility and cloze ratings is presented in table 3.4. Overall, the predictors could explain more variance in GPT-2 surprisal (R^2=0.13) than in LSTM surprisal (R^2=0.02). It is to note that the F-statistic for the LSTM model was not significant (F(3,266)=1.94, p=0.12). For GPT-2, all predictors were negatively associated with surprisal, but only semantic association was significant (t=-4.26, p<0.001). Still, cloze (t=-1.8, p=0.07) was a stronger predictor than plausibility (t=-0.01, p=0.99). For LSTM surprisal, no predictor was significant, though association (t=-1.67, p=0.10) was stronger than cloze (t=-1.09, p=0.28), while the influence of plausibility was weakest (t=0.48, p=0.63).

	GPT-2			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	11.49	0.31	36.59	<0.001
Association	-1.61	0.38	-4.26	<0.001
Plausibility	-0.01	0.41	-0.01	0.99
Cloze	-0.66	0.37	-1.8	0.07
	R^2=0.13	F(3,266)=12.97, p<0.001		
	LSTM			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	13.93	0.44	31.48	<0.001
Association	-0.89	0.53	-1.67	0.10
Plausibility	0.27	0.57	0.48	0.63
Cloze	-0.56	0.52	-1.09	0.28
	R^2=0.02	F(3,266)=1.94, p=0.12		

Table 3.4.: **DBC19** Linear regression model summary for LM surprisal.

To recap, DBC19 observed a negativity for condition C relative to A and B in a time window of 300-500 ms, and a positivity of condition B relative to A, which was significant in the 800-1000 ms time window . Another late positivity of condition C relative to A was not observable, potentially due to component overlap (see Figure 3.3, left).

Considering the patterns of mean values for association, plausibility, cloze and LM surprisal per condition in conjunction with the overall correlations and results of the linear regression models leads to the following expectations towards the outcome of the rERP analysis.

N400: LM surprisal will predict the N400 effect in condition C (unrelated) relative to A and B (both related). It will not predict an N400 effect between A and B.

P600: LM surprisal will not predict the P600 effect in B relative to A.

Next, the rERP results for both models are presented. The observed ERP is displayed in both Figure 3.3 and Figure 3.6 on the left side.

GPT-2 surprisal. Relative to the intercept, the surprisal coefficient evolves negatively in the 350-500 ms window but not positively in the 800-1000 ms window (Figure 3.3). Inspecting the models' forward solutions (Figure 3.4, left), surprisal can predict the N400 in condition C (relative to A and B). Though, the magnitude of the N400 effect is underestimated, as can be observed from the large residuals for condition C in this time window (Figure 3.4, right). The P600 effect in B (relative to A) appears to be not predicted by GPT-2 surprisal, as there is no visually observable difference between conditions in the models' estimates and large residuals for condition B in the 800-100 ms window. As can be seen in Figure 3.1, GPT-2 surprisal was a significant predictor in the N400 window but never significant in the P600 window across all 9 central electrodes.

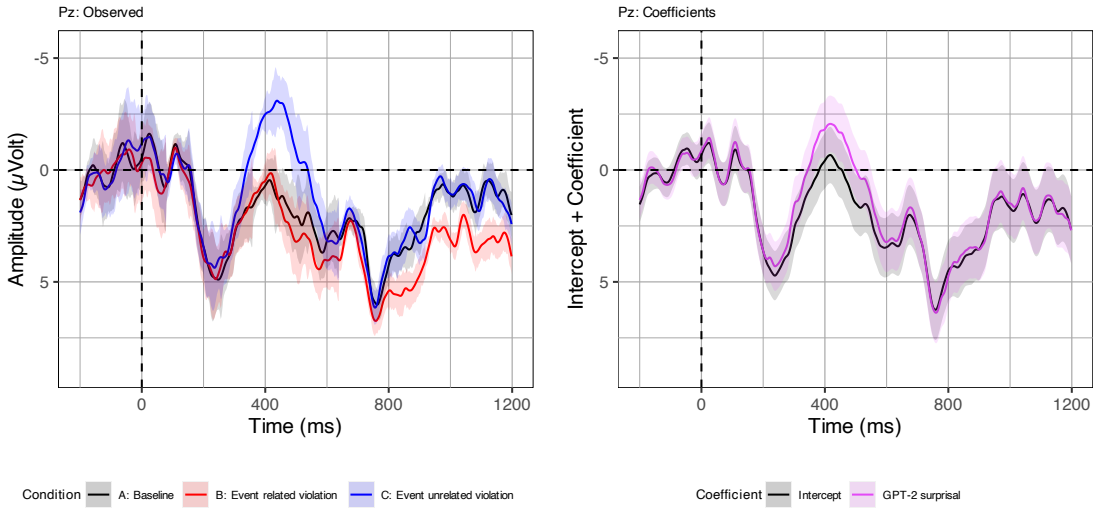

Figure 3.3.: **DBC19**, **GPT-2**: observed voltages per condition (left) and the surprisal coefficient over time (right).

LSTM surprisal. Relative to the intercept, the surprisal coefficient evolves slightly negatively in the 300-500 ms window but only minimally positively in the 800-1000 ms window (Figure 3.6). Inspecting the models' forward solutions (Figure 3.7, left), there is hardly any observable difference between conditions across time, though condition C appears to be minimally more negative relative to A and B in the N400 time window. Though, if there is any predicted N400 effect in C vs. A/B, its magnitude is considerably underestimated, as can be observed from the large residuals for condition C in this time window (Figure 3.7, right). The P600 effect in B (relative to A) is not predicted by LSTM surprisal, as there is no visually observable difference between conditions in the models' estimates and large residuals for condition B in

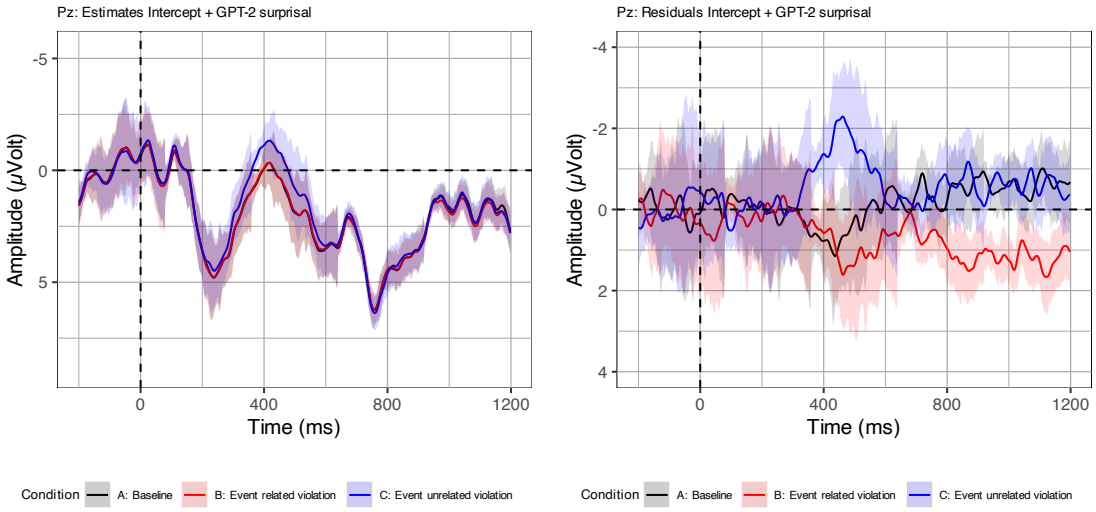

Figure 3.4.: **DBC19**, **GPT-2**: estimated voltages (left) and residuals (right) per condition.

the 800-100 ms window. Though the presence of a predicted N400 in condition C is debatable, LSTM surprisal was a significant predictor in the N400 window consistently throughout time across most of the central electrodes, except for F3, Fz and F4 (Figure 3.1). On the other hand, LSTM surprisal was not significant in the P600 window.

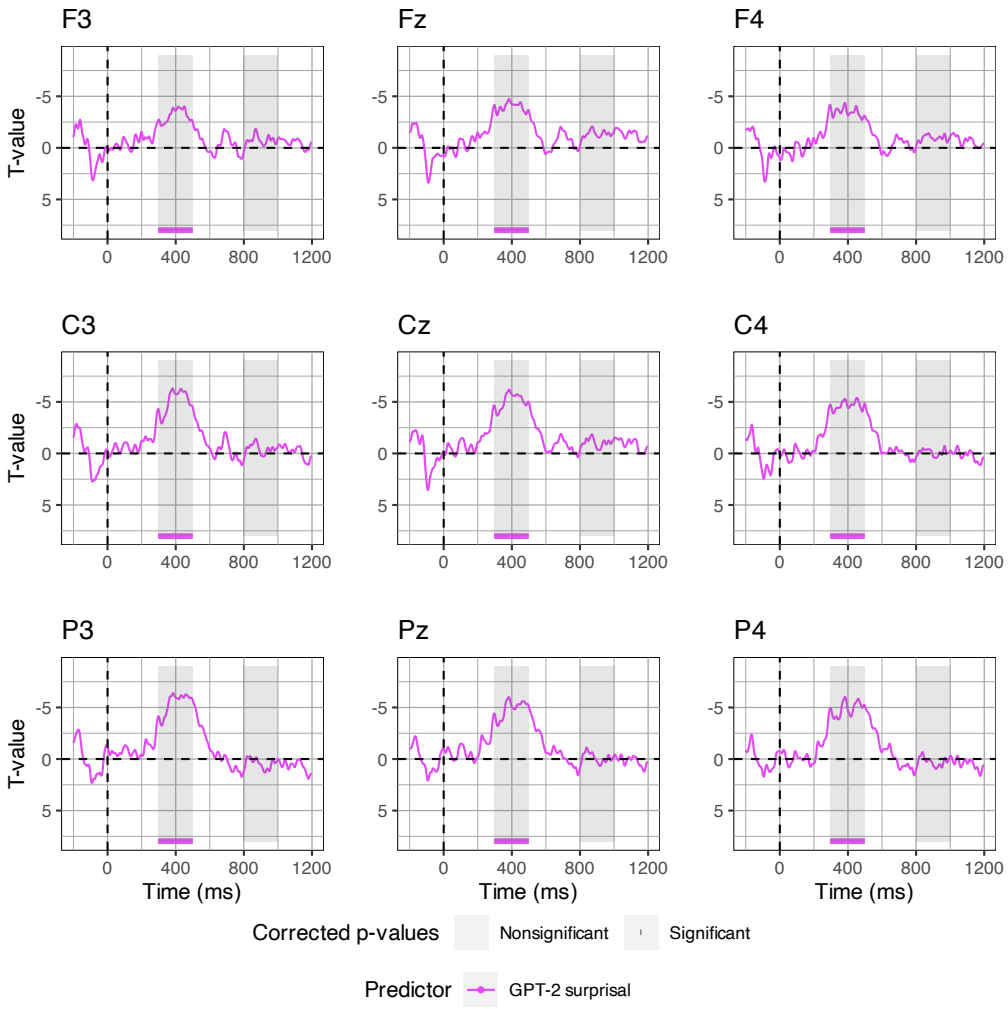

Figure 3.5.: **DBC19**, **GPT-2**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

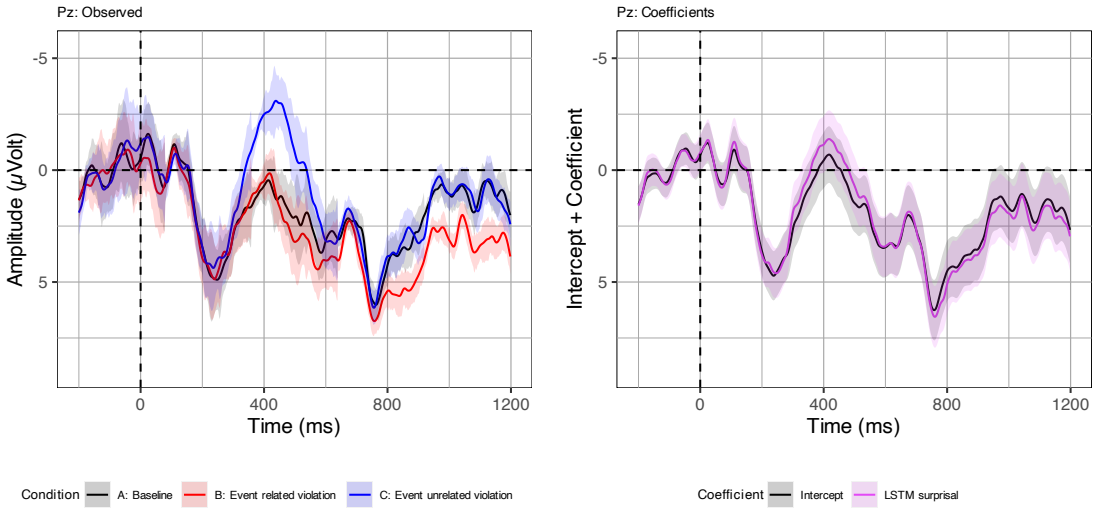

Figure 3.6.: **DBC19**, **LSTM**: observed voltages per condition (left) and the surprisal coefficient over time (right).

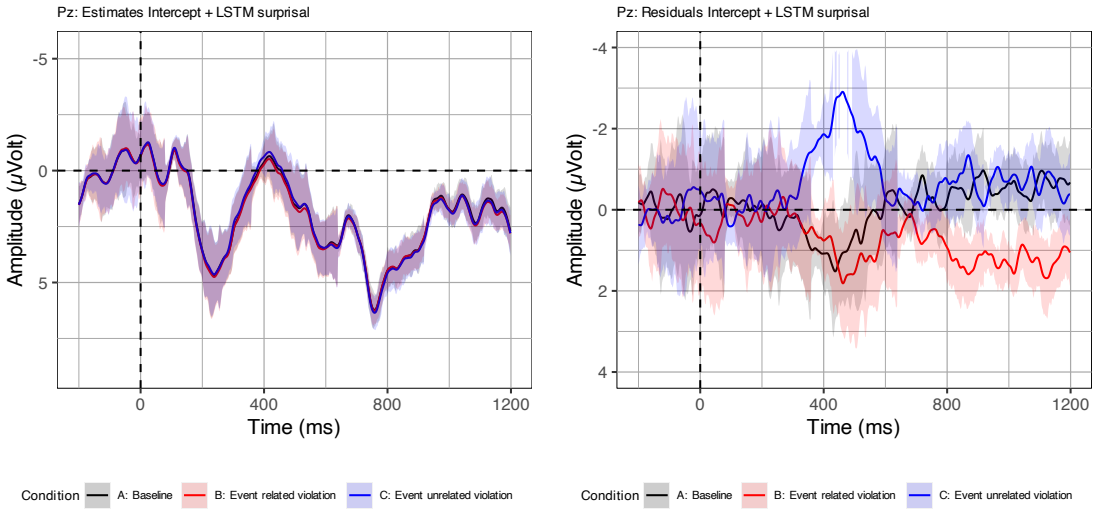

Figure 3.7.: **DBC19**, **LSTM**: estimated voltages (left) and residuals (right) per condition.

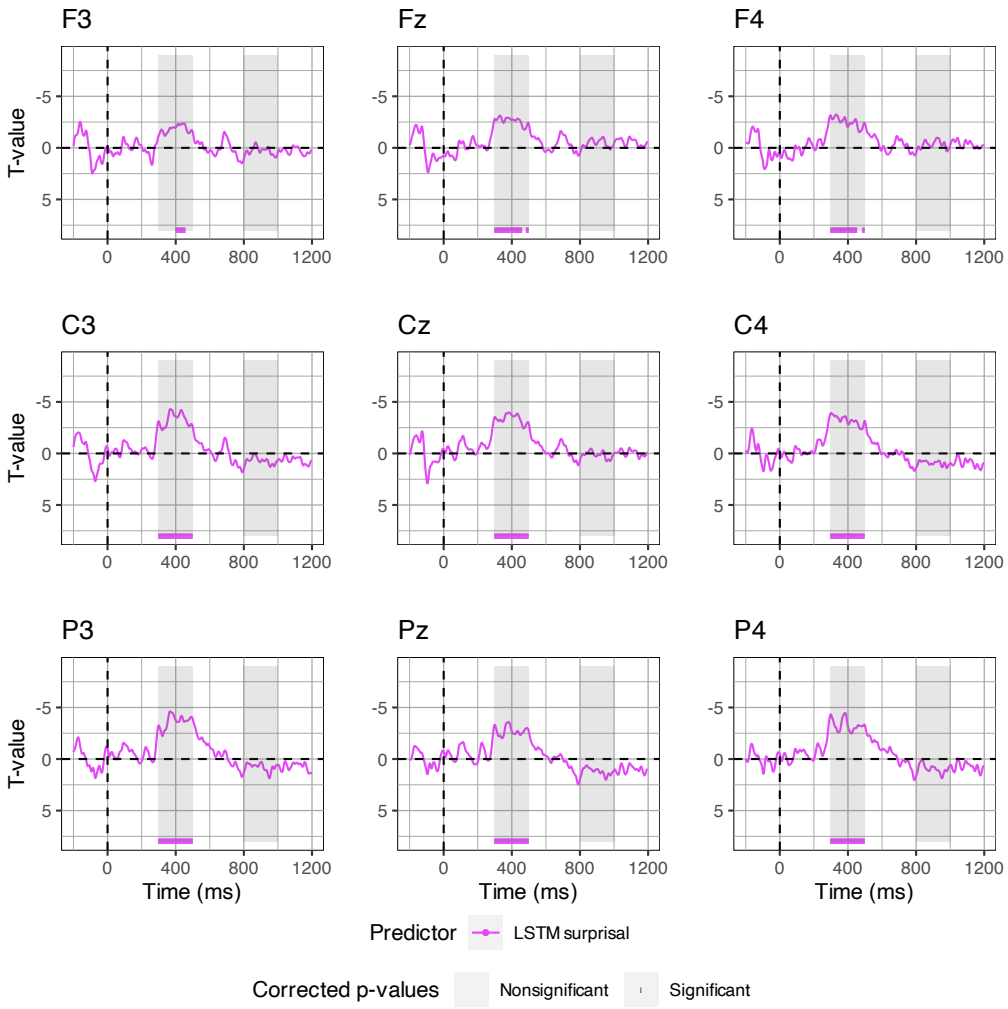

Figure 3.8.: **DBC19**, **LSTM**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

3.2. DBC21

The mean LM surprisal for both models per condition is displayed in Table 3.5. Densities are plotted in figure 3.9.

Overall, mean surprisal of both models patterns with the mean association ratings within conditions (see table 2.3). Conditions B and C share the same prime and target combination with a weak association rating, while A features a different prime with a strong association rating. Analogously, mean LM surprisals appear similar to each other (though not identical) and relatively higher in conditions B and C, compared to a lower and more distinct mean surprisal in A. This holds for both models. Both plausibility ratings and cloze probability on the other hand show a different pattern, with gradually lower ratings for B and C compared to A, which LM mean surprisal does not seem to strongly follow. It is to note though, that mean surprisal in C appears slightly higher compared to B for GPT-2., while for the LSTM reversely mean surprisal in B appears slightly higher than in C.

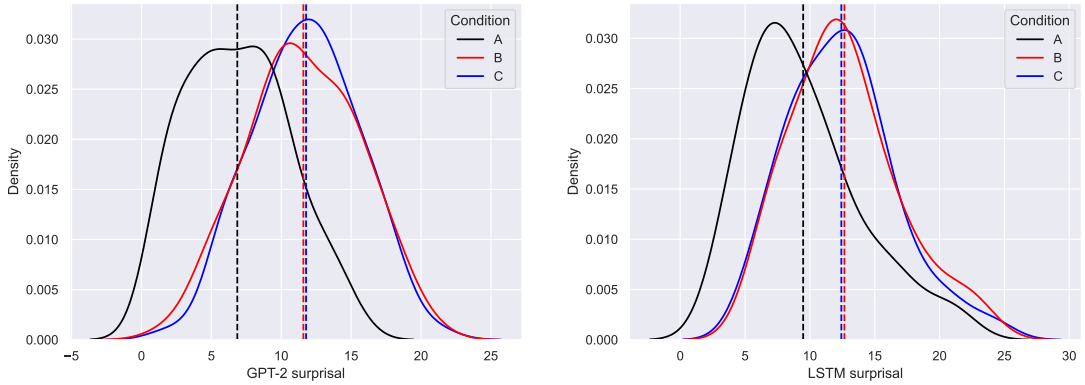

Figure 3.9.: **DBC21**: densities for the surprisal estimates from both the GPT-2 and LSTM models within conditions. Vertical lines indicate condition mean study.

Correlations are reported for respectively each model in Table 3.6. There is a strong positive correlation (0.59) between the surprisal estimates of the language models. For GPT-2, there is a strong negative correlation with association (-0.4), cloze probability (-0.39) and plausibility (-0.31). For the LSTM, there is a moderate negative correlation with association (-0.29), cloze

	GPT-2			**LSTM**		
	Mean	**SD**	**Range**	**Mean**	**SD**	**Range**
A	6.86	3.65	0.79-15.03	9.48	4.43	3.06-21.51
B	11.58	4.00	1.83-20.78	12.66	4.20	5.46-23.78
C	11.78	3.74	2.04-21.18	12.42	4.13	5.23-24.33

Table 3.5.: **DBC21** mean, standard deviations and ranges for the distributions of surprisal estimates for the GPT-2 and LSTM model per condition.

	GPT-2	LSTM	Association	Plausibility	Cloze
GPT-2	1	0.59	-0.4	-0.31	-0.39
LSTM	0.59	1	-0.29	-0.22	-0.28
Association	-0.4	-0.29	1	0.48	0.63
Plausibility	-0.31	-0.22	0.48	1	0.61
Cloze	-0.39	-0.28	0.63	0.61	1

Table 3.6.: **DBC21** Kendall correlation between human ratings and LM surprisal.

	GPT-2			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	10.07	0.22	44.95	<0.001
Association	-1.69	0.47	-3.57	<0.001
Plausibility	-0.02	0.30	-0.08	0.94
Cloze	-0.85	0.45	-1.90	0.06
	R^2=0.31	F(3,266)=40.61, p<0.001		
	LSTM			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	11.52	0.25	45.44	<0.001
Association	-1.04	0.54	-1.94	0.053
Plausibility	0.13	0.34	0.39	0.70
Cloze	-0.81	0.51	-1.59	0.11
	R^2=0.14	F(3,266)=14.96, p<0.001		

Table 3.7.: **DBC21** Linear regression model summary for LM surprisal.

probability (-0.28) and plausibility (-0.22). Similarly to DBC19, both models show the strongest correlation with association, followed by cloze probability and the weakest correlation with plausibility. Overall, correlation strengths are weaker for the LSTM than for the GPT-2 model.

A summary of the linear models of LM surprisal predicted by association, plausibility and cloze ratings is presented in Table 3.7. The pattern of results is very similar to DBC19. Overall, the predictors could explain more variance in GPT-2 surprisal (R^2=0.31) than in LSTM surprisal (R^2=0.14). For GPT-2, all predictors were negatively associated with surprisal, but only semantic association was significant (t=-3.57, p<0.001). Still, cloze (t = -1.90, p=0.06) was a stronger predictor than plausibility (t=-0.08, p=0.94). For LSTM surprisal, no predictor was significant, though, association (t=-1.94, p=0.053) was negatively associated more strongly than cloze (t=-1.59, p=0.11), while the influence of plausibility was weakest (t=0.39, p=0.70).

To recap, DBC21 observed a negativity for conditions B and C (both weak association) relative to A (strong association) in a time window of 300-500 ms, and a positivity of condition C (implausible) relative to B (plausible) in the 800-1000 ms time window. Another late positivity of condition C relative to A was not directly observable due to component overlap but was revealed in an rERP analysis.

Considering the patterns of mean values for association, plausibility, cloze and LM surprisal

per condition in conjunction with the overall correlations and results of the linear regression models leads to the following expectations towards the outcome of the rERP analysis.

N400: LM surprisal of both models will predict the N400 effect in conditions B and C (both unrelated) relative to A (related).

P600: LM surprisal of both models will not predict the P600 effect in C (implausible) relative to B (plausible).

Next, the rERP results for both models are presented. The observed ERP is displayed in both Figure 3.10 and Figure 3.13 on the left side.

GPT-2 surprisal. Relative to the intercept, the surprisal coefficient evolves negatively in the 350-500 ms window but not positively in the 600-1000 ms window (Figure 3.10). Inspecting the models' forward solutions (Figure 3.11, left), surprisal appears to predict the N400 in conditions B and C (relative to A). Though, the magnitude of the N400 effect appears to be underestimated specifically for condition B, as can be observed from the large residuals(Figure 3.11, right). The P600 effect in C (relative to B) appears to be not predicted by GPT-2 surprisal, as there is no visually observable difference between conditions in the models' estimates and large residuals for condition B in the 800-100 ms window. As can be seen in Figure 3.2, GPT-2 surprisal was a significant predictor in the N400 window across electrodes (except for some time samples on Pz and P3), but never significant in the P600 window across all 9 central electrodes.

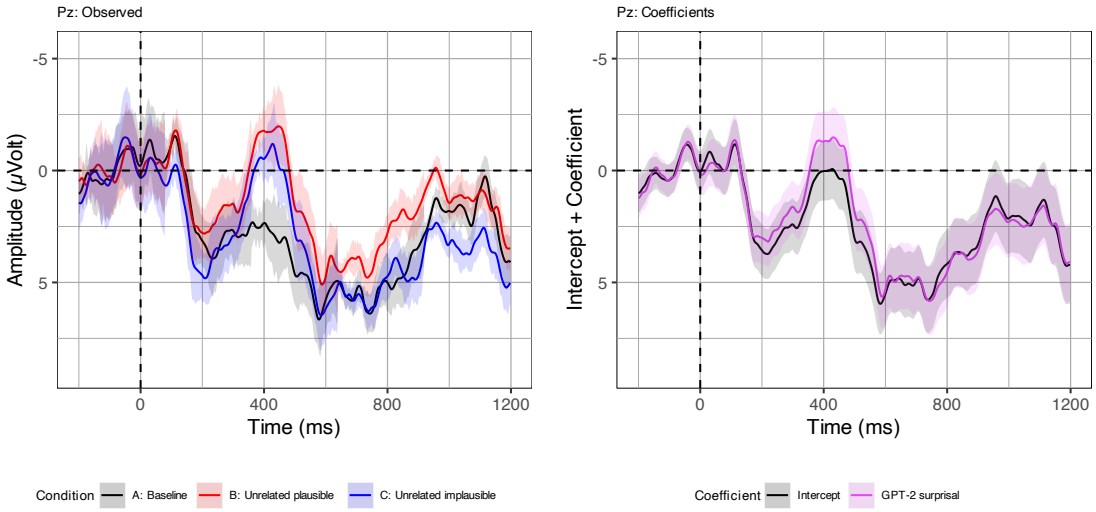

Figure 3.10.: **DBC21**, **GPT-2**: observed voltages per condition (left) and the surprisal coefficient over time (right).

LSTM surprisal. Relative to the intercept, the surprisal coefficient evolves slightly negatively in the 300-500 ms window but only minimally positively in the 600-1000 ms window (Figure 3.13). Inspecting the models' forward solutions (Figure 3.14, left), surprisal can predict the N400 effect in B and C relative to A. Though, the magnitude of the effect appears to be underestimated specifically for condition C, as can be seen in the larger residuals (Figure 3.14, right). The P600 effect in C relative to B is not predicted by LSTM surprisal, as there is no

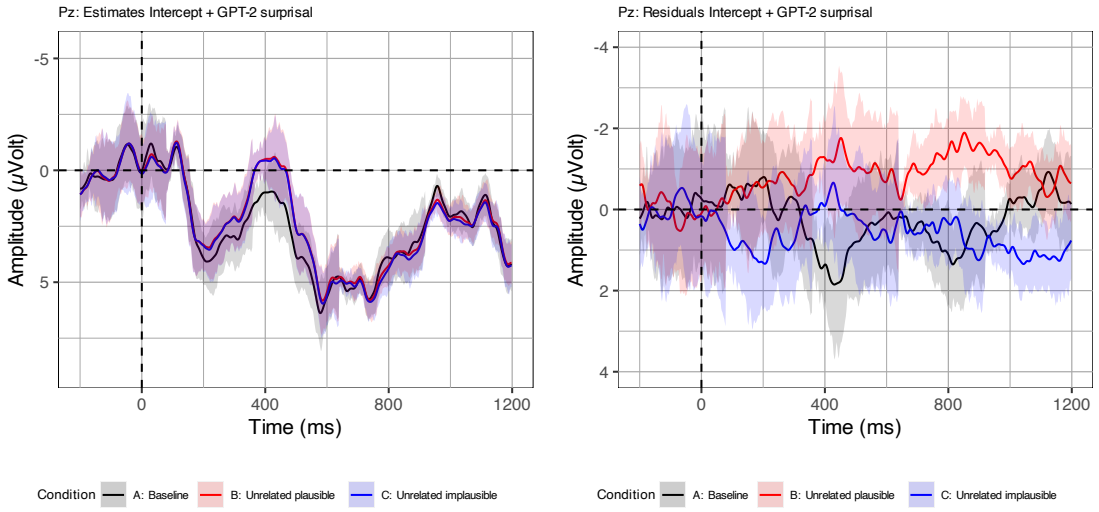

Figure 3.11.: **DBC21**, **GPT-2**: estimated voltages (left) and residuals (right) per condition.

visually observable difference between conditions in the models' estimates and large residuals for both conditions in the 800-1000 ms window. Analogue to GPT-2 surprisal, LSTM surprisal was a significant predictor in the N400 across most of the central electrodes, except for some time samples on Pz and P3 (Figure 3.2). On the other hand, LSTM surprisal was not significant in the P600 window.

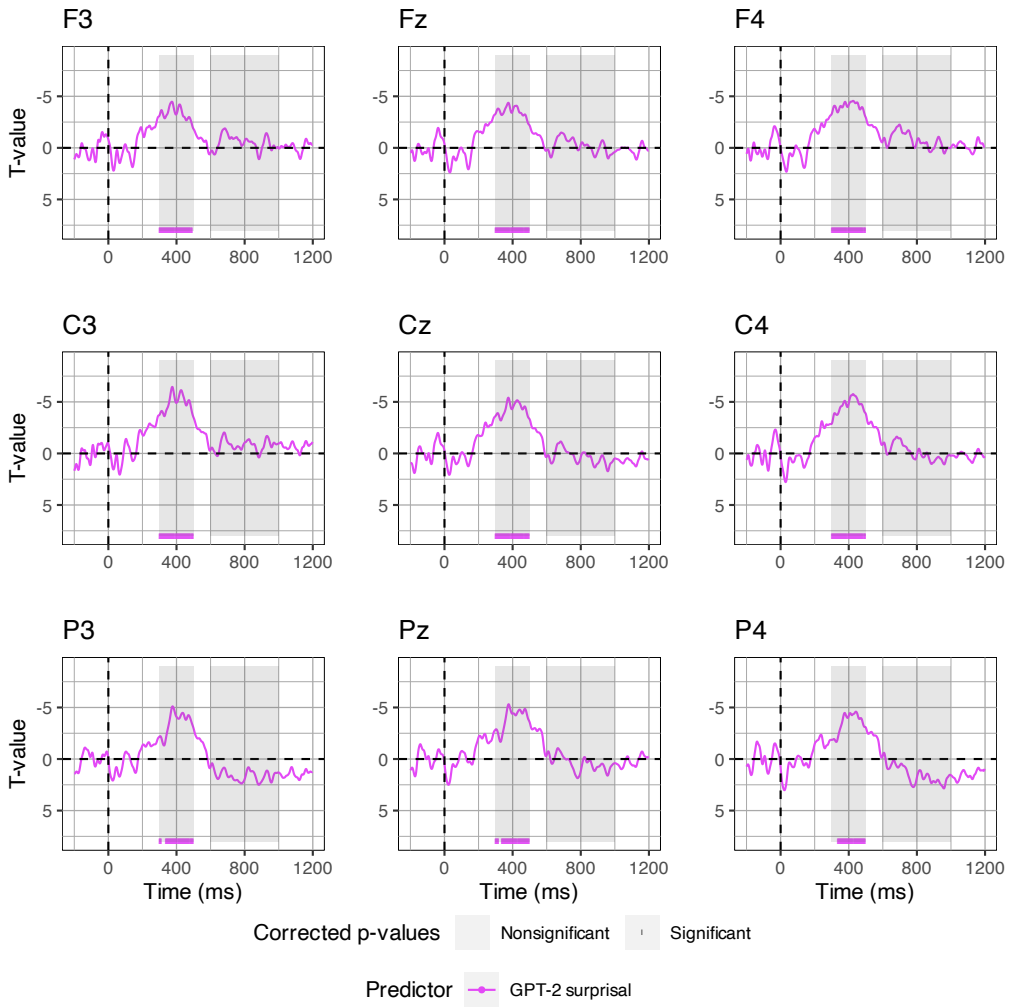

Figure 3.12.: **DBC21**, **GPT-2**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

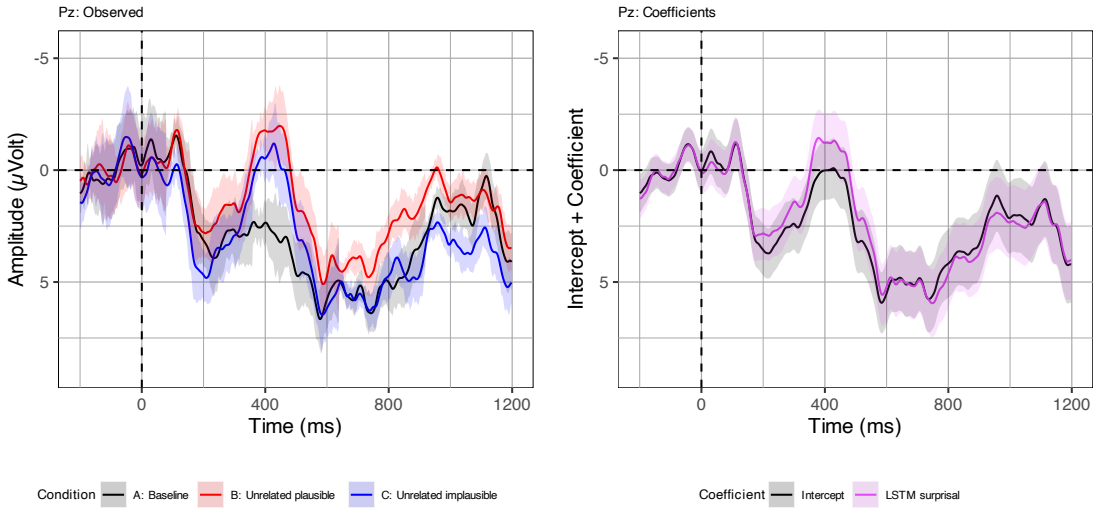

Figure 3.13.: **DBC21**, **LSTM**: observed voltages per condition (left) and the surprisal coefficient over time (right).

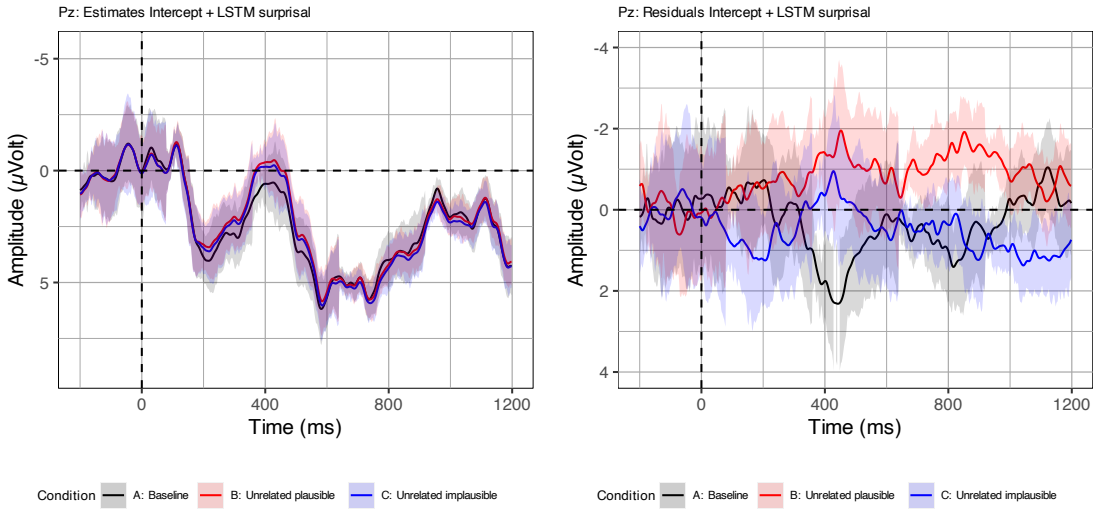

Figure 3.14.: **DBC21**, **LSTM**: estimated voltages (left) and residuals (right) per condition.

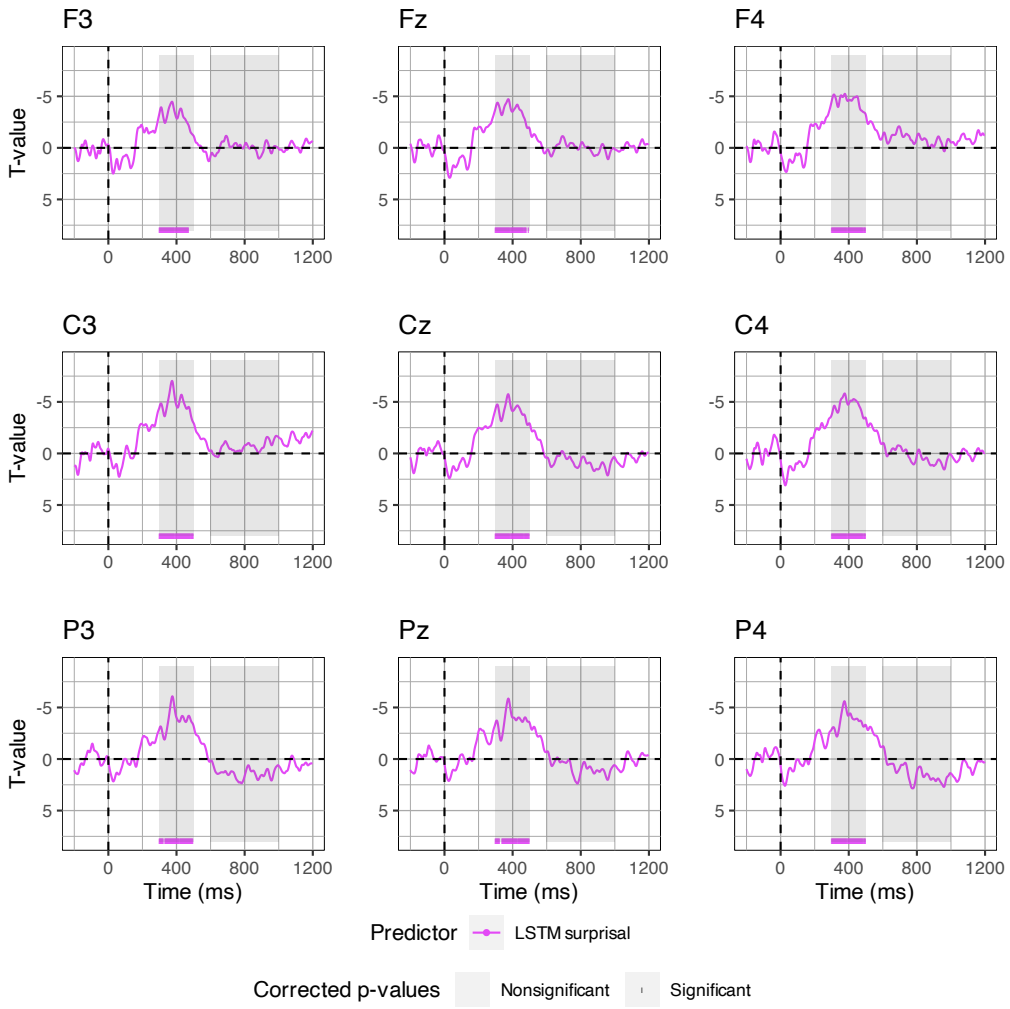

Figure 3.15.: **DBC21**, **LSTM**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

3.3. ADSBC21

The mean LM surprisal for both models per condition is displayed in table 3.8. Densities are plotted in figure 3.16.

In contrast to DBC19 and DBC21, the GPT-2 and LSTM model show a different pattern with respect to surprisal means (see table 2.4). Conditions B and C have both gradually lower plausibility ratings and cloze probabilities relative to A. For GPT-2, mean surprisal appears higher in B and C compared to A, but there seems to be almost no difference between B and C, that is, the gradedness in B vs. C is missing. LSTM mean surprisal doesn't appear to exhibit any notable difference between conditions.

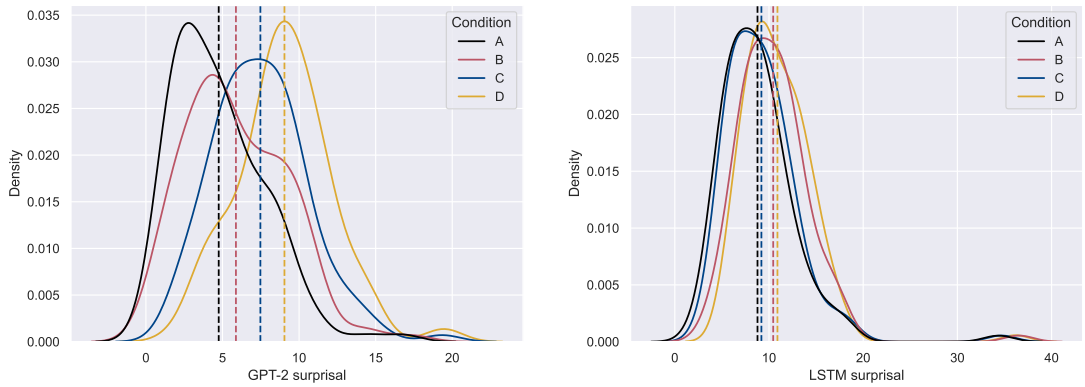

Figure 3.16.: **ADSBC21**: densities for the surprisal estimates from both the GPT-2 and LSTM models within conditions. Vertical lines indicate condition mean.

Correlations are reported for respectively each model in table 3.9. Since ADSBC21 found the association difference between conditions A & C (strong) vs. B & D (weak) to be stronger for Noun-target than Verb-target, only the former is considered here. There is a strong positive correlation (0.46) between the surprisal estimates of the language models. For GPT-2, there is a weak negative correlation with association (-0.18) and a strong negative correlation with cloze probability (-0.31). For the LSTM, there is a moderate negative correlation with association (-0.22) and only a very weak negative correlation with cloze probability (-0.07). Thus, while for GPT-2 surprisal the correlation with cloze is relatively stronger than with association, the

	GPT-2			**LSTM**		
	Mean	**SD**	**Range**	**Mean**	**SD**	**Range**
A	4.75	3.00	0.18-16.78	8.78	4.03	2.13-34.37
B	5.88	3.29	0.23-17.35	10.45	4.00	3.34-36.53
C	7.47	3.07	1.56-19.35	9.19	3.97	3.40-34.69
D	9.04	3.12	1.88-19.71	10.90	3.78	5.49-36.16

Table 3.8.: **ADSBC21** mean, standard deviations and ranges for the distributions of surprisal estimates for the GPT-2 and LSTM model per condition.

	GPT-2	LSTM	Association	Cloze
GPT-2	1	0.46	-0.18	-0.31
LSTM	0.46	1	-0.22	-0.07
Association	-0.18	-0.22	1	0.02
Cloze	-0.31	-0.07	0.02	1

Table 3.9.: **ADSBC21** Kendall correlation between human ratings and LM surprisal. Note that association refers to association between the target and the noun of the adverbial clause.

	GPT-2			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	6.79	0.14	47.67	<0.001
Association	-0.74	0.14	-5.20	<0.001
Cloze	-1.42	0.14	-9.93	<0.001
	R^2=0.21	F(2,477)=64.36, p<0.001		
	LSTM			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	9.83	0.19	55.18	<0.001
Association	-0.97	0.18	-5.46	<0.001
Cloze	-0.31	0.18	-1.74	0.08
	R^2=0.07	F(2,477)=40.61, p<0.001		

Table 3.10.: **ADSBC21** Linear regression model summary for LM surprisal.

reverse is true for the LSTM.

A summary of the linear models of LM surprisal predicted by association and cloze is presented in Table 3.10. Overall, the predictors could explain more variance in GPT-2 surprisal (R^2=0.21) than in LSTM surprisal (R^2=0.07). For GPT-2, both predictors were significant(p<0.001) and negatively associated with surprisal, with cloze (t=-9.93) having a stronger influence than association (t=-5.20). For LSTM surprisal on the other hand, only association was significant (t=-5.46, p<0.001) while cloze was not (t=-1.74, p=0.08).

To recap, in the N400 time window (350ms-450ms), ADSBC21 observed an increased negativity for the less associated conditions, B relative to A and D relative to C, and for the lower cloze conditions, C relative to A and D relative to B (see Figure 3.17 left). In the P600 window, the low cloze conditions C and D elicited a larger positivity compared to the high cloze conditions, while semantic association appeared to have no influence.

Considering the patterns of mean values for association, plausibility, cloze and LM surprisal per condition in conjunction with the overall correlations and results of the linear regression models leads to the following expectations towards the outcome of the rERP analysis.

N400: GPT-2 surprisal may predict all N400 effects. LSTM surprisal may predict the association-driven N400 effects in B versus A and D versus C, but it won't predict the cloze-driven effects in C versus A and D versus B.

P600: GPT-2 surprisal may predict the P600 effect in C/D relative to A/B. LSTM surprisal will not predict the P600 effect.

Next, the rERP results for both models are presented. The observed ERP is displayed in both Figure 3.17 and Figure 3.20 on the left side.

GPT-2 surprisal. Relative to the intercept, the surprisal coefficient evolves negatively in the 350-450 ms window and also slightly positively in the 600-800 ms window (Figure 3.17, right). Inspecting the models' forward solutions (Figure 3.18, left), surprisal can predict all of the N400 effects: B vs. A, D vs. C, C vs. A and D vs. B. Though, the magnitudes of the effects are underestimated, as can be observed from the residuals (Figure 3.18, right). If the P600 effect in conditions C/D relative to A/B is predicted by GPT-2 surprisal is more difficult to visually assess, though it appears to be present. As can be seen in Figure 3.3, GPT-2 surprisal was a significant predictor in the N400 window across all electrodes. On electrodes Pz, P3 and P4 it was also partially significant in the P600 window.

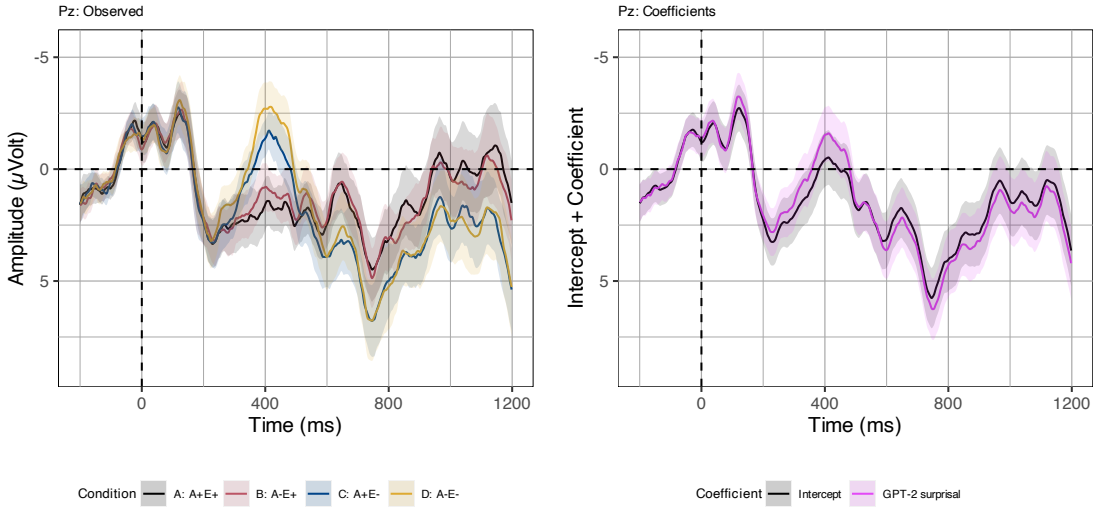

Figure 3.17.: **ADSBC21**, **GPT-2**: observed voltages per condition (left) and the surprisal coefficient over time (right).

LSTM surprisal. Relative to the intercept, the surprisal coefficient evolves slightly negatively in the 350-500 ms window but only minimally positively in the 600-800 ms window (Figure 3.20). Inspecting the models' forward solutions (Figure 3.21, left), there is hardly any observable difference between conditions across time and overall large residuals (Figure 3.21, right). As shown in Figure 3.3, LSTM surprisal was not significant in either the N400 or P600 window across time and electrodes.

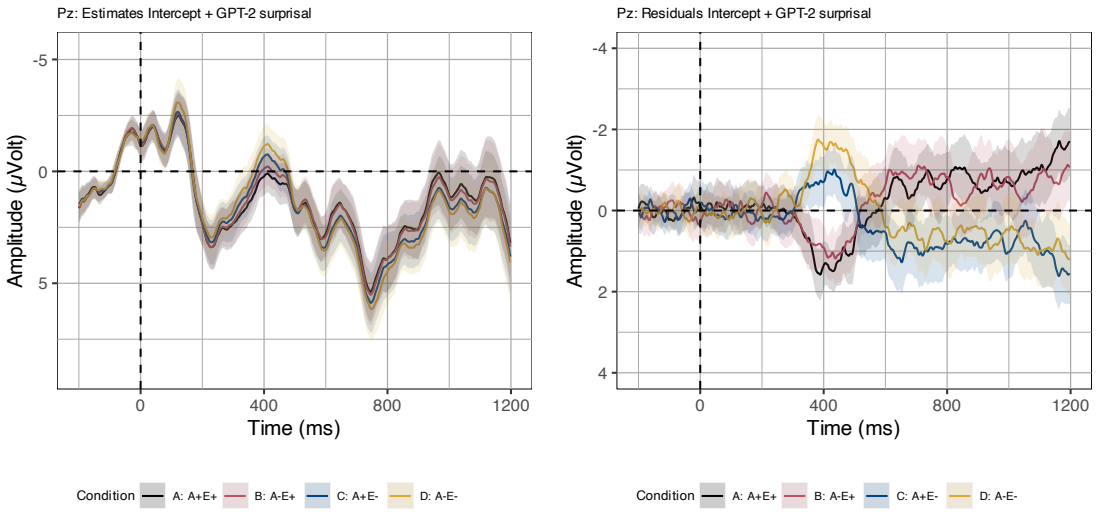

Figure 3.18.: **ADSBC21**, **GPT-2**: estimated voltages (left) and residuals (right) per condition.

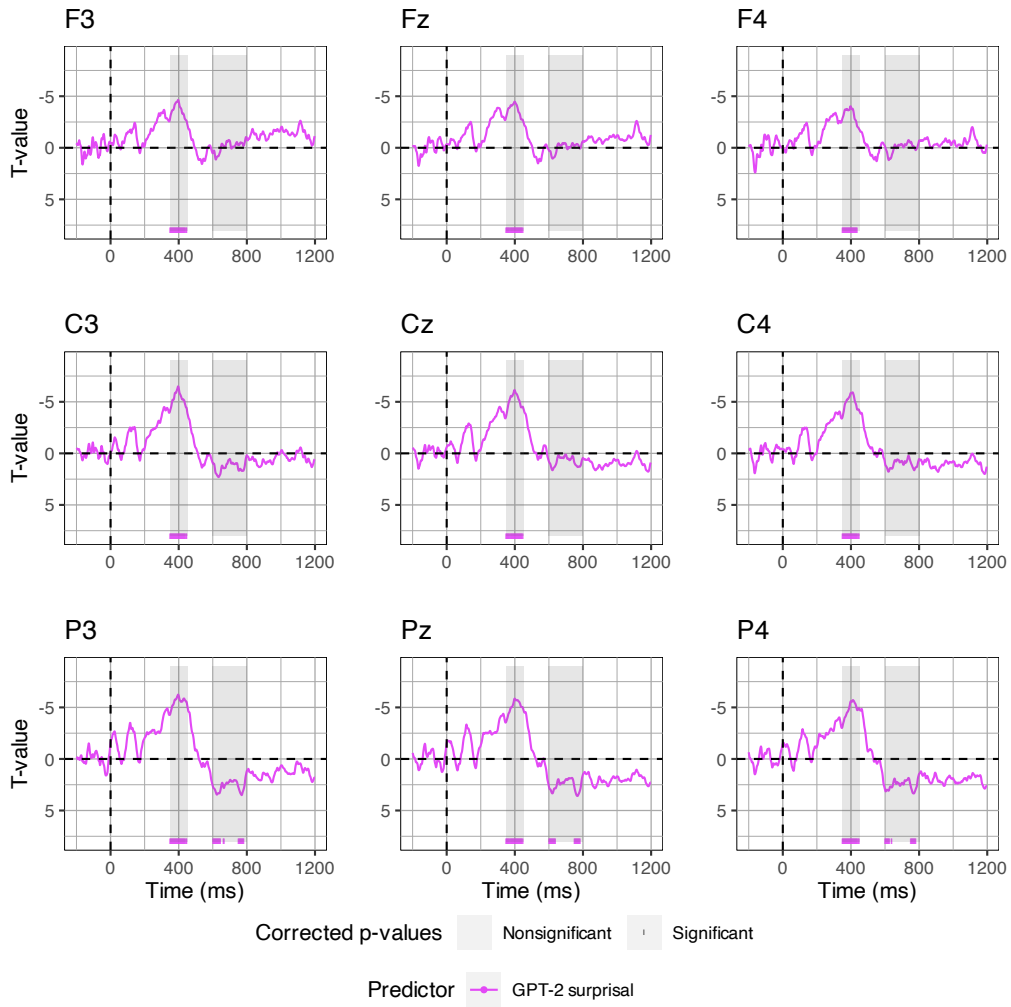

Figure 3.19.: **ADSBC21**, **GPT-2**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

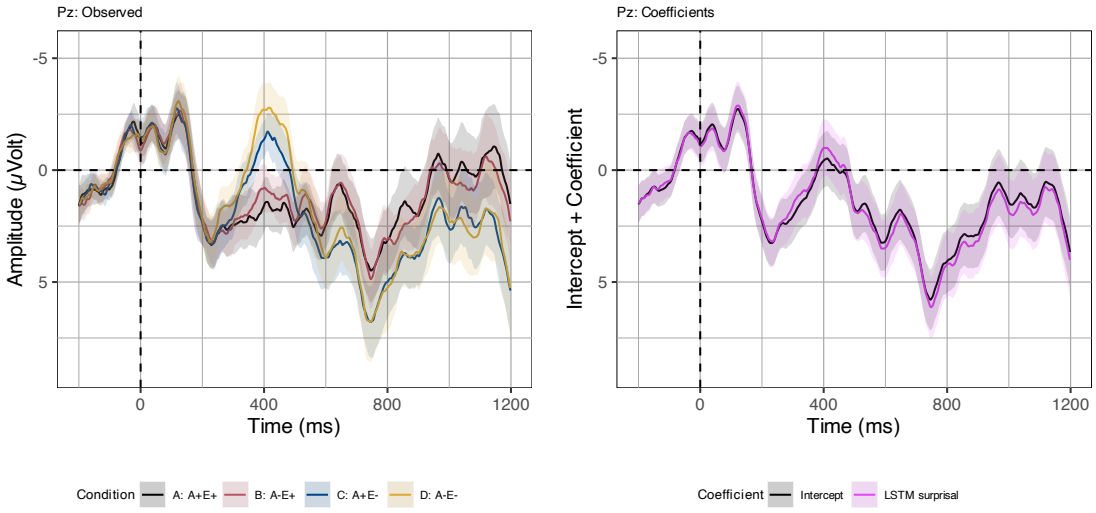

Figure 3.20.: **ADSBC21**, **LSTM**: observed voltages per condition (left) and the surprisal coefficient over time (right).

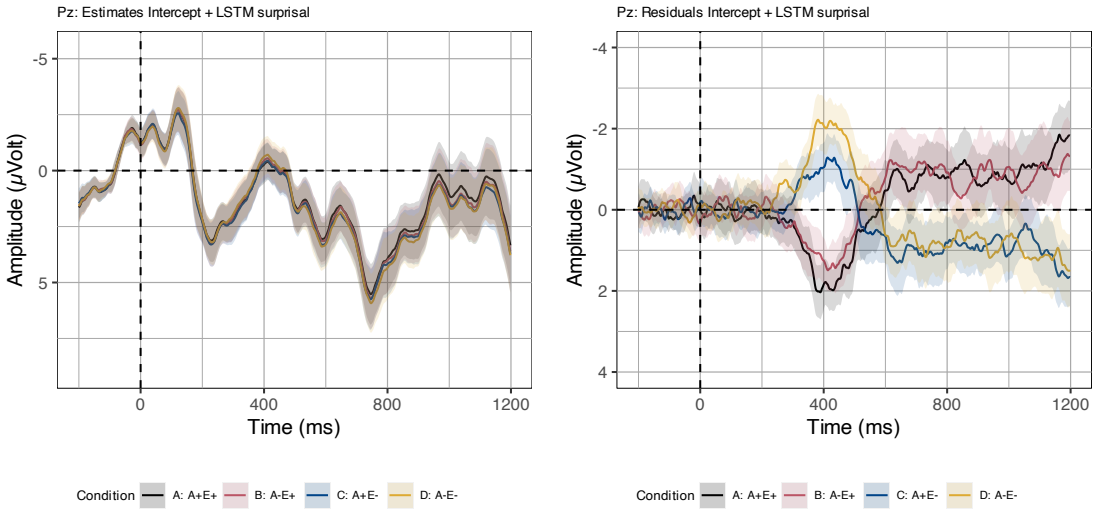

Figure 3.21.: **ADSBC21**, **LSTM**: estimated voltages (left) and residuals (right) per condition.

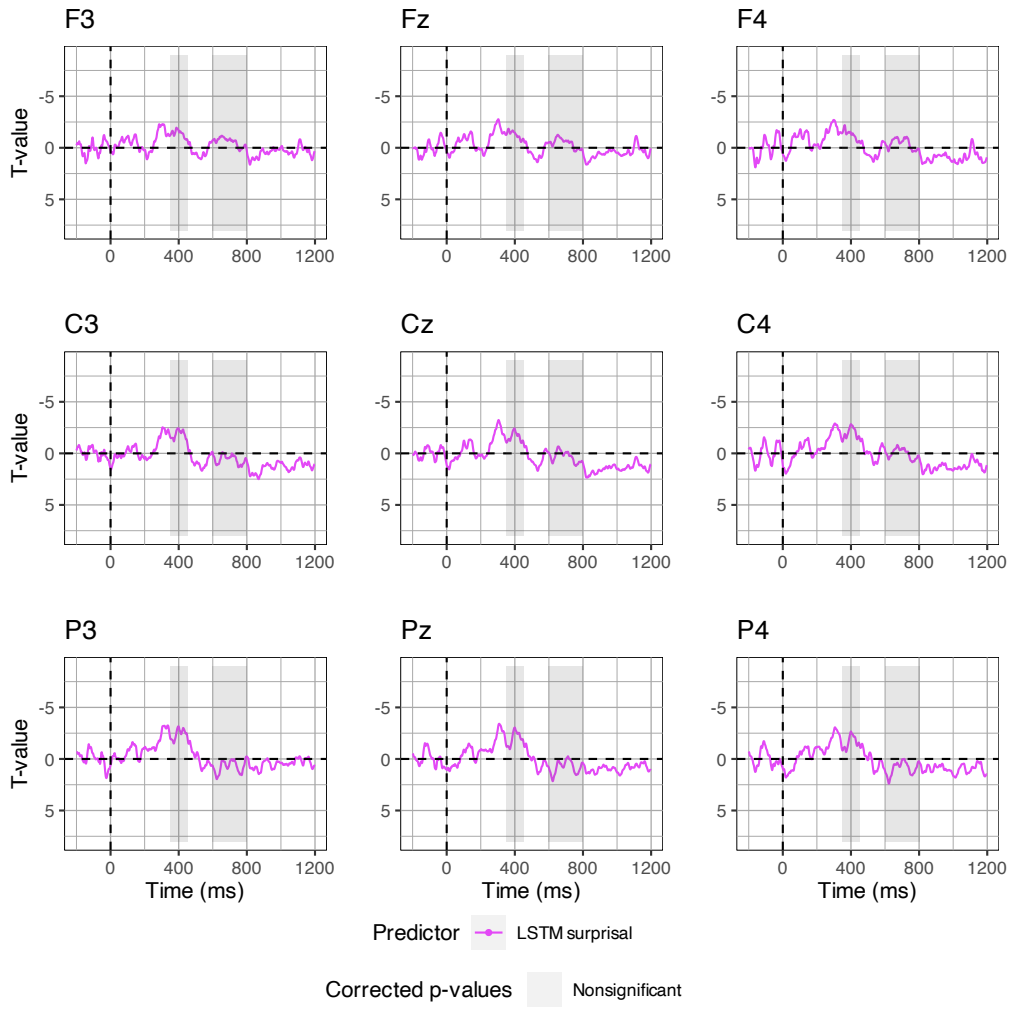

Figure 3.22.: **ADSBC21**, **LSTM**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

3.4. ADBC23

The mean LM surprisal for both models per condition is displayed in table 3.11. Densities are plotted in figure 3.23.

As for ADSBC21, the GPT-2 and LSTM model show a different pattern with respect to surprisal means (see table 2.5). Conditions A and C have a high semantic association rating (they feature the same adverbial clause) for the target, but the target's cloze probability is high in A and low in C. Conditions B and D on the other hand have both a low semantic association rating but cloze probability is high in B and low in C. For GPT-2 mean surprisal is graded in ascending order from A (lowest surprisal) to D (highest surprisal). While LSTM mean surprisal is also lowest in A and highest in D, the low association conditions (B and D) show a higher mean surprisal than the high association conditions (A and C). Hence, while for GPT-2 surprisal there seems to be no clear pattern leaning towards either cloze probability or association, LSTM surprisal seems to pattern with association. It should be noted though, that differences between conditions are less pronounced for LSTM surprisal.

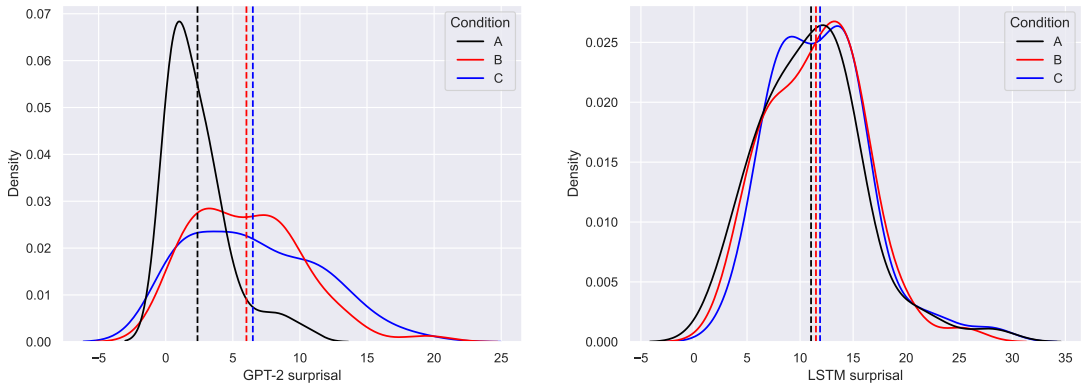

Figure 3.23.: **ADBC23**: densities for the surprisal estimates from both the GPT-2 and LSTM models within conditions. Vertical lines indicate condition mean.

Correlations are reported for respectively each model in table 3.12. In contrast to the other RI-studies, there is only a weak positive correlation (0.12) between the surprisal estimates of the language models. For GPT-2, there is a moderate negative correlation with plausibility

	GPT-2			**LSTM**		
	Mean	**SD**	**Range**	**Mean**	**SD**	**Range**
A	2.37	2.33	0.07-10.53	11.03	4.90	2.33-28.10
B	6.01	4.02	0.09-19.42	11.50	4.49	2.84-25.39
C	6.49	4.74	0.14-18.71	11.88	4.62	3.47-28.10

Table 3.11.: **ADBC23** mean, standard deviations and ranges for the distributions of surprisal estimates for the GPT-2 and LSTM model per condition.

	GPT-2	LSTM	Plausibility	Cloze
GPT-2	1	0.12	-0.27	-0.39
LSTM	0.12	1	-0.06	-0.09
Plausibility	-0.27	-0.06	1	0.63
Cloze	-0.39	-0.09	0.63	1

Table 3.12.: **ADBC23** Kendall correlation between human ratings and LM surprisal.

	GPT-2			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	4.96	0.28	17.63	<0.001
Plausibility	-0.44	0.46	-0.97	0.33
Cloze	-1.60	0.46	-3.52	0.001
	R^2=0.22	F(2,177)=24.47, p<0.001		
	LSTM			
	Estimate	**SE**	**t-value**	**p-value**
Intercept	11.47	0.35	32.98	<0.001
Plausibility	-0.47	0.56	-0.84	0.40
Cloze	0.04	0.56	0.07	0.94
	R^2=0.01	F(2,177)=0.81, p=0.45		

Table 3.13.: **ADBC23** Linear regression model summary for LM surprisal.

(-0.27) and a strong negative correlation with cloze probability (-0.39). For the LSTM, there is only a very weak negative correlation with both plausibility (-0.06) and cloze probability (-0.09). Thus, for both models the negative correlation with cloze probability is stronger than with plausibility.

A summary of the linear models of LM surprisal predicted by plausibility and cloze is presented in table 3.13. Overall, the predictors could explain more variance in GPT-2 surprisal (R^2=0.22) than in LSTM surprisal (R^2=0.01). It has to be noted that, as in DBC19, the F-statistic of the model for LSTM surprisal was not significant (F(2,177)=0.81, p=0.45). For GPT-2, cloze was significant (t=-3.52, p=0.001) while plausibility was not (t=-0.97, p=0.33). For LSTM surprisal, no predictor was significant, though the influence of plausibility (t=-0.84, p=0.40) was stronger than the inluence of cloze (t=0.07, p=0.94).

To recap, in the P600 time window (600ms-1000ms), ADBC23 observed an increasingly large effect of plausibility, in B relative to A and C relative to both B and A (see Figure 3.24 left).

Considering the patterns of mean values for association, plausibility, cloze and LM surprisal per condition in conjunction with the overall correlations and results of the linear regression models leads to the following expectations towards the outcome of the rERP analysis.

N400: LM surprisal from both models should not predict an effect here.

P600: GPT-2 surprisal may predict the P600 effect in B and C relative to A but not the effect of C relative to B. LSTM surprisal will not predict any P600 effect.

Next, the rERP results for both models are presented. The observed ERP is displayed in both

Figure 3.24 and Figure 3.27 on the left side.

GPT-2 surprisal. Relative to the intercept, the surprisal coefficient evolves positively in the 600-1000 ms window(Figure 3.24). Inspecting the models' forward solutions (Figure 3.25, left), surprisal can indeed predict the P600 in condition B and C relative to A but not the effect between C and B. This can also be observed from the large residuals for condition C (Figure 3.25, right). As can be seen in Figure 3.4, GPT-2 surprisal was a significant predictor in the P600 window across all 9 central electrodes but not consistently throughout the complete time window.

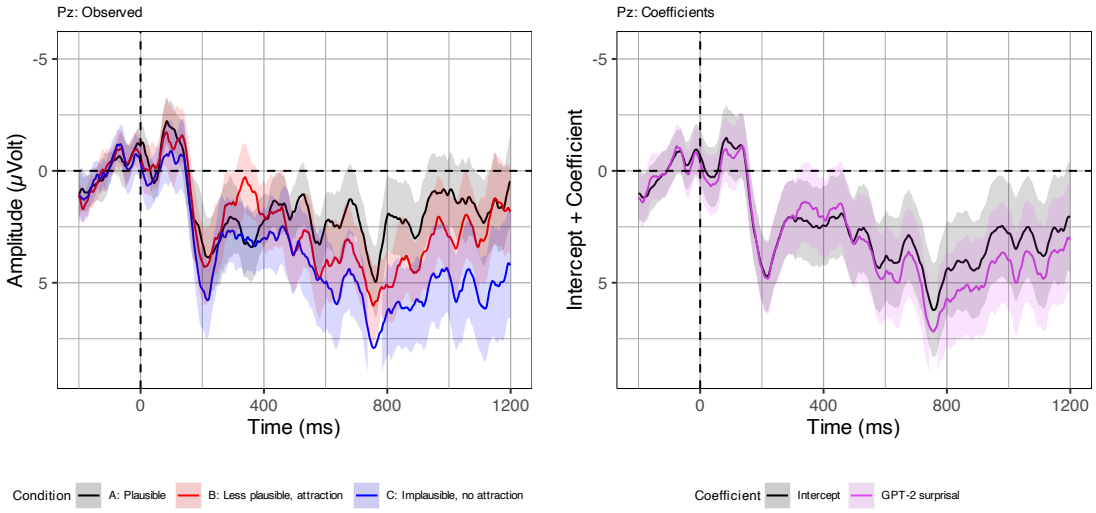

Figure 3.24.: **ADBC23**, **GPT-2**: observed voltages per condition (left) and the surprisal coefficient over time (right).

LSTM surprisal. Relative to the intercept, the surprisal coefficient does not evolve positively in the 600-1000 ms(Figure 3.27). Inspecting the models' forward solutions (Figure 3.28, left), there is no observable difference between conditions across time, and residuals grow increasingly large for condition C (Figure 3.28, right). Figure 3.4 shows that LSTM surprisal was not a significant predictor in the P600 window at any electrode or time sample.

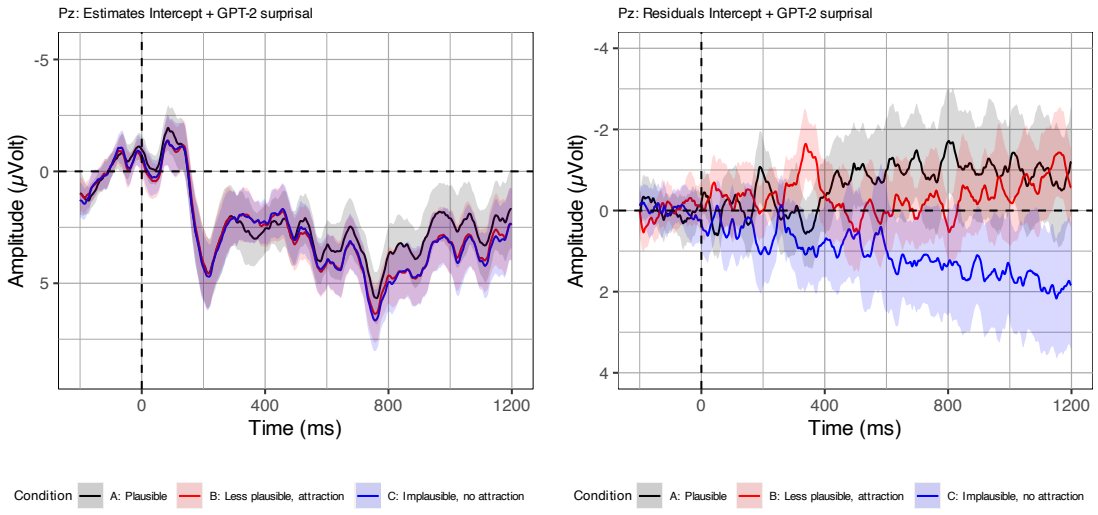

Figure 3.25.: **ADBC23**, **GPT-2**: estimated voltages (left) and residuals (right) per condition.

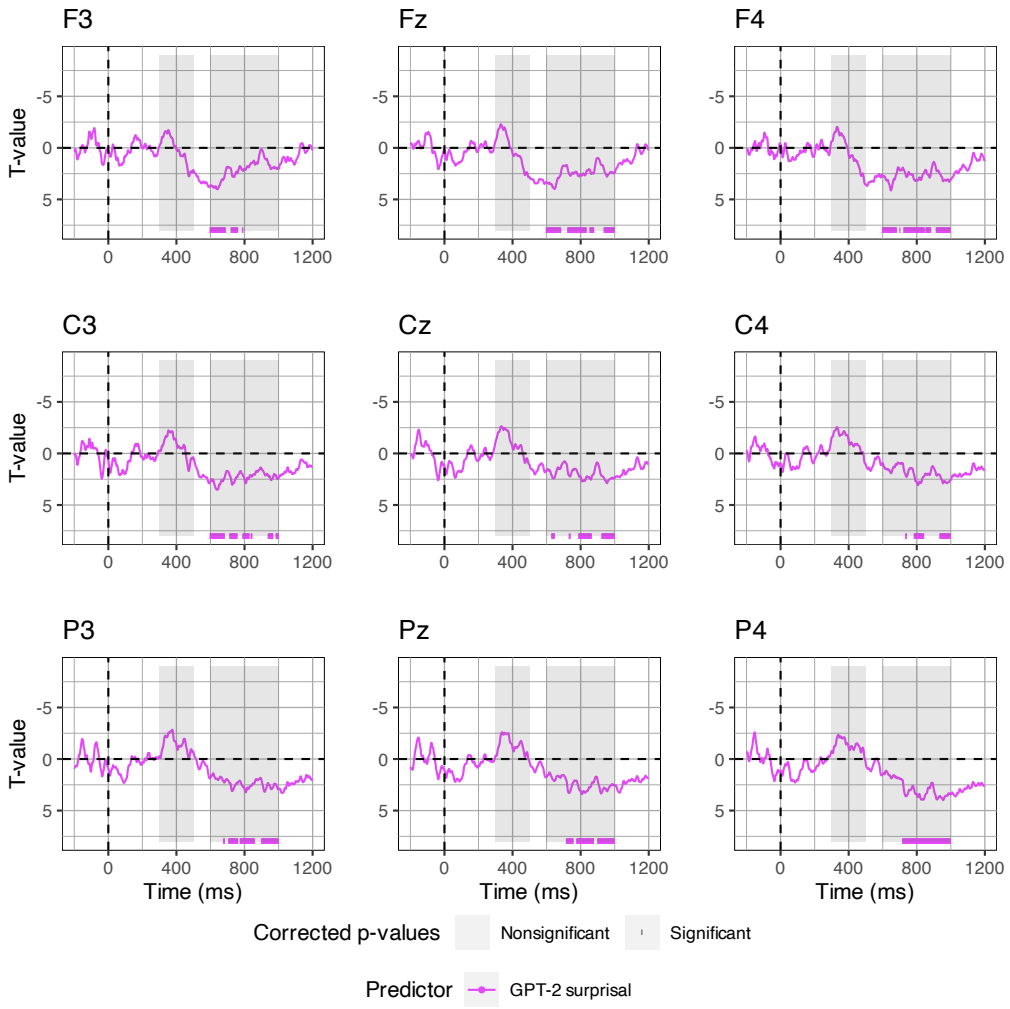

Figure 3.26.: **ADBC23**, **GPT-2**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

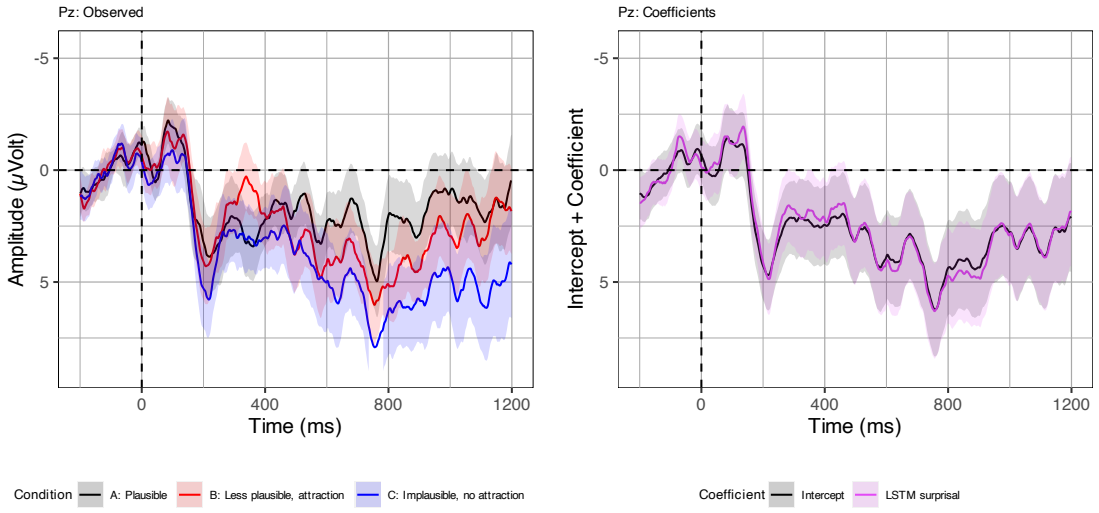

Figure 3.27.: **ADBC23**, **LSTM**: observed voltages per condition (left) and the surprisal coefficient over time (right).

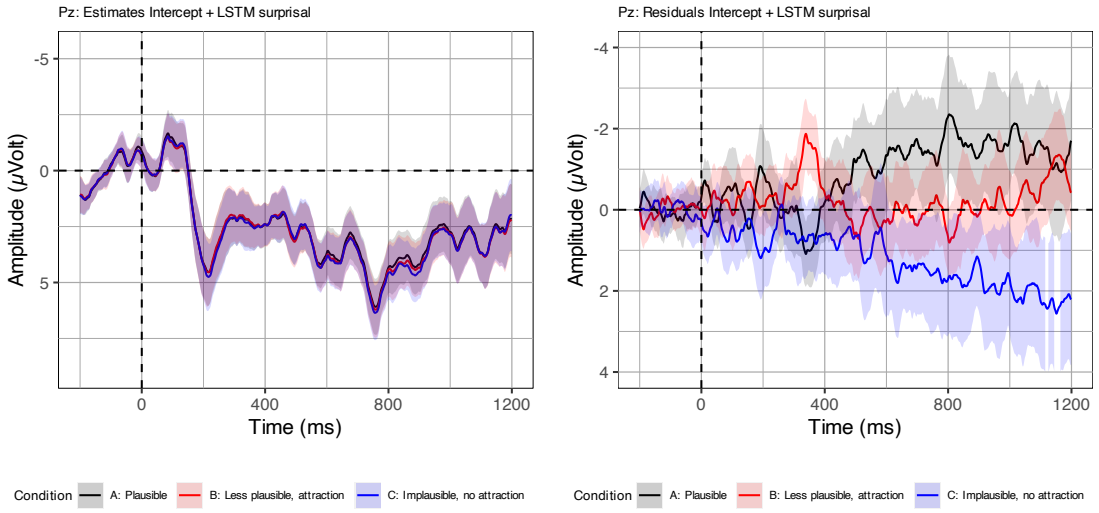

Figure 3.28.: **ADBC23**, **LSTM**: estimated voltages (left) and residuals (right) per condition.

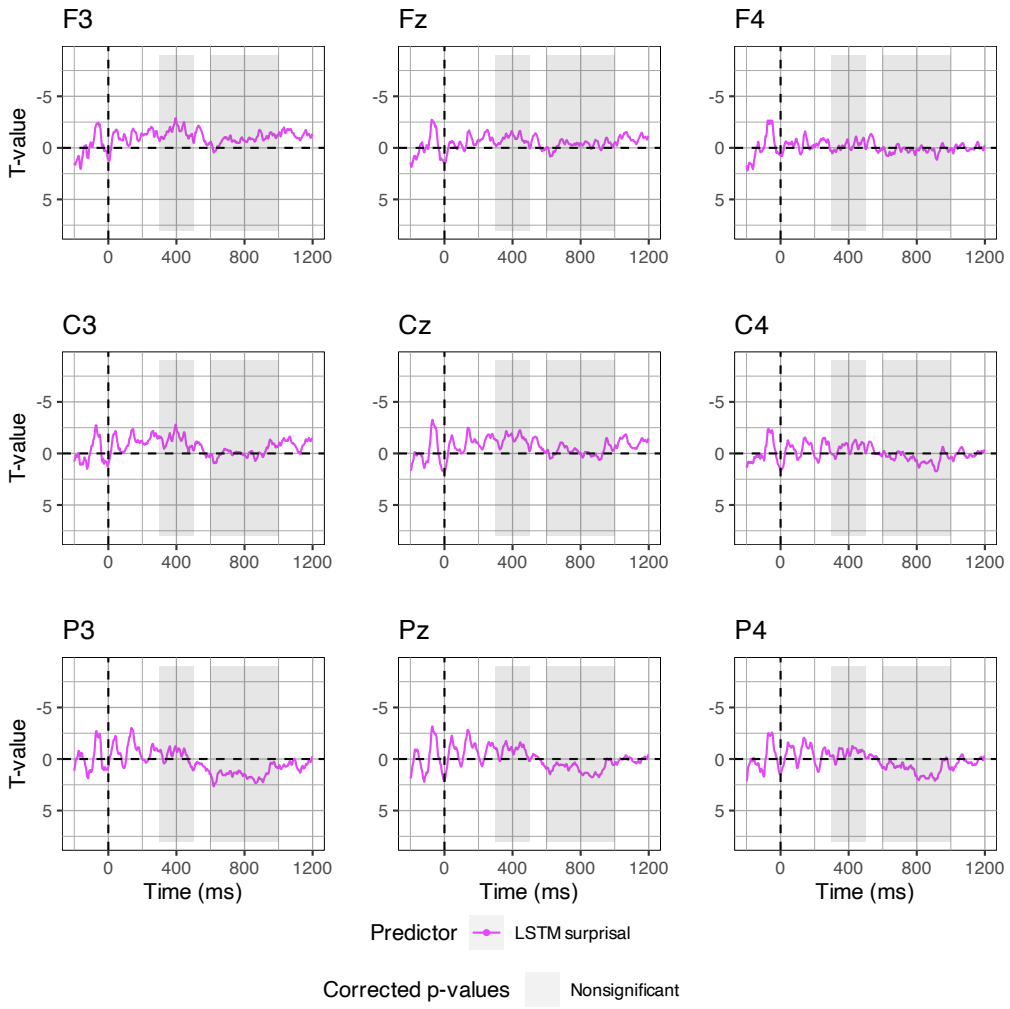

Figure 3.29.: **ADBC23**, **LSTM**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

3.5. Summary of Results

When inspecting the surprisal distributions between conditions within the RI-studies and comparing surprisal means with the respective means for ratings for association, plausibility and cloze, some interesting patterns can be observed.

For DBC19 and DBC21, mean surprisal of both language models aligns with mean association. That is, DBC19 features two conditions with high mean association ratings in which mean LM surprisal is low (A & B), and one condition in which mean association is low and LM surprisal is high (C). DBC21 features one condition in which mean association is high and mean LM surprisal is low (A) and two conditions in which mean association is low and mean LM surprisal is high (B & C). At the same time, mean plausibility and cloze lower gradually from A over B to C in both DBC19 and DBC21 while LM surprisal does not show this trend.

In ADSBC21 a similar pattern shows for LSTM surprisal: in conditions A & C, mean association is high and mean LSTM surprisal is low, while in conditions B & D, mean association is low and mean LSTM surprisal is high. Mean cloze probability is high in A & B and low in C & D and though plausibility ratings have not been collected for this study it appears intuitive that they would inform cloze probabilities and show the same pattern here. Hence, LSTM surprisal patterns with mean association but not cloze probability. GPT-2 surprisal on the other hand gradually lowers from A to D. In condition A, where both mean association and cloze are high, GPT-2 mean surprisal is lowest. In condition B, where mean association is low but mean cloze is high, mean GPT-2 surprisal is higher relative to A. In condition C, where mean association is high and mean cloze is low, mean GPT-2 surprisal is higher relative to both A & B. And in condition C, where both mean association and cloze are low, GPT-2 surprisal is highest.

Lastly, in ADBC23 another interesting observation can be made. While mean plausibility and cloze gradually lower from A over B to C, LM surprisal behaves differently for both models. For GPT-2, surprisal is low in A and high in B & C, but with no noticeable difference between B & C. LSTM surprisal on the other hand does not show any noticeable difference across any of the three conditions.

The Kendall correlations between LM surprisal and the human ratings as well as the linear regression models using the ratings to predict LM surprisal confirm these observations for DBC19 and DBC21. For both studies, LM surprisal had the strongest correlation with association which also was the only predictor that was significant for GPT-2 surprisal in both studies. In ADSBC21 on the other hand, GPT-2 surprisal was more strongly correlated with cloze probability than with association, and while both were significant predictors of surprisal, cloze had a stronger influence. LSTM surprisal on the other hand was only moderately correlated with association but not correlated with cloze and only association was a significant predictor of surprisal. In ADBC23, GPT-2 surprisal was correlated with Cloze more strongly than with plausibility and again only cloze was a significant predictor in the linear regression model. LSTM surprisal was not correlated with either cloze or plausibility and no reliable linear regression model could be fitted to predict LSTM surprisal.

In the rERP analysis the two language models show results different to each other. GPT-2 surprisal was able to predict all the N400 effects found by DBC19, DBC21 and ADSBC21. That is, the N400 effects in DBC19, DBC21 and ADSBC21 that were elicited through manipulations of association as well as the N400 effect in ADSBC21 that was elicited through a manipulation

of expectancy (operationalized through cloze probability). Importantly though, the magnitude was underestimated for all effects. Moreover, though not strongly visible, GPT-2 surprisal may have captured to some smaller extent the P600 effect between conditions A & B versus C & D in ADSBC21. It was also able to predict the P600 effect in ADBC23 for conditions B & C relative to A, but not the effect in C relative to B.

In contrast, LSTM surprisal could only predict the N400 effect in DBC21, but none of the other N400 effects, though it was still significant across electrodes for DBC19. It could not predict any of the P600 effects.

3.6. Additional rERP Analyses

To account for the fact that some conditions in DBC19 and DBC21 were subject to component overlap, that is, a sustained negativity in the 400ms window led to spatio-temporally concealing P600 effects, additional rERP analyses were conducted for these studies. For both LMs additional models were fit, respectively featuring association or plausibility as a second predictor alongside surprisal. The key idea for the association models was to capture the sustained negativity by the means of association, which should then better enable surprisal to predict P600 effects. Moreover, when pairing surprisal with plausibility the goal was to see if surprisal would make similar predictions as association across time.

The results were not included into the main analyses but are instead listed in Appendix B. They corroborate the main results by showing, that even when paired with association, surprisal does not capture any of the P600 effects in DBC19 & DBC21. Moreover, when paired with plausibility, surprisal stays a significant predictor for negativity in the N400 window, as association would.

Chapter 4.

Discussion

With the results of the previous chapter available, this chapter will now return to the theoretical considerations and the research questions formulated in chapter 1.

An underlying overall assumption about language comprehension is, that humans build to some extent expectations about upcoming units, such as words, as they incrementally process an input stream of language. As laid out earlier, the information-theoretic notion of surprisal provides a linking hypothesis between these overall expectations about a word in context and the cognitive effort that is required to process it as a listener during online language comprehension. That is, a low probability of word w_{t+1} given the context of previous words $w_{1...t}$ results in high surprisal which is then taken to indicate high processing cost. Crucially, the theory itself does not specify how the underlying probability distribution used for the computation of surprisal comes about. Therefore, a multitude of operationalizations to arrive at probability estimates for words in context are conceivable, which can be either based on predictions by humans or by models. These estimates of word surprisal (or probability) can be used to predict neurobehavioural human processing effort as indicated by ERP-components, specifically the N400 and the P600.

The thesis conducted a closer examination of the predictions that surprisal estimates of two distinct language models would make on the ERP data of four studies providing evidence for the retrieval-integration account. A core finding of three of these studies (DBC19, DBC21 and ADSBC21) was, that not only the overall expectancy of a word but also its semantic relatedness to the prior context can modulate the N400. The first research question built upon this finding:

1. Can language model surprisal capture N400 effects that have been elicited by a manipulation of association but *not* overall expectancy?

If we take LM surprisal to reflect overall expectancy but not pure semantic association, then the answer should be *no*. That is, in cases where expectancy between two conditons was kept constant but association differed and an N400 effect was elicited, LM surprisal should not be able to predict this effect (see table 1.4). Conversely, in cases where expectancy differed while association was kept constant and *no* N400 effect was observed, LM surprisal should falsely predict an N400 effect (see table 1.5).

Moreover, we should observe that LM surprisal patterns with cloze probability as another operationalization of general expectancy. Expectancy, and thus cloze probability should also be strongly influenced by plausibility. That is, under a rational communication approach, plausible continuations should have a higher probability within listeners than such continuations that violate the assumptions made by common world knowledge. ADSBC21 and ADBC23

found that exactly such violations of plausibility, leading to a diminished expectancy for the target word, can lead to P600 effects. Moreover, ADSBC21 established that semantic association had no impact on the P600 and ADBC23 showed that the P600 is continuously graded for plausibilty. This led to the second research question:

2. Does language model surprisal provide a strong predictor of the P600 amplitude?

Again, if LM surprisal reflects the notion of overall expectancy, then the answer should be *yes*. It should behave similarly to cloze probability and plausibility and be able to predict the P600 effects found by ADSBC21 and ADBC23.

The results however do not confirm these theoretical considerations. In fact, GPT-2 surprisal clearly aligned with semantic association but not with cloze probability or plausibility in DBC19 and DBC21. This has already been indicated when comparing mean surprisal with the means of the human-based metrics between conditions and was additionally confirmed by the correlations. GPT-2 surprisal showed the strongest correlation with association, followed by cloze and the weakest correlation with plausibility. Moreover, association was the only significant predictor in the linear regression model for GPT-2 surprisal.

Based on these observations it was already anticipated in chapter 3, that in the rERP analysis LM surprisal might actually predict the opposite of what was assumed from the theoretical background with respect to the research questions. Indeed, GPT-2 surprisal predicted the same association-driven N400 effects but not the plausibility-driven P600 effects found by DBC19 and DBC21.

For the data of ADSBC21 however the pattern becomes less distinguishable. Mean GPT-2 surprisal was graded so that condition **D** (-A,-E) > **C** (+A,-E) > **B** (-A,+E) > **A** (+A,+E). That is, surprisal was highest when the target was both unassociated and unexpected, and it was higher when expectancy was low than when association was low. Moreover, its correlation with cloze probability was stronger than with association and, analogously cloze was a stronger predictor in the linear regression model for GPT-2 surprisal. While GPT-2 surprisal predicted the association-driven N400 effects, it also predicted the N400 effects driven by expectancy and, although only to a smaller extent, the expectancy-driven P600 effect.

Finally, for ADBC23 mean GPT-2 surprisal aligned to a certain extent with the mean plausibility ratings and mean cloze probability in that it was low for the plausible baseline condition and high for the two implausible conditions. But there appeared to be no noticeable difference in mean GPT-2 surprisal between conditions B & C, a difference which was present in mean plausibility and cloze. Moreover, GPT-2 surprisal correlated more strongly with cloze probability than with plausibility and cloze was a more influential predictor in the linear regression model for surprisal. As anticipated from the within condition means, in the rERP GPT-2 surprisal predicted the P600 effect between the implausible conditions and the plausible baseline, but it did not predict the gradedness between the former ones.

To summarize, it appears to be the case that GPT-2 surprisal aligns more with association than with cloze probability or plausibility in DBC19 & DBC21, by predicting all the N400 but none of the P600 effects. In ADSBC21 on the other hand it patterns with both association and cloze probability in predicting all N400 efects and to some degree the P600 effect. And in ADBC23, where association was not manipulated between conditions, it only partially patterns

with cloze and plausibility in predicting the P600 effect between plausible and implausible conditions, but not in a graded manner.

In contrast to GPT-2, the surprisal estimates obtained from the LSTM were not able to predict most of the observed effects from the studies. In fact, only the association-driven N400 effects of DBC21 were predicted by the LSTM. Still, the different mean estimates between conditions and the correlations with the human-derived metrics provide interesting information. For DBC19 & DBC21, the overall pattern for mean surprisal and the correlations is very similar to the GPT-2 model, but diverges for ADSBC21 & ADBC23. For ADSBC21, LSTM surprisal is lower in the associated vs. un-associated conditions, regardless of cloze probability. Moreover, there is only a correlation with association but not with cloze and only association was significant in predicting LSTM surprisal. For ADBC23, mean LSTM surprisal did not show any noticeable difference between conditions and was neither correlated with plausibility or cloze probability.

This means that experimental manipulations of semantic association appear to have been well captured by both language models at least at the distributional level. Cloze probability, hence overall expectancy, was only partially captured by the GPT-2 model but not by the LSTM and plausibility appears to be difficult to capture for both models.

Two more obvious sources can be identified in order to explain the observed variance in the results: linguistic idiosyncracies in the stimulus material of the four RI-studies and architectural idiosyncracies within the language models. That is, which differences between the stimuli of DBC19, DBC21, ADSBC21 and ADBC23 on the one side, and which differences between GPT-2 and the LSTM on the other side could have led to the observed pattern of results? Importantly, this then connects to a more fundamental question: which factors drive in general the underlying probability distributions of language models? And how does this relate to the observation that the surprisal estimates appeared to align rather well with the human ratings for semantic relatedness, but only partially with cloze probability and not so well with the plausibility ratings? Finally, it needs to be considered what the results may imply for the application of LM surprisal within psycholinguistic studies and specifically ERP-studies.

The remainder of this chapter seeks to address these topics by first qualitatively assessing some of the token predictions that GPT-2 would make for the target word position in the stimuli. This assessment then will also be subject of discussing overall mechanisms that may drive the predictions of neural language models, model-specific architectural differences and also stimulus-related differences between the RI-studies.

4.1. GPT-2 Target Predictions

In order to gain a better understanding of the nature of the probability distribution that the language models generated, GPT-2 was queried for the five tokens with the highest probability at the target word position w_{t+1}. That is, the five token-ids with the highest value from the softmax-output at the output layer at position w_t were collected for all items. The full results for all four RI-studies in German are listed in Appendix A. Importantly, this analysis was conducted post-hoc and therefore did not enter chapter 3 as a main result. Rather, it provides an assessment that supports the discussion of the results in a qualitative way, but as such, no further reaching conclusions can be drawn and the following observations need to be viewed

with caution.

In the Appendix, as well as in the translated examples below, the predictions are sorted for their probabilities in descending order, that is, the highest probability was assigned to the first item of the list. Although a more thorough analysis is beyond the scope of this thesis and needs to remain subject of future work, some insightful observations can be made here. It has to be noted, that it is overall difficult to judge the syntactic adequacy of the predictions, since they are constrained to be the single next token in the sequence, which could always be a subword-token, functioning as prefix to further subword-tokens. This is also the reason why some of the listed predictions feature single characters or character combinations that can't be interpreted as words (see Table A.1, item 6 as an example).

Across studies it appears mostly to be the case that at least some of the predictions have a clear semantic relatedness to one or more words of the preceding context. In DBC19 & DBC21 the prime noun sometimes occurs in the predictions, as can be seen in the example item at the top of Table 4.1, taken from DBC19. Moreover, the predictions often appear to be simply continuations that form collocations with the immediately preceding content of the target sentence. An instance of this pattern can be seen in item 8 of DBC19 (Table 4.1, top). In German, the word 'end', that appears in the predictions in all three conditions, forms a common collocation with the preceding verb phrase, i.e. *"Wenig später war sie am Ende"* could translate to the idiomatic phrase *"A little later, she was at the end of her tether"* or simply to *"A little later, she was at the end."'*. In some cases, the top predictions are almost identical to each other across conditions, even though the context describes a very different situation between at least one condition compared to the others, such as in the bottom example item of Table 4.1.

Con	Target sentence	Target	Predictions
A	Lisa entered the station. A little later, she was at the	platform	['station', 'end', 'main station', 'airport', 'exit']
B	Lisa left the station. A little later, she was at the	platform	['station', 'airport', '.', 'end', 'morning']
C	Lisa entered the house. A little later, she was at the	platform	['floor', 'beach', 'morning', 'end', 'window']
A	Jenni washed the cucumbers. Then she made a	salad	['small', 'cut', 'walk', 'large', 'short']
B	Jenni got out of the bathtub. Then she made a	salad	['small', 'walk', 'jump', 'big', 'step']
C	Jenni sat down in the bathtub. Then she made a	salad	['small', 'walk', 'big', 'deep', 'short']

Table 4.1.: Five highest ranking predictions of GPT-2 for the target word position of item 8 of **DBC19** (top) and of item 30 of **DBC21** (bottom). Items and predictions are translated to English.

Table 4.2 showcases two further interesting examples from ADSBC21. In the top item, the actual target word was predicted to be the most likely continuation in all 4 conditions, even though it is a clearly implausible continuation in conditions C & D. The bottom item on the other hand is an instance of cases, where the verb of the main clause appeared to have a direct influence on the predictions. This latter example could potentially hint to how the surprisal estimates, though still driven by semantic relatedness, captured the diminished cloze probability (and implicitly plausibility) in the C and D conditions of ADSBC21. While most of the predictions for C & D seem intuitively plausible, at the same time they also provide collocative, that is, statistically frequent completions to the verb of the main clause (*"to sense the fire/absence/problem/extent"*).

Thus, it is conceivable, that GPT-2 assigned a low probability to the target word in conditions

C & D simply due to the fact that it provides a statistically very infrequent completion for the verbal phrase of the main clause. This would indeed align here with the predictions of general expectancy *and* plausibility, that is, "*sensing the goal*" would be overall unexpected and implausible, and the items in ADSBC21 were explicitly designed for this feature.

Con	Target sentence	Target	Predictions
A	Today the doctor inoculates, after filling the syringe, the	patient	['patient'[1], 'blood', 'patient'[2], 'doctor', 'dog']
B	Today the doctor inoculates, after closing the drawer, the	patient	['patient'[1], 'children', 'dog', 'man', 'people']
C	Today the doctor fires, after filling the syringe, the	patient	['patient'[1], 'man', 'doctor', 'ambulance', 'boy']
D	Today the doctor fires, after closing the drawer, the	patient	['patient'[1], 'man', 'key', 'doctor', 'wheelchair']
A	Quickly scored the soccer player, who had crossed the soccer field, the	goal	['goal', 'match', 'first', 'foul', 'animal']
B	Quickly scored the soccer player, who had taken off his jacket, the	goal	['goal', 'match', 'first', 'team', 'victim']
C	Quickly sensed the soccer player, who had crossed the soccer field, the	goal	['fire', 'Un'[3], 'misfortune', 'extent', 'Auf'[3]]
D	Quickly sensed the soccer player, who had taken off his jacket, the	goal	['fire', 'absence', 'Un'[3], 'problem', 'extent']

[1] ACC/DAT
[2] NOM
[3] not translated because of subword status

Table 4.2.: **ADSBC21** Five highest ranking predictions of GPT-2 for the target word position of items 24 (top) and 53 (bottom). Items and predictions are transliterated to English, preserving German word order.

In the highest ranking predictions for ADBC23, the impact of semantic relatedness becomes apparent once again. Two example items are shown in Table 4.3. In order to test the predictions of a multi-stream account of the N400 & P600, ADBC23 created items that maximally prime not only the target but also a distractor word in the context paragraph preceding the target sentence[1]. Importantly, in condition A the target is more plausible than the distractor, in condition B the reverse is true, and in C both target and distractor are implausible. Accordingly, the examples given here as well as the full list of predictions given in the appendix feature an additional column, indicating the distractor for the items of ADBC23.

For many items, the top five predictions for A & B feature both the target and the distractor[2]. In some cases, even condition C features both words, as can be seen in the top example in Table 4.3. Crucially, this implies, that it may be possible to "trick" GPT-2 into predicting an implausible continuation by strongly priming it in the preceding context. This being said, specifically in the C condition one can also observe predictions that provide highly frequent and locally plausible completions for the preceding verb of the target sentence. This can be seen in the C condition of the bottom example in Table 4.3, where in German "process" and "farewell" provide frequent complements to the verb "simplify".

To summarize, inspecting the highest-probability predictions that GPT-2 provides for the target word position, it appears to be the case that these often feature words (or subwords) that are strongly semantically related to the preceding context, regardless of whether they are plausible or not. This even goes so far that prime words appear in the predictions, though they usually provide less plausible continuations. Though it has to be noted, that the plausibility of the prediction may be difficult to judge because it has to be considered that the predicted token may not necessarily constitute a complete word but rather a subword, potentially followed by

[1] The authors built upon stimulus material from Nieuwland and Van Berkum (2005).
[2] Either completely or as recognizable subword unit.

Con	Target sentence	Target	Distractor	Predictions
A	In the end *consulted* the judge the	prosecutor	defendant	['prosecutor', 'defendant', 'witness', 'judge', 'attorney']
B	In the end *questioned* the judge the	prosecutor	defendant	['defendant', 'prosecutor', 'witness', 'accused', 'man']
C	In the end *copied* the judge the	prosecutor	defendant	['application', 'prosecutor', 'trial', 'defendant', 'report']
A	Happily *kissed* the bride the	groom	ring	['ring', 'Bräutig'[1], 'bride', 'Trau'[2], 'altar']
B	Happily *admired* the bride the	groom	ring	['ring', 'Bräutig'[1], 'jewelry', 'bride', 'golden']
C	Happily *simplified* the bride the	groom	ring	['process', 'gaze', 'Hochzeits'[2], 'farewell', 'moment']

[1] German subword prefix for "groom".
[2] German subword prefixes semantically related to the wedding topic.

Table 4.3.: **ADBC23** Five highest ranking predictions of GPT-2 for the target word position of item 30 (top) and 31 (bottom). Items and predictions are transliterated to English, preserving German word order.

further subwords which could then still result in a plausible word. For example, the token prediction "ring" for condition A in the bottom example of Table 4.3 could be continued with "bearer", leading to "Happily kissed the bride the ringbearer" (with *ring bearer* being a single word in German), which would constitute possibly a more plausible continuation than *ring*.

To summarize further, the predictions often appear to feature high-frequent collocations alongside either the immediately preceding verb in DBC19, DBC21 and ADBC23, or the more distant verb of the main clause in ADSBC21. Together the target word predictions qualitatively complement the finding, that LM surprisal patterns well and is correlated with the human ratings of semantic association throughout the RI-studies. Moreover, the predictions hint towards how language models may be distinctively influenced by only certain aspects of either immediate or more distant context.

4.2. Training Data and Vocabulary Organization

When it comes to neural network based models of any type, there exist multiple challenges and peculiarities with respect to processing natural language, as indicated in chapter 1. And although the two models that have been used in this thesis differ substantially with respect to their architecture, and this fact may have led to the partially different pattern of results as will be discussed below, there also two properties that they share and which may have impacted the overall quality of surprisal estimates: the training data and the way how input words are tokenized.

The training data for the models was described in chapter 2 and consists of several smaller sub-corpora.[3] To recap, the data comprises the following sources: Wikipedia, NewsCrawl, ParaCrawl, EU Bookshop and Open Subtitles. As to be expected, the Wikipedia set appears to contain descriptive language in terms of encyclopedic articles, communicating factual knowledge. The NewsCrawl set contains German news articles, that also follow a characteristic reporting style. The EU Bookshop set contains more formal, political and legal language. The ParaCrawl set appears to contain a wider variety of language styles as it contains data crawled from various webpages. Finally, the Open Subtitles set appears to contain, as suggested by the name, almost exclusively written conversations. To summarize, the characteristics of these

[3]Note, that from the provided sources it was not possible to verify the precise origin of the sub-corpora.

sub-corpora are considerably different, and this can be a desired feature to introduce the trained language model to a certain amount of linguistic variety.

As comparing language models that differ in their basic architecture is naturally very challenging, the aim was to align the training procedure of the LSTM as much as possible with the one of the pre-trained GPT-2 model. For this reason, the training data was not further preprocessed in any form. Therefore, it should be noted at this point, that the data itself appears to vary greatly in terms of features such as treatment of punctuation and whitespace. Furthermore, there appears to be a lot of variance in what constitutes a single line in the training files. Some lines consist of single sentences, some of only a few characters and some may even display hyperlinks and other strings that arguably wouldn't count as examples for typical natural language.

Although a more thorough discussion may lie beyond the scope of this thesis, specifically when it comes to language modeling in the context of psycholinguistic research, the question arises of how psychologically plausible the data is as a foundation to model aspects of human language comprehension. That is, the question needs to be asked to which extent the training data may capture the amount and characteristics of language input an average adult comprehender may have encountered. As a matter of fact, the predictions of language models will always be biased towards the patterns of language that occur in their training data. This was a considerable factor that was, due to the fact of using a pre-trained language and training another language model accordingly, not controlled for here.

Another theoretical challenge, that has been described in chapter 2, comes in the form of Byte-Pair Encoding (Sennrich et al., 2016). While BPE brings undoubted advantages with respect to computational memory efficiency, it also introduces a potential theoretical problem for estimating word level probabilities and surprisal, specifically for the purpose of psycholinguistic research. The problem arises due to the fact that the splits into subwords happen solely on the basis of frequencies of characters and character combinations and are not linguistically motivated, that is, for example on the basis of morphemes. Crucially, for the units that the BPE-algorithm has derived, it is not distinguishable, whether they represent subword units that only occur alongside other subword units, or standalone words that can not further be composed to higher-level units. To give an example, the tokenizer used in this thesis would encode the word *" Haus"*(house)[4] to the token id *1046*. Further, *" Hauseingang"* (house entrance) would be decoded into ids *1046* and *17,871*, the latter being the id for *"eingang"*. That is, the id *1046* encodes both uses of *Haus* as standalone, $Haus_s$, and as compound, $Haus_c$.

Now, consider we would like to obtain surprisal estimates for the words of the sentence *"Sie wollte das Haus nicht betreten. Also blieb sie im Hauseingang."* (She didn't want to enter the house. So, she stayed in the house entrance.). Using a model that is trained with a BPE-tokenizer, we would obtain a list of surprisal estimates, containing two separate values for *" Haus"* and for *"eingang"*. To obtain a surprisal estimate for *Hauseingang*, usually the individual surprisals for both subwords would be summed. This is the method that has been used here and is commonly used in related work (see Nair and Resnik, 2023). What we would need to distinguish here is $S(Haus_s)$ versus $S(Haus_c) + S(eingang_c)$, that is, there might be cases where the occurrence of the word *Haus* is maybe more surprising than *Hauseingang*. But, since the

[4]Importantly, with a trailing whitespace.

model can not differentiate between $Haus_s$ and $Haus_c$, we get

$$P(Haus) = P(Haus_s) + P(Haus_c) \tag{4.1}$$

and therefore, model surprisal will always be higher for *Hauseingang* relative to *Haus* even if we would expect it to be different in human listeners. For a linguistically motivated tokenizer this may be less of a problem, since the language unit in question would be broken down from words into morphemes, and there exists a clear notion about how lexical words can be comprised of morphemes. Critically though, it is unclear how completely distributionally derived subunits may constitute words and this problem may also be exacerbated for agglutinating languages (Park et al., 2021). As reported in the beginning of chapter 3, a considerable amount of target words of the RI-studies was split into at least two subword units by the tokenizer. Nair and Resnik (2023) find that even though in their overall analyses surprisal estimates do not significantly differ between a BPE-tokenizer and a morpheme-based tokenizer in their aggregate ability of predicting reading times, qualitatively and in terms of psycholinguistic plausibility morpheme based tokenizers might be preferable.

4.3. Similarity-driven Probability Estimates

Fundamentally, all types of neural language models require their input words to be mapped to a numeric representation, which is typically, and in case of the two language models used here, a high-dimensional vector representation, most often learned during training. In fact, all model internal representations in neural language models revolve around these vector representations and performing mathematical operations on them, most prominently the multiplication of weight matrices for forward and backward propagation of activation and the application of activation functions that impose the non-linearities needed for generalization. As described earlier, a considerable benefit of these representations lies in their comparability on a mathematical level. *Cosine similarity* is a metric that reflects the cosine of the angle between two vectors and is computed by dividing the dot product of these vectors by the product of their respective Euclidian norm (Han et al., 2012):

$$\text{Cosine Similarity}(x, y) = \frac{x \bullet y}{\|x\|\|y\|} \tag{4.2}$$

It is most often cosine similarity or a closely related variant that is used when deriving a metric for (semantic) similarity between words or sentences (see Frank, 2017; Michaelov et al., 2023 as examples). The question that then arises is: how might this property of comparability in terms of a mathematical similarity metric find its way into the predictions of the next token? That is, how and at which point may language models make use of such properties inherent to their internal representations? For transformer models, a potential answer may lie in the attention mechanism. When predicting a next word w_{t+1}, a query vector q is computed for w_t and key vectors k are computed for all w_1,...w_t. Then, a similarity score is computed between the query and each of the keys and used to scale a value v. The German GPT-2 version used in this thesis (Schweter, n.d.-b) shares the same architectural features as the original GPT model (Radford & Narasimhan, 2018). Thus, the similarity metric used during the computation of the

attention values should be the same as originally proposed by Vaswani et al. (2017), which is a scaled version of the dot product between the query and key vectors, as displayed in chapter 1, equation 1.2.

Considering now, that the scaled dot product similarity to each of the preceding tokens and the current token itself is what directly impacts the decoding of the next token, it appears plausible, that past tokens that frequently co-occur with the current token shape the probability distribution for the next token towards stronlgy related predictions. That is, tokens less related to the current one receive a smaller score and therefore contribute less to the decoding of the current token, while more strongly related words would make a greater impact in predicting the next token. Importantly, these links can also be established stably across longer distances with intervening words. This was demonstrated in the attention visualizations of Vaswani et al. (2017). It has to be noted, that the original transformer architecture described by the authors features both an encoder and a decoder module, and that the visualizations depict self-attentions of the encoder, that is, attention is not masked and hence the words for which attention is illustrated have access to all words of the sequence, not only the preceding ones as the GPT-2 model would have to. Nonetheless, it can be observed in Figure 3 of Vaswani et al. (2017, p.13), how several attention heads at the words *"more"* and *"difficult"* appear to specifically focus on the verb *"making"* which occurs earlier in the sequence.

Thus, it is conceivable, that the GPT-2 model, when presented with the stimulus material of the RI-studies, may have shown similar attention patterns at the target word position, strongly attending to the semantically related prime words and/or specifically to context verbs for which the actual encountered target word then provided a statistically more or less frequent completion. As a purely hypothetical example, this can be illustrated for item 53 of ADSBC21 (displayed in Table 4.2, bottom). At the target word position, the model has access to all the preceding tokens, and when encoding the next token, attention scores between *"the"* and all the preceding tokens will be computed, as a relative weight to determine to which extent each token should impact the prediction for the next token. It seems plausible how this could lead to predictions that are either semantically related to or form collocations with the words with higher similarity scores. This could therefore explain why GPT-2 appeared to have in conjunction with the verb *"scored"* leaned more towards the semantic theme of *soccer* in conditions A & B. On the other hand, in conditions C & D, which feature the main clause verb *"sense"*, the model may have favored predictions providing frequent collocations specifically with this verb and the determiner *"the"*, e.g. *"sensed...misfortune"*.

Considering that in this way different words of the context may have had a distinct influence on the probability distribution underlying the target word surprisal, this may have led to the observed pattern of mean surprisal between conditions. That is, if the model focused in ADSBC21 most strongly on the verb of the main clause but simultaneously to a lesser extent on the noun of the main clause and the noun and verb of the adverbial clause, the observed pattern of surprisal, with **A** < **B** < **C** < **D** could potentially be explained, alongside with the results of the rERP analysis. In ADBC23, it would also appear plausible that selectively the verb preceding the target word could have led to a high surprisal in condition C compared to A. Though it is less clear why under this assumption the model would show higher surprisal in B relative to A.

But these are only preliminary considerations that could be incorporated into future re-

search. To gain further insight into the distinct role of the attention mechanism in determining the probability and hence surprisal estimates for the target words within the conditions of the experimental stimuli, a visualization similar to the one used by Vaswani et al. (2017) could prove to be helpful. This could be conducted with an appropriate implementation (Vig, 2019) and show specifically for the target words to which of the context words the model attended to. However, although ongoing research is being conducted with the aim to uncover the mode of operation of the attention mechanism, it appears to be very challenging to assign certain functional linguistic interpretations to specific layers and attention heads (Jo & Myaeng, 2020).

Unlike the GPT-2 model, the LSTM does not incorporate attention but rather recurrence to generate probability estimates. This involves that contextual information about past input is passed along in the hidden states sequentially across time. At each time step, the new input is combined with the previous hidden state[5]. Although LSTMs feature a sophisticated mechanism to better store distant information from the past than other types of RNNs, the entire memory is compressed into a single hidden-state representation. The gating mechanisms allow the model to alter the vector representation as a whole, that is, specific values of the context representation may be forgotten and updated with new values to a certain extent. Critically though, the LSTM does not have the option of accessing specific tokens and to differentiate which of those tokens may be more or less relevant in predicting the next token. Thus, it is conceivable that an LSTM has in general more difficulties in noticing subtle differences that are cued by specific words.

Moreover, a known phenomenon when comparing word embedding representations is, that words with contrary meanings show a high cosine similarity to each other due to the fact that they often appear in similar contexts. As an example, comparing the fastText (Bojanowski et al., 2017) embeddings for the words *"betrat"* (*entered*) and *"verließ"* (*left*) yields a relatively high cosine similarity of 0.7. In fact, this feature may be desirable when it comes to operationalizing semantic association, since words of contrasting meanings can be semantically associated to each other via antonymia. If the learned input embeddings that enter the LSTM are very similar, then this would presumably also lead to very similar hidden state representations, thus leading to similar output probabilities. This could be a possible explanation for why LSTM surprisal patterned well with the semantic association ratings. Again, this is a theory that would need to be tested in future work, for example by a more thorough examination of the hidden state representations of the LSTM on the basis of the experimental stimuli. Another point of notice for both models is the fact, that distributional information about statistical co-occurrences of words is one to be accessible on the surface level of language. Plausibility on the other hand, since it is dependent on world knowledge, is not easy to capture and may be more implicitly be reflected in the distributional properties by rarer occurrences of implausible material.

Furthermore, as a rather straightforward but maybe not very profound explanation for why the LSTM overall exhibited higher mean surprisals and was only able to capture one of the N400 effects, one could point to the overall smaller scale and general lesser capabilities as a language model. That is, the sheer size advantage of the transformer model in terms of number of parameters and layers led to an overall better generalization, which is also reflected in the perplexities on the test set.

[5] Again, both input and hidden state are represented as high-dimensional vector representations

4.4. Study-related Differences

In addition to architectural differences between the language models, another important source for the variance observed in the results naturally lies in the stimulus material of the four RI studies.

One observation from the results is, that GPT-2 was only partially able and the LSTM not able at all to capture the P600 effects of the RI studies, that were all elicited by manipulations of expectancy and plausibility, but not association. The question arises, why GPT-2 surprisal was not able to capture the expectancy differences between conditions in DBC19 and DBC21 but at least to some extent in ADSBC21 and ADBC23. One reason might be that due to its sensitivity to semantic relatedness, the model was effectively primed by the context sentences. That is, the strong relatedness may have overwritten the influence of overall expectancy. Simultaneously, it is to note that the plausibility violations that led to reduced expectancy qualitatively differ between studies. It could be argued, that the violations of event structure in DBC19 and DBC21 are rather subtle compared to the violations of selectional restrictions of the verb in ADSBC21 and ADBC23. Moreover, as mentioned in the last section, the verbs that manipulated the expectancy difference in DBC19 and DBC21 may indeed have had very similar embedding representations, leading respectively to similar similarity scores in GPT-2 as well as similar hidden state embeddings in the LSTM. Thus, the difference between those conditions may just have been overall small.

Moreover, ADSBC21 is special in the sense that it manipulates expectancy (and plausibility) through the verb of the main sentence and manipulates association by the intervening adverbial clause. The models may have been impacted differently by this sentence structure. As discussed in the previous section, for GPT-2, the attention mechanism may potentially have led to a stronger focus on the verb of the main clause and a relatively weaker focus on the content of the adverbial clause and this could have resulted in the observed mean surprisal pattern between conditions, that was **A** < **B** < **C** < **D**. The LSTM on the other hand can not weigh specific context words with different strength for its upcoming predictions, but has to compress all of the context into a single representation. Moreover, even with the gating mechanism to remember long term information it is more likely to be impacted by the more recent context compared to transformer models. Therefore, it is conceivable, that in ADSBC21 the LSTM was overall impacted by the content of the adverbial clause, rather than the main clause. This could potentially explain the LSTM mean surprisal between conditions, that is, conditions B and D (unassociated adverbial clause) showed higher surprisal relative to conditions A and C (associated adverbial clause).

An important difference between ADBC23 and the other studies is, that the authors made heavy use of priming. Both the target and a distractor word are mentioned frequently in the context paragraph. This apparently led both target and distractor to often occur in the top 5 predictions of GPT-2 for the target word position. An interesting further analysis would be, if the target systematically ranks higher in the GPT-2 predictions than the distractor in condition A and if the reverse is the case in condition B. This could clarify further how sensitive the model is to the verb manipulation. Given the result that mean target surprisal was higher in condition B relative to A, it may well be the case that the opposite is true for the distractor.

Finally, considering that GPT-2 often appeared to favor collocative predictions with respect

to the verbs throughout the studies, there may be further item-induced variance to uncover across studies. For example, it is conceivable that the model was lured by verbs that have a more narrow, restrictive set of high-frequent completions into predicting those completions and disregarding further cues from the context.

4.5. Summary of Discussion

The results show that the probability estimates of language models may be driven by different factors than predictions made by human listeners during online language comprehension. This is in fact not surprising given the sources of information that both have access to. Human listeners can make use of a life-long fund of profound knowledge of the world and have access to pragmatic reasoning as part of their linguistic expertise, which itself inclcudes syntactical and semantic knowledge. Language models on the other side have only the information encoded in the data they were trained on at their disposal. While this data can be linguistically rich, it is also most often messy and skewed towards containing certain linguistic structures. Controlling for this confound is an immensely challenging task, since the amount of training data needed for a neural language model to generalize well goes beyond what is feasible for manual editing.

Crucially though, contemporary language models are not equipped to capture information that goes beyond distributional properties of language in an explicit way. Even modern large-scale models like GPT-3 do not connect to any knowledge base our build an internal discourse representation which they then query and update according to new incoming words. But expectancy in online language comprehension may considerably be driven by exactly such reasoning. The four RI-studies showed how plausibility violations can lead to both N400 and P600 effects, and the results of this thesis show that exactly those effects are challenging to capture by surprisal estimates collected from language models. Simultaneously, the estimates appeared to align better with the predictions of human-rated semantic relatedness.

Now, when we turn back to the notion of surprisal as a linking hypothesis which links cognitive processing effort to the amount of contextual unexpectedness an encountered word carries, then the question arises as to how psychologically plausible the estimates that language models can provide us with are. That is, can we refer to these model estimates really as *surprisal* in a cognitive sense that they reflect outright predictions a listener might generate for an upcoming word? Answering these questions may have considerable implications for the application of LM surprisal within psycholinguistic research and specifically ERP-research. This is especially crucial when referring to LM surprisal in any explanatory way such as to inform specific theories and mechanisms underlying human language processing.

At the same time, however, specifically more recent large-scale language models offer a promising resource for future research. As the results show, the GPT-2 model was partially able to predict the expectancy-driven N400 and P600 effects observed in the RI-studies. The important question arises as to *(*how) the model may have achieved this, that is, which model-internal mechanisms may have been responsible to generate surprisal estimates that partially align with the cloze probabilities and plausibility judgements collected for the experiments? One potential answer may lie in the transformer's attention mechanism which allows the model

to selectively weigh the importance of long-distance tokens when predicting the next word. Considering that plausibility may to a certain extent manifest itself within distributional information, potentially across longer distances, it appears logical that more advanced and larger models may have a better chance at capturing these relationships.

Therefore, a future goal is to repeat the analyses conducted here with another, architecturally more advanced model, with LeoLM as open-source LLM being a promising candidate (Plüster, 2023). Visualization techniques could then be used systematically to gain insights about the behavior of different attention heads in different attention layers, specifically at the target word position (Vig, 2019). Moreover, a more thorough analysis of the model predictions at the target word positions could confirm the hypothesis that language model probabilities may to a considerable extent be driven by similarity metrics. Finally, it may be considerable to also implement a model that improves upon some potential problems of the models used here by featuring a morpheme-based tokenizer (Smit et al., 2014) and a different training corpus and regime that may aim to increase psychological plausibility.

Chapter 5.

Conclusion

The thesis evaluated the predictions of language model surprisal on the data of four ERP-studies. These studies have found the N400 to be sensitive to manipulations of both expectancy and semantic association, while the P600 was shown to be sensitive to expectancy alone. Given the theoretical conception of surprisal as reflecting expectancy, it was anticipated that LM surprisal would pattern with human-based metrics such as cloze probability and plausibility. Moreover, it was expected that in an rERP analysis LM surprisal would predict the expectancy induced N400 and P600 effects found by the studies, but not the association-driven N400 effects.

However, the results show that this was not the case. First, surprisal estimates of both a GPT-2 and an LSTM model were collected for the target words of all experimental items. Then, mean surprisal of the conditions was compared to the mean ratings which the studies originally collected for semantic association, plausibility and cloze probability. It could be observed that overall LM surprisal aligned the most with the association ratings and only in some cases with cloze probability and plausibility. This relationship was confirmed by computing correlations between LM surprisal and the human judgements as well as using the judgements as predictors for LM surprisal in simple linear regression. Furthermore, an rERP analysis was conducted for each of the models and each of the studies, where LM surprisal was used as single predictor. GPT-2 surprisal predicted all of the N400 effects driven by both expectancy and association and to a smaller extent the expectancy-driven P600 effects. LSTM surprisal on the other hand was only able to predict one of the association-driven N400 effects.

Overall, the results show how LM surprisal estimates may be influenced to a stronger degree by contextual similarity while simultaneously being less able to capture expectancy violations. Potential reasons could lie in the model-internal representations and mechanisms such as attention as well as originate from the training data and the organization of the vocabulary. Further research should be aimed at confirming these observations and gaining a deeper understanding of the model-internal mechanisms that might lead to such a behavior. Moreover, it was discussed how large-scale language models featuring an attention mechanism could better be equipped to detect violations of plausibility that could implicitly show themselves on a distributional level of language.

Importantly, the results indicate that surprisal estimates derived from language models may not adequately capture the concept of theoretical surprisal and its application in psycholinguistics and specifically ERP-research, that is, reflecting overall expectancy within human listeners rather than semantic association. Therefore, the interpretation of LM surprisal may need to be re-evaluated, especially when linking it to theories and mechanisms underlying language comprehension.

Appendix A.

GPT-2 Target Predictions

Table A.1.: **DBC19** GPT-2 top 5 predictions for the target word position.

I	Con	Target Sentence	Target	Predictions
1	a	Johann betrat das Restaurant. Wenig später öffnete er die	Speisekarte	['Tür', 'Küche', 'Eingang', 'Bar', 'Türen']
1	b	Johann verließ das Restaurant. Wenig später öffnete er die	Speisekarte	['Gaststätte', 'Tür', '"', 'Wohnung', 'Bar']
1	c	Johann betrat die Wohnung. Wenig später öffnete er die	Speisekarte	['Tür', 'Wohnung', 'Haustür', 'Balk', 'Fenster']
2	a	Sabine betrat das Kino. Schnell ging sie zur	Kasse	['Film', 'Schauspiel', 'Sache', 'Arbeit', 'Polizei']
2	b	Sabine verließ das Kino. Schnell ging sie zur	Kasse	['Schauspiel', 'Film', 'Arbeit', 'Polizei', 'Musik']
2	c	Sabine betrat die Schule. Schnell ging sie zur	Kasse	['Schule', 'Polizei', 'Uni', 'Schauspiel', 'Arbeit']
3	a	Kevin betrat den Bauernhof. Ohne zu zögern nahm er eine	Mistgabel	['Schüssel', 'Flasche', 'Pistole', 'Stelle', 'Waffe']
3	b	Kevin verließ den Bauernhof. Ohne zu zögern nahm er eine	Mistgabel	['Stelle', 'Arbeit', 'Anstellung', 'Stellung', 'neue']
3	c	Kevin betrat die Kirche. Ohne zu zögern nahm er eine	Mistgabel	['Waffe', 'der', 'Leiter', 'kleine', 'Flasche']
4	a	Susi betrat die Arztpraxis. Freundlich sprach sie mit der	Arzthelferin	['Patientin', 'Mutter', 'Frau', 'Ärztin', 'Schwester']
4	b	Susi verließ die Arztpraxis. Freundlich sprach sie mit der	Arzthelferin	['Mutter', 'Patientin', 'Frau', 'Familie', 'Schwester']
4	c	Susi betrat das Gehege. Freundlich sprach sie mit der	Arzthelferin	['Mutter', 'jungen', 'Kleinen', 'kleinen', 'anderen']
5	a	Roman betrat die Bücherei. Kurz danach ging er zum	Regal	['ersten', 'Studium', 'Unterricht', 'zweiten', '"']
5	b	Roman verließ die Bücherei. Kurz danach ging er zum	Regal	['Studium', 'ersten', '"', 'Journalismus', 'Militär']
5	c	Roman betrat die Kneipe. Kurz danach ging er zum	Regal	['ersten', 'Friseur', 'Auto', 'zweiten', 'Arzt']
6	a	Lea begann Spaghetti zu kochen. Vorsichtig öffnete sie die	Nudelpackung	['Tür', 'Pf', 'T', 'Schüssel', 'Sch']
6	b	Lea war fertig Spaghetti zu kochen. Vorsichtig öffnete sie die	Nudelpackung	['Tür', 'Schüssel', 'T', 'Pf', 'Sch']
6	c	Lea begann die Wände zu streichen. Vorsichtig öffnete sie die	Nudelpackung	['Tür', 'Fenster', 'Schub', 'Türen', 'Sch']
7	a	Peter erreichte das Theater. Wenig später ging er zur	Loge	['Armee', 'Schauspiel', 'Polizei', 'Theater', '"']
7	b	Peter verließ das Theater. Wenig später ging er zur	Loge	['Armee', 'Schauspiel', 'Polizei', '"', 'BBC']
7	c	Peter erreichte den Supermarkt. Wenig später ging er zur	Loge	['Polizei', 'Schule', 'Arbeit', 'Post', 'Bank']
8	a	Lisa betrat den Bahnhof. Wenig später war sie am	Gleis	['Bahnhof', 'Ende', 'Hauptbahnhof', 'Flughafen', 'Ausgang']
8	b	Lisa verließ den Bahnhof. Wenig später war sie am	Gleis	['Bahnhof', 'Flughafen', ':', 'Ende', 'Morgen']
8	c	Lisa betrat das Haus. Wenig später war sie am	Gleis	['Boden', 'Strand', 'Morgen', 'Ende', 'Fenster']
9	a	Tim begann im Fitnessstudio zu trainieren. Umgehend war er auf dem	Laufband	['Weg', 'College', 'Sprung', 'neuesten', 'Platz']
9	b	Tim hörte auf im Fitnessstudio zu trainieren. Umgehend war er auf dem	Laufband	['Weg', 'Sprung', 'neuesten', 'Boden', 'Parkplatz']
9	c	Tim begann seine Fahrstunde. Umgehend war er auf dem	Laufband	['Weg', 'Parkplatz', 'Fahrrad', 'Platz', 'Dach']
10	a	Maria erreichte den Flughafen. Sofort ging sie zum	Check-in	['Flughafen', 'Arzt', 'Bahnhof', 'Hotel', 'Einkaufen']
10	b	Maria verließ den Flughafen. Sofort ging sie zum	Check-in	['Flughafen', 'Bahnhof', 'Arzt', 'Einkaufen', 'Hotel']
10	c	Maria erreichte das Einkaufszentrum. Sofort ging sie zum	Check-in	['Arzt', 'Einkaufen', 'Friseur', 'ersten', 'Bahnhof']
11	a	Lukas kam zur Geburtstagsfeier. Nach einer Weile überreichte er den	Kuchen	['Gästen', 'Kindern', 'beiden', 'Eltern', 'Anwesenden']
11	b	Lukas verließ die Geburtstagsfeier. Nach einer Weile überreichte er den	Kuchen	['Gästen', 'beiden', 'Kindern', 'Anwesenden', 'Geburtstag']
11	c	Lukas kam zum Unterricht. Nach einer Weile überreichte er den	Kuchen	['Schülern', 'Kindern', 'Jungen', 'beiden', 'Eltern']
12	a	Marie ging zum Friseur. Nach einem kurzen Moment fragte sie die	Stylistin	['Friseur', 'Verkä', 'Frise', 'Frau', 'anderen']
12	b	Marie verließ den Friseur. Nach einem kurzen Moment fragte sie die	Stylistin	['Friseur', 'Frau', 'anderen', 'Frise', 'Mutter']
12	c	Marie ging zur Uni. Nach einem kurzen Moment fragte sie die	Stylistin	['Studentin', 'Lehrerin', 'Studenten', 'anderen', 'Professorin']
13	a	Thorsten begann zu duschen. Als erstes benutzte er das	Shampoo	['Bad', 'Wasser', 'Handt', 'Waschbecken', '"']
13	b	Thorsten hörte auf zu duschen. Als erstes benutzte er das	Shampoo	['Telefon', 'Wasser', 'Bad', 'Handy', 'Handt']
13	c	Thorsten begann zu grillen. Als erstes benutzte er das	Shampoo	['Grill', '"', 'Feuer', 'Gas', 'Messer']
14	a	Claudia betrat den Blumenladen. Schnell fragte sie nach einer	Rose	['Bestellung', 'Flasche', 'neuen', 'Lieferung', 'Tasse']
14	b	Claudia verließ den Blumenladen. Schnell fragte sie nach einer	Rose	['neuen', 'Wohnung', 'anderen', 'Stelle', 'Firma']
14	c	Claudia betrat das Tiergeschäft. Schnell fragte sie nach einer	Rose	['Bestellung', 'Katze', 'Zigarette', 'neuen', 'Frau']
15	a	Jonathan ging zum Friedhof. Nachdenklich betrachtete er den	Grabstein	['Tod', 'Grabstein', 'Toten', 'Friedhof', 'Sarg']
15	b	Jonathan verließ den Friedhof. Nachdenklich betrachtete er den	Grabstein	['Tod', 'Friedhof', 'Ort', 'Verlust', 'Anblick']
15	c	Jonathan ging in die Küche. Nachdenklich betrachtete er den	Grabstein	['Tisch', 'Herd', 'Teller', 'Spiegel', 'Koch']
16	a	Lara ging in den Garten. Genau prüfte sie die	Erde	['Pflanzen', 'Blumen', 'Pflanze', 'Obst', 'Stelle']
16	b	Lara verließ den Garten. Genau prüfte sie die	Erde	['Pflanzen', 'Blumen', 'Pflanze', 'verschiedenen', 'Stelle']
16	c	Lara ging zum Herd. Genau prüfte sie die	Erde	['Zutaten', 'Koch', 'Temperatur', 'Küche', 'Eier']
17	a	Roman ging ins Museum. Akribisch betrachtete er die	Skulpturen	['Geschichte', 'Werke', 'Bilder', 'Funde', 'Zeichnungen']
17	b	Roman verließ das Museum. Akribisch betrachtete er die	Skulpturen	['Geschichte', 'Sammlung', 'Ausstellung', 'Arbeit', 'Werke']
17	c	Roman ging in die Buchhandlung. Akribisch betrachtete er die	Skulpturen	['Bücher', 'Werke', 'Geschichte', 'Buch', 'Literatur']
18	a	Annika ging zum Konditor. Schnell fragte sie nach einem	Kuchen	['Rezept', 'neuen', 'guten', 'Kuchen', 'Job']

to be continued on the next page

Table A.1.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
18	b	Annika verließ den Konditor. Schnell fragte sie nach einem	Kuchen	['Job', 'neuen', 'Rezept', 'anderen', 'guten']
18	c	Annika ging in die Boutique. Schnell fragte sie nach einem	Kuchen	['Kleid', 'neuen', 'Mann', 'Job', 'passenden']
19	a	Emil betrat das Spa. Nach einer Weile bekam er eine	Massage	['neue', 'Stelle', 'E', 'Menge', 'kleine']
19	b	Emil verließ das Spa. Nach einer Weile bekam er eine	Massage	['Stelle', 'Anstellung', 'neue', 'Einladung', 'Rolle']
19	c	Emil betrat die Apotheke. Nach einer Weile bekam er eine	Massage	['Nachricht', 'E', 'neue', 'Erk', 'Sprit']
20	a	Susanne betrat die Reitschule. Schnell setzte sie den	Sattel	['Kurs', 'Reit', 'Unterricht', 'Weg', 'Kopf']
20	b	Susanne verließ die Reitschule. Schnell setzte sie den	Sattel	['Reit', 'Weg', 'Unterricht', 'Kurs', 'Kontakt']
20	c	Susanne betrat das Bad. Schnell setzte sie den	Sattel	['Notruf', 'Kopf', 'Gang', 'Weg', 'Mund']
21	a	Martin kam zur Sicherheitskontrolle. Zügig öffnete er den	Koffer	['Kofferraum', 'Koffer', 'Wagen', 'Umschlag', 'Transporter']
21	b	Martin verließ die Sicherheitskontrolle. Zügig öffnete er den	Koffer	['Kofferraum', 'Koffer', 'Wagen', 'Safe', 'Tresor']
21	c	Martin kam zur Autowerkstatt. Zügig öffnete er den	Koffer	['Kofferraum', 'Wagen', 'Laden', 'Motor', 'Koffer']
22	a	Jutta ging zur Hotelrezeption. Sofort bekam sie ein	Zimmer	['Angebot', 'Schreiben', 'Foto', 'Paket', 'neues']
22	b	Jutta verließ die Hotelrezeption. Sofort bekam sie ein	Zimmer	['Angebot', 'Schreiben', 'neues', 'Foto', 'Post']
22	c	Jutta ging zur Post. Sofort bekam sie ein	Zimmer	['Paket', 'Schreiben', 'Angebot', 'Foto', 'Kind']
23	a	Jens ging zum Zahnarzt. Ängstlich ging er zum	Röntgen	['Arzt', 'Zahnarzt', 'Friseur', 'Psychiater', 'Tierarzt']
23	b	Jens verließ den Zahnarzt. Ängstlich ging er zum	Röntgen	['Zahnarzt', 'Arzt', 'Psychiater', 'Friseur', 'Tierarzt']
23	c	Jens ging zur Zollstelle. Ängstlich ging er zum	Röntgen	['Zoll', 'Hafen', 'Markt', 'Schiff', 'Wagen']
24	a	Elena begann Poker zu spielen. Schnell setzte sie ihr	Geld	['Poker', 'Interesse', 'Studium', 'Spiel', 'Geld']
24	b	Elena hörte auf Poker zu spielen. Schnell setzte sie ihr	Geld	['Poker', 'Leben', 'Geld', 'Spiel', 'Studium']
24	c	Elena begann zu zeichnen. Schnell setzte sie ihr	Geld	['Studium', 'Interesse', 'Talent', 'künstler', 'ganzes']
25	a	Bruno begann seinen Urlaub. Endlich kam er am	Meer	['1', 'Sonntag', '23', '17', '21']
25	b	Bruno beendete seinen Urlaub. Endlich kam er am	Meer	['Sonntag', 'Abend', 'Samstag', 'Morgen', 'Montag']
25	c	Bruno begann seinen Filmabend. Endlich kam er am	Meer	['Set', 'Ende', 'Abend', '1', 'nächsten']
26	a	Jana betrat das Bürgeramt. Umgehend zeigte sie ihr	Visum	['Gesicht', 'Foto', 'Handy', 'Bild', 'Baby']
26	b	Jana verließ das Bürgeramt. Umgehend zeigte sie ihr	Visum	['Foto', 'Gesicht', 'Bild', 'wahres', 'Handy']
26	c	Jana betrat das Gartencenter. Umgehend zeigte sie ihr	Visum	['neues', 'wahres', 'Talent', 'Baby', 'die']
27	a	Nico fuhr auf die Kirmes. Sofort ging er zum	Auto-Scooter	['Arzt', 'ersten', 'Auto', 'Training', 'Friseur']
27	b	Nico verließ die Kirmes. Sofort ging er zum	Auto-Scooter	['Karneval', 'ersten', '"', 'FC', '1']
27	c	Nico fuhr zur Pizzeria. Sofort ging er zum	Auto-Scooter	['Supermarkt', 'Restaurant', 'Pizza', 'Auto', 'Arzt']
28	a	Beate ging wandern. Nach einer Weile sah sie ein	Edelweiß	['Mädchen', 'kleines', 'Kind', 'Bild', 'weißes']
28	b	Beate kam zurück vom Wandern. Nach einer Weile sah sie ein	Edelweiß	['Mädchen', 'kleines', 'Foto', 'paar', ',']
28	c	Beate ging schwimmen. Nach einer Weile sah sie ein	Edelweiß	['Mädchen', ',', 'kleines', 'Kind', 'Foto']
29	a	Hubert nahm ein Taxi. Freundlich begrüßte er den	Fahrer	['jungen', 'Fahrer', 'Taxi', 'Bus', 'neuen']
29	b	Hubert verließ das Taxi. Freundlich begrüßte er den	Fahrer	['Fahrer', 'Taxi', 'jungen', 'neuen', 'Bus']
29	c	Hubert nahm ein Flugzeug. Freundlich begrüßte er den	Fahrer	['Piloten', 'Kapitän', 'jungen', 'Flug', 'neuen']
30	a	Jenni begann zu baden. Sofort verteilte sie den	Schaum	['ganzen', 'Inhalt', 'Müll', 'Wein', 'Duft']
30	b	Jenni hörte auf zu baden. Sofort verteilte sie den	Schaum	['Müll', 'ganzen', 'Rest', 'Wein', 'Inhalt']
30	c	Jenni begann die Vorlesung. Sofort verteilte sie den	Schaum	['Text', 'Zettel', 'ersten', 'Artikel', 'Brief']
31	a	Hans betrat die Eishalle. Flott schnürte er seine	Schlittschuhe	['Schuhe', 'Hose', 'Eishockey', 'Handschuhe', 'Hosen']
31	b	Hans verließ die Eishalle. Flott schnürte er seine	Schlittschuhe	['Eishockey', 'Schuhe', 'Eis', 'Handschuhe', 'Schl']
31	c	Hans betrat die Kletterhalle. Flott schnürte er seine	Schlittschuhe	['Kletter', 'Hose', 'Schuhe', 'Ski', 'Jacke']
32	a	Clara ging bowlen. Gekonnt erzielte sie einen	Strike	['zweiten', 'Vorsprung', 'Rekord', 'guten', 'Platz']
32	b	Clara hörte auf zu bowlen. Gekonnt erzielte sie einen	Strike	['Rekord', 'Touch', 'Vorsprung', 'großen', 'neuen']
32	c	Clara ging zum Schießstand. Gekonnt erzielte sie einen	Strike	['Vorsprung', 'Treffer', 'guten', 'Hatt', 'Schuss']
33	a	Georg nahm teil an einem Feuerwehreinsatz. Sofort benutzte er den	Schlauch	['Namen', 'Begriff', 'Notruf', 'Spitznamen', 'Wagen']
33	b	Georg beendete einen Feuerwehreinsatz. Sofort benutzte er den	Schlauch	['Notruf', 'Namen', 'Begriff', 'Wagen', 'Dienst']
33	c	Georg nahm teil an Yoga. Sofort benutzte er den	Schlauch	['Namen', 'Begriff', 'Ausdruck', '"', 'Yoga']
34	a	Johanna begann das Fußballspiel. Nach kurzer Zeit machte sie ein	Tor	['Praktikum', 'Sport', 'Angebot', 'Foto', 'Pro']
34	b	Johanna beendete das Fußballspiel. Nach kurzer Zeit machte sie ein	Tor	['Praktikum', 'Foto', 'Angebot', 'Fotos', 'paar']
34	c	Johanna begann die Jagd. Nach kurzer Zeit machte sie ein	Tor	['Foto', 'paar', 'Jagd', 'gutes', 'Angebot']
35	a	Jürgen betrat die Metzgerei. Sofort fragte er nach einem	Steak	['neuen', 'Rezept', 'guten', 'Brot', 'Stück']
35	b	Jürgen verließ die Metzgerei. Sofort fragte er nach einem	Steak	['neuen', 'anderen', 'Job', 'geeigneten', 'guten']
35	c	Jürgen betrat die Bäckerei. Sofort fragte er nach einem	Steak	['Brot', 'Stück', 'Rezept', 'neuen', 'kleinen']
36	a	Frauke betrat den Zoo. Wenig später fotografierte sie ein	Zebra	['Foto', 'Tier', 'junges', 'paar', 'Mädchen']
36	b	Frauke verließ den Zoo. Wenig später fotografierte sie ein	Zebra	['Foto', 'Bild', 'Tier', 'Mädchen', 'weiteres']
36	c	Frauke ging ins Stadion. Wenig später fotografierte sie ein	Zebra	['Foto', 'Bild', 'paar', 'Mädchen', 'Video']
37	a	Bernd betrat den Waschsalon. Schnell wählte er das	Waschprogramm	['Wort', 'Pseudonym', 'Café', '"', 'Gebäude']
37	b	Bernd verließ den Waschsalon. Schnell wählte er das	Waschprogramm	['Pseudonym', 'Leben', 'Wort', 'Café', '"']
37	c	Bernd betrat den Kiosk. Schnell wählte er das	Waschprogramm	['Wort', 'Wasser', 'Lokal', '"', 'Pseudonym']
38	a	Nicole ging in den Zirkus. Bald setzte sie sich auf die	Tribüne	['Bühne', 'Bank', 'Tanz', 'Straße', 'Knie']
38	b	Nicole verließ den Zirkus. Bald setzte sie sich auf die	Tribüne	['Bühne', 'Straße', 'Bank', 'Suche', 'Terrasse']
38	c	Nicole ging ins Büro. Bald setzte sie sich auf die	Tribüne	['Terrasse', 'Bank', 'Couch', 'Toilette', 'Treppe']
39	a	Viktor ging in die Sauna. Nach einer Weile machte er einen	Aufguss	['Spaziergang', 'kleinen', 'Fehler', 'sehr', 'Sprung']
39	b	Viktor verließ die Sauna. Nach einer Weile machte er einen	Aufguss	['Spaziergang', 'Ausflug', 'Fehler', 'kleinen', 'sehr']
39	c	Viktor ging ins Labor. Nach einer Weile machte er einen	Aufguss	['Fehler', 'kleinen', 'sehr', 'Rück', 'selts']
40	a	Gabi ging zum Tierarzt. Einen Moment später fragte sie die	Arzthelferin	['Mutter', 'Frau', 'Katze', 'anderen', 'Heb']
40	b	Gabi verließ den Tierarzt. Einen Moment später fragte sie die	Arzthelferin	['Mutter', 'Ärzte', 'Frau', 'anderen', 'Polizei']
40	c	Gabi ging in die Umkleide. Einen Moment später fragte sie die	Arzthelferin	['anderen', 'Frau', 'Mutter', 'beiden', 'Kinder']
41	a	Jakob betrat den Tatort. Kurz darauf fotografierte er die	Blutspritzer	['beiden', 'Frau', 'Leiche', 'Tat', 'Wohnung']
41	b	Jakob verließ den Tatort. Kurz darauf fotografierte er die	Blutspritzer	['Leiche', 'beiden', 'Frau', 'Szene', 'Tat']
41	c	Jakob betrat die Tankstelle. Kurz darauf fotografierte er die	Blutspritzer	['Tankstelle', 'beiden', 'Frau', 'junge', 'drei']
42	a	Peter begann Bäume zu fällen. Nach einer Weile benutzte sie die	Axt	['Bäume', 'Straße', 'Säge', 'Erde', 'Steine']
42	b	Peter hörte auf Bäume zu fällen. Nach einer Weile benutzte sie die	Axt	['Bäume', 'Erde', 'Steine', 'Straße', 'Zeit']
42	c	Peter begann die Kutsche zu fahren. Nach einer Weile benutzte sie die	Axt	['Kut', 'Pferde', 'Straße', 'Dampf', 'erste']
43	a	Tom begann ein Lagerfeuer. Sofort stapelte er das	Holz	['Holz', 'Essen', 'Lager', 'Feuer', 'Zelt']
43	b	Tom beendete das Lagerfeuer. Sofort stapelte er das	Holz	['Holz', 'Essen', 'Zelt', 'Lager', 'Feuer']
43	c	Tom begann Frühstück zu machen. Sofort stapelte er das	Holz	['Frühstück', 'Brot', 'Paket', 'Essen', 'Bett']
44	a	Kim ging zum Weihnachtsmarkt. Kurz darauf erreichte er die	Weihnachtskiosks	['Nachricht', 'Stadt', 'Botschaft', 'Polizei', 'Grenze']

to be continued on the next page

Table A.1.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
44	b	Kim verließ den Weihnachtsmarkt. Kurz darauf erreichte er die	Weihnachtskiosks	['Stadt', 'Nachricht', 'Grenze', 'Botschaft', 'Polizei']
44	c	Kim ging ins Krankenhaus. Kurz darauf erreichte er die	Weihnachtskiosks	['Nachricht', 'Diagnose', 'Klinik', 'Polizei', 'Grenze']
45	a	Michael begann sein Mittagessen zu kochen. Schnell schnitt er das	Gemüse	['Essen', 'Gemüse', 'Fleisch', 'Frühstück', 'Brot']
45	b	Michael war fertig sein Mittagessen zu kochen. Schnell schnitt er das	Gemüse	['Gemüse', 'Essen', 'Frühstück', 'Fleisch', 'Rezept']
45	c	Michael begann einen Kuchen zu backen. Schnell schnitt er das	Gemüse	['Brot', 'Rezept', 'Mehl', 'Kuchen', 'Stück']
46	a	Johann betrat die Wohnung. Sofort setzte er sich aufs	Bett	['Sofa', 'Bett', 'Pferd', 'Dach', 'Klo']
46	b	Johann verließ die Wohnung. Sofort setzte er sich aufs	Bett	['Sofa', 'Pferd', 'Bett', 'Fahrrad', 'Rad']
46	c	Johann betrat das Kino. Sofort setzte er sich aufs	Bett	['Sofa', 'Bett', 'Dach', 'Klavier', 'Podium']
47	a	Sabine betrat die Schule. Schnell ging sie zum	Klassensaal	['Studium', 'ersten', 'Sport', 'Schwimmen', 'Unterricht']
47	b	Sabine verließ die Schule. Schnell ging sie zum	Klassensaal	['Studium', 'ersten', 'Militär', 'Sport', 'Theater']
47	c	Sabine betrat den Bauernhof. Schnell ging sie zum	Klassensaal	['ersten', 'Haus', 'Hof', 'Stall', 'Bauernhof']
48	a	Kevin betrat die Kirche. Vorsichtig nahm er eine	Kerze	['kleine', 'der', 'Pf', 'Flasche', 'Decke']
48	b	Kevin verließ die Kirche. Vorsichtig nahm er eine	Kerze	['kleine', 'Pf', 'der', 'Leiter', 'Flasche']
48	c	Kevin betrat die Arztpraxis. Vorsichtig nahm er eine	Kerze	['Sprit', 'kleine', 'Flasche', 'Nadel', 'Tabl']
49	a	Susi betrat das Gehege. Präzise begutachtete sie das	Kalb	['Tier', 'Geb', 'Geh', 'Verhalten', 'Jung']
49	b	Susi verließ das Gehege. Präzise begutachtete sie das	Kalb	['Tier', 'Geh', 'Geb', 'Verhalten', 'Ges']
49	c	Susi betrat die Bücherei. Präzise begutachtete sie das	Kalb	['Buch', 'Bild', 'neue', 'Bücher', 'Manuskript']
50	a	Thomas betrat die Kneipe. In Eile rief er den	Kellner	['Wirt', 'Kellner', '""', 'Herrn', 'Bar']
50	b	Thomas verließ die Kneipe. In Eile rief er den	Kellner	['""', 'Freund', 'Club', 'Laden', ',,']
50	c	Thomas begann Spaghetti zu kochen. In Eile rief er den	Kellner	['Koch', 'Chef', '""', 'Küchen', 'Sohn']
51	a	Lea begann die Wände zu streichen. Schnell wählte sie einen	Pinsel	['Ort', 'anderen', 'Weg', 'Stil', 'der']
51	b	Lea war fertig die Wände zu streichen. Schnell wählte sie einen	Pinsel	['anderen', 'Ort', 'Weg', 'neuen', 'Raum']
51	c	Lea erreichte das Theater. Schnell wählte sie einen	Pinsel	['anderen', 'neuen', 'Weg', 'Beruf', 'Ort']
52	a	Peter erreichte den Supermarkt. Schnell kaufte er etwas	Gemüse	['Obst', '.', 'zu', ',', 'und']
52	b	Peter verließ den Supermarkt. Schnell kaufte er etwas	Gemüse	[',', 'von', 'zu', 'Geld', 'und']
52	c	Peter erreichte den Bahnhof. Schnell kaufte er etwas	Gemüse	['Geld', 'später', 'von', 'zu', 'Land']
53	a	Lisa betrat das Haus. Kurz danach checkte sie den	Kühlschrank	['Fernseher', 'Schlüssel', 'Besitzer', 'Computer', 'Mann']
53	b	Lisa verließ das Haus. Kurz danach checkte sie den	Kühlschrank	['Computer', 'Fernseher', 'Wagen', 'Besitzer', 'neuen']
53	c	Lisa begann im Fitnessstudio zu trainieren. Kurz danach checkte sie den	Kühlschrank	['Fitness', 'Fit', 'ersten', 'neuen', 'Trainer']
54	a	Tim begann seine Fahrstunde. Hektisch drückte er aufs	Gaspedal	['Gas', 'Lenkrad', 'Pedal', 'Brem', 'Handy']
54	b	Tim beendete seine Fahrstunde. Hektisch drückte er aufs	Gaspedal	['Gas', 'Lenkrad', 'Pedal', 'Brem', 'Handy']
54	c	Tim ging zum Flughafen. Hektisch drückte er aufs	Gaspedal	['Gas', 'Handy', 'Roll', 'Dach', 'Bein']
55	a	Maria erreichte das Einkaufszentrum. Sofort begrüßte sie die	Verkäuferin	['Gäste', 'ersten', 'vielen', 'Ankunft', 'Kunden']
55	b	Maria verließ das Einkaufszentrum. Sofort begrüßte sie die	Verkäuferin	['Kunden', 'neuen', 'Gäste', 'ersten', 'neue']
55	c	Maria erreichte die Geburtstagsfeier. Sofort begrüßte sie die	Verkäuferin	['Gäste', 'Anwesenden', 'Kinder', 'anwesenden', 'Familie']
56	a	Lukas begann die Vorlesung. Eilig begrüßte er seine	Studenten	['Schüler', 'Zuhörer', 'Studenten', 'Mitschüler', 'Gäste']
56	b	Lukas beendete die Vorlesung. Eilig begrüßte er seine	Studenten	['Schüler', 'Studenten', 'Zuhörer', 'Gäste', 'Kollegen']
56	c	Lukas ging zum Friseur. Eilig begrüßte er seine	Studenten	['Freundin', 'Frau', 'Mutter', 'Kunden', 'Braut']
57	a	Marie ging zur Uni. In Eile betrat sie den	Hörsaal	['Raum', 'Konferenz', 'Flur', 'Lehrstuhl', 'ersten']
57	b	Marie verließ die Uni. In Eile betrat sie den	Hörsaal	['Raum', 'Laden', 'Konferenz', 'Lehrstuhl', 'Campus']
57	c	Marie ging duschen. In Eile betrat sie den	Hörsaal	['Raum', 'Salon', 'Saal', 'Laden', 'Pool']
58	a	Thorsten begann zu grillen. Behutsam platzierte er den	Grillanzünder	['Grill', 'Topf', 'Teller', 'Deckel', 'Ball']
58	b	Thorsten hörte auf zu grillen. Behutsam platzierte er den	Grillanzünder	['Grill', 'Topf', 'Deckel', 'Teller', 'Ball']
58	c	Thorsten betrat den Blumenladen. Behutsam platzierte er den	Grillanzünder	['Schlüssel', 'Blumenstrauß', 'Beutel', 'Strauß', 'kleinen']
59	a	Claudia betrat das Tiergeschäft. Eine Zeit lang stand sie am	Käfig	['Fenster', 'Eingang', 'Rande', 'Rand', 'Zaun']
59	b	Claudia verließ das Tiergeschäft. Eine Zeit lang stand sie am	Käfig	['Rande', 'Fenster', 'Telefon', 'Eingang', 'Ende']
59	c	Claudia betrat den Friedhof. Eine Zeit lang stand sie am	Käfig	['Rande', 'Rand', 'Grab', 'Eingang', 'Fuße']
60	a	Jonathan ging in die Küche. Eine Weile säuberte er den	Ofen	['Raum', 'Tisch', 'Boden', 'Kühlschrank', 'Müll']
60	b	Jonathan verließ die Küche. Eine Weile säuberte er den	Ofen	['Raum', 'Kühlschrank', 'Tisch', 'Boden', 'Herd']
60	c	Jonathan ging in den Garten. Eine Weile säuberte er den	Ofen	['Garten', 'Rasen', 'Boden', 'Platz', 'Baum']
61	a	Lara ging zum Herd. Schnell reinigte sie die	Herdplatte	['Küche', 'Wohnung', 'Wäsche', 'Wunde', 'Haut']
61	b	Lara ging weg vom Herd. Schnell reinigte sie die	Herdplatte	['Wäsche', 'Wohnung', 'Küche', 'Hände', 'Haut']
61	c	Lara ging ins Museum. Schnell reinigte sie die	Herdplatte	['Wände', 'Bilder', 'Kleider', 'Haare', 'Kleidung']
62	a	Roman ging in die Buchhandlung. Schnell suchte er einen	Roman	['Verlag', 'Verleger', 'neuen', 'Partner', 'Job']
62	b	Roman verließ die Buchhandlung. Schnell suchte er einen	Roman	['neuen', 'Verlag', 'Verleger', 'Job', 'Partner']
62	c	Roman ging zum Konditor. Schnell suchte er einen	Roman	['neuen', 'Partner', 'Job', 'Ort', 'Arbeitsplatz']
63	a	Annika ging in die Boutique. Eine Zeit lang suchte sie nach	Stiefeln	['einem', 'einer', 'dem', 'neuen', 'der']
63	b	Annika verließ die Boutique. Eine Zeit lang suchte sie nach	Stiefeln	['einem', 'einer', 'neuen', 'dem', 'der']
63	c	Annika ging ins Spa. Eine Zeit lang suchte sie nach	Stiefeln	['einem', 'einer', 'dem', 'der', 'ihrem']
64	a	Emil betrat die Apotheke. Schnell kaufte er das	Medikament	['Haus', 'Medikament', 'Rezept', 'nötige', '""']
64	b	Emil verließ die Apotheke. Schnell kaufte er das	Medikament	['Haus', 'Grundstück', 'Geschäft', '""', 'Buch']
64	c	Emil betrat die Reitschule. Schnell kaufte er das	Medikament	['Pferd', 'Anwesen', 'Reit', '""', 'Grundstück']
65	a	Susanne betrat das Bad. Sofort entfernte sie ihr	Makeup	['Gesicht', 'Haar', 'Kleid', 'Make', 'Kopftuch']
65	b	Susanne verließ das Bad. Sofort entfernte sie ihr	Makeup	['Haar', 'Gesicht', 'Kleid', 'Make', 'Bad']
65	c	Susanne ging zur Sicherheitskontrolle. Sofort entfernte sie ihr	Makeup	['Handy', 'Gesicht', 'Auto', 'Kleid', 'Mobiltelefon']
66	a	Martin kam zur Autowerkstatt. Umgehend fragte er nach den	Reifen	['Reifen', 'Kennzeichen', 'Preisen', 'Fahr', 'Autos']
66	b	Martin verließ die Autowerkstatt. Umgehend fragte er nach den	Reifen	['neuen', 'Gründen', 'Kosten', 'Preisen', 'anderen']
66	c	Martin kam zur Hotelrezeption. Umgehend fragte er nach den	Reifen	['Zimmern', 'Preisen', 'Zimmer', 'Hotel', 'Öffnungszeiten']
67	a	Jutta ging zur Post. Umgehend bekam sie das	Paket	['Angebot', 'Schreiben', 'Paket', 'Post', 'Foto']
67	b	Jutta verließ die Post. Umgehend bekam sie das	Paket	['Angebot', 'Schreiben', 'Geld', 'Versprechen', 'erste']
67	c	Jutta ging zum Zahnarzt. Umgehend bekam sie das	Paket	['Angebot', 'Rezept', 'erste', 'Medikament', 'Gefühl']
68	a	Jens ging zur Zollstelle. Schnell zeigte er sein	Formular	['wahres', 'Können', 'Talent', 'Interesse', 'Geschick']
68	b	Jens verließ die Zollstelle. Schnell zeigte er sein	Formular	['wahres', 'Talent', 'Können', 'Interesse', 'Potenzial']
68	c	Jens begann Poker zu spielen. Schnell zeigte er sein	Formular	['Talent', 'Können', 'Poker', 'wahres', 'Potenzial']

to be continued on the next page

Table A.1.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
69	a	Elena begann zu zeichnen. Umgehend nahm sie den	Stift	['Namen', 'Künstlern', 'Pinsel', 'ersten', 'Auftrag']
69	b	Elena hörte auf zu zeichnen. Umgehend nahm sie den	Stift	['Namen', 'Job', 'Pinsel', 'Auftrag', 'Brief']
69	c	Elena begann den Urlaub. Umgehend nahm sie den	Stift	['Flug', 'Bus', 'Namen', 'nächsten', 'Job']
70	a	Bruno begann seinen Filmabend. Schnell besorgte er das	Popcorn	['Drehbuch', 'nötige', 'Visum', 'Geld', 'Material']
70	b	Bruno beendete seinen Filmabend. Schnell besorgte er das	Popcorn	['Drehbuch', 'Geld', 'Visum', 'nötige', 'notwendige']
70	c	Bruno betrat das Bürgeramt. Schnell besorgte er das	Popcorn	['Geld', 'Amt', 'Zimmer', 'nötige', 'erste']
71	a	Jana betrat das Gartencenter. Nach einer Weile fragte sie nach einer	Palme	['Bestellung', 'neuen', 'Tasse', 'Wohnung', 'Tasche']
71	b	Jana verließ das Gartencenter. Nach einer Weile fragte sie nach einer	Palme	['neuen', 'Wohnung', 'Stelle', 'anderen', 'Arbeit']
71	c	Jana ging auf die Kirmes. Nach einer Weile fragte sie nach einer	Palme	['neuen', 'Zigarette', 'kleinen', 'Flasche', 'Tasse']
72	a	Nico fuhr zur Pizzeria. Sofort griff er die	Speisekarte	['Frau', 'junge', 'beiden', 'Polizisten', 'Verkä']
72	b	Nico verließ die Pizzeria. Sofort griff er die	Speisekarte	['Frau', 'beiden', 'anderen', 'junge', 'Polizei']
72	c	Nico fuhr wandern. Sofort griff er die	Speisekarte	['Frau', 'anderen', 'Polizei', 'Vorder', 'Katze']
73	a	Beate ging schwimmen. Schnell prüfte sie das	Becken	['Wasser', 'Schwimm', 'Boot', 'Schwimmen', 'Bad']
73	b	Beate kam zurück vom schwimmen. Schnell prüfte sie das	Becken	['Wasser', 'Boot', 'Schwimm', 'Schwimmen', 'Meer']
73	c	Beate fuhr mit dem Taxi. Schnell prüfte sie das	Becken	['Auto', 'Taxi', 'Fahrzeug', 'Angebot', 'Verhalten']
74	a	Hubert nahm ein Flugzeug. Voller Begeisterung genoss er die	Höhe	['Flug', 'Schönheit', 'Landschaft', 'Atmosphäre', 'Aussicht']
74	b	Hubert verließ das Flugzeug. Voller Begeisterung genoss er die	Höhe	['Aussicht', 'Atmosphäre', 'Flug', 'Fahrt', 'Schönheit']
74	c	Hubert begann zu baden. Voller Begeisterung genoss er die	Höhe	['Schönheit', 'Natur', 'Bade', 'Wasser', 'Aussicht']
75	a	Jenni begann den Unterricht. Sofort öffnete sie die	Tafel	['Augen', 'Tür', 'Türen', 'Schule', 'ersten']
75	b	Jenni hörte auf mit dem Unterricht. Sofort öffnete sie die	Tafel	['Tür', 'Augen', 'Schule', 'Türen', 'Wohnung']
75	c	Jenni betrat die Eishalle. Sofort öffnete sie die	Tafel	['Tür', 'Augen', 'Eis', 'Türen', 'Luft']
76	a	Hans betrat die Kletterhalle. Wenig später griff er das	Seil	['Mädchen', 'Kind', 'Seil', 'Kletter', 'erste']
76	b	Hans verließ die Kletterhalle. Konzentriert griff er das	Seil	['Kletter', 'Kind', 'Seil', 'Mädchen', 'Material']
76	c	Hans begann zu bowlen. Konzentriert griff er das	Seil	['Fahrrad', 'Gewicht', 'Kind', 'Pferd', 'Gleichgewicht']
77	a	Clara ging zum Schießstand. Sofort bekam sie einen	Revolver	['Brief', 'Anruf', 'Schlag', 'Schock', 'neuen']
77	b	Clara verließ den Schießstand. Sofort bekam sie einen	Revolver	['Anruf', 'Brief', 'neuen', 'Schlag', 'Schock']
77	c	Clara ging auf einen Feuerwehreinsatz. Sofort bekam sie einen	Revolver	['Anruf', 'Schock', 'Brief', 'Notruf', 'Schlag']
78	a	Georg nahm teil an Yoga. Nach einer Weile setzte er sich auf seine	Yogamatte	['Füße', 'Beine', 'Knie', 'Matte', 'eigene']
78	b	Georg kam zurück von Yoga. Nach einer Weile setzte er sich auf seine	Yogamatte	['Knie', 'Matte', 'Beine', 'Füße', 'Schultern']
78	c	Georg nahm teil am Fußballspiel. Nach einer Weile setzte er sich auf seine	Yogamatte	['linke', 'Bank', 'eigene', 'alte', 'rechte']
79	a	Johanna begann die Jagd. Nach einer Weile erblickte sie den	Hochsitz	['Hund', 'Wolf', 'Mann', 'jungen', 'König']
79	b	Johanna kam zurück von der Jagd. Nach einer Weile erblickte sie den	Hochsitz	['Hund', 'Mann', 'Wald', 'Wolf', 'kleinen']
79	c	Johanna betrat die Metzgerei. Nach einer Weile erblickte sie den	Hochsitz	['Mann', 'Laden', 'Metzger', 'Herrn', 'Sohn']
80	a	Jürgen betrat die Bäckerei. Wenig später zeigte er auf das	Croissant	['Schild', 'Bild', 'Foto', 'Display', 'Gesicht']
80	b	Jürgen verließ die Bäckerei. Wenig später zeigte er auf das	Croissant	['Schild', 'Bild', 'Foto', 'Wappen', 'neue']
80	c	Jürgen betrat den Zoo. Wenig später zeigte er auf das	Croissant	['Schild', 'Bild', 'Tier', 'Foto', 'Mask']
81	a	Frauke ging ins Stadion. Schnell wählte sie eine	Tribüne	['andere', 'neue', 'Abkürzung', 'der', 'Nummer']
81	b	Frauke verließ das Stadion. Schnell wählte sie eine	Tribüne	['andere', 'neue', 'Freundin', 'Abkürzung', 'der']
81	c	Frauke ging zum Waschsalon. Schnell wählte sie eine	Tribüne	['andere', 'neue', 'Frau', 'der', 'Wohnung']
82	a	Bernd betrat den Kiosk. Sofort kaufte er eine	Zeitung	['Flasche', 'Pistole', 'Schach', 'Cola', 'Karte']
82	b	Bernd verließ den Kiosk. Sofort kaufte er eine	Zeitung	['Flasche', 'neue', 'kleine', 'gebrauch', 'Kiste']
82	c	Bernd betrat den Zirkus. Sofort kaufte er eine	Zeitung	['kleine', 'neue', 'Zirkus', 'große', 'Rolle']
83	a	Nicole betrat das Büro. Zuerst ging sie zum	Schreibtisch	['Telefon', 'Schreibtisch', 'Büro', 'Computer', 'Friseur']
83	b	Nicole verließ das Büro. Zuerst ging sie zum	Schreibtisch	['Büro', 'Telefon', 'Schreiben', 'Friseur', 'Fernsehen']
83	c	Nicole betrat die Sauna. Zuerst ging sie zum	Schreibtisch	['Bad', 'Pool', 'Arzt', 'Wasser', 'Friseur']
84	a	Viktor ging ins Labor. Schnell begann er das	Experiment	['Geheimnis', 'Blut', 'Experiment', 'Problem', 'Labor']
84	b	Viktor verließ das Labor. Schnell begann er das	Experiment	['Projekt', 'Experiment', 'Leben', 'Buch', 'Studium']
84	c	Viktor ging zum Tierarzt. Schnell begann er das	Experiment	['Leiden', 'Leben', 'Problem', 'Verhalten', 'Blut']
85	a	Gabi ging in die Umkleide. Schnell probierte sie die	Kleidung	['neuen', 'neue', 'verschiedenen', 'Möglichkeiten', 'Technik']
85	b	Gabi verließ die Umkleide. Schnell probierte sie die	Kleidung	['neuen', 'neue', 'verschiedenen', 'Möglichkeiten', 'Möglichkeit']
85	c	Gabi betrat den Tatort. Schnell probierte sie die	Kleidung	['verschiedenen', 'neuen', 'Tricks', 'neue', 'Tat']
86	a	Jakob betrat die Tankstelle. Sofort sagte er seine	Nummer	['Meinung', 'Bestellung', 'Reise', 'Frau', 'Arbeit']
86	b	Jakob verließ die Tankstelle. Sofort sagte er seine	Nummer	['Arbeit', 'Meinung', 'Reise', 'Frau', 'Teilnahme']
86	c	Jakob begann Bäume zu fällen. Sofort sagte er seine	Nummer	['Arbeit', 'Meinung', 'Familie', 'Bäume', 'Hilfe']
87	a	Hilde begann die Kutsche zu fahren. Schnell bestieg sie den	Kutschbock	['Zug', 'Berg', 'Wagen', '"', 'Gipfel']
87	b	Hilde war fertig die Kutsche zu fahren. Schnell bestieg sie den	Kutschbock	['Zug', 'Berg', 'Wagen', 'Bus', 'Bahnhof']
87	c	Hilde begann ein Lagerfeuer. Schnell bestieg sie den	Kutschbock	['Gipfel', 'Berg', 'Mount', 'höchsten', 'Kil']
88	a	Tom begann Frühstück zu machen. Sofort öffnete er die	Butter	['Tür', 'Augen', 'Türen', 'Küche', 'Fenster']
88	b	Tom hörte auf Frühstück zu machen. Sofort öffnete er die	Butter	['Tür', 'Augen', 'Haustür', 'Fenster', 'Türen']
88	c	Tom ging zum Weihnachtsmarkt. Sofort öffnete er die	Butter	['Tür', 'Türen', 'Haustür', 'Tore', 'Pfor']
89	a	Kim ging ins Krankenhaus. Gleich ging sie zur	Rezeption	['Schule', 'Polizei', 'Arbeit', 'Post', 'Toilette']
89	b	Kim verließ das Krankenhaus. Gleich ging sie zur	Rezeption	['Schule', 'Polizei', 'Arbeit', 'Post', 'Armee']
89	c	Kim fing an sein Mittagessen zu kochen. Gleich ging sie zur	Rezeption	['Küche', 'Toilette', 'Arbeit', 'Schule', 'Tür']
90	a	Michael fing an einen Kuchen zu backen. Als erstes nahm er einen	Schneebesen	['Kuchen', 'Löffel', 'kleinen', 'Apfel', 'Stein']
90	b	Michael war fertig einen Kuchen zu backen. Als erstes nahm er einen	Schneebesen	['Kuchen', 'kleinen', 'Löffel', 'mit', 'Apfel']
90	c	Michael betrat das Restaurant. Als erstes nahm er einen	Schneebesen	['Schluck', 'Teller', 'Drink', 'Stuhl', 'Burger']

Table A.2.: **DBC21** GPT-2 top 5 predictions for the target word position.

I	Con	Target Sentence	Target	Predictions
1	A	Johann ging raus in den Regen . Sofort öffnete er seinen	Schirm	['Mund', 'Mantel', 'Sack', 'Garten', 'Wagen']
1	B	Johann verließ das Restaurant . Sofort öffnete er seinen	Schirm	['Laden', 'eigenen', 'Mund', 'neuen', 'ersten']
1	C	Johann betrat das Restaurant . Sofort öffnete er seinen	Schirm	['Mund', 'Blick', 'Mantel', 'Bauch', 'Kopf']
2	A	Philipp nahm seinen Radhelm . Nach einem Moment setzte er sich auf sein	Fahrrad	['Fahrrad', 'Rad', 'Motorrad', 'Knie', 'Bike']
2	B	Philipp kam aus dem Schwimmbecken . Nach einem Moment setzte er sich auf sein	Fahrrad	['Fahrrad', 'Bett', 'Knie', 'Rad', 'Brett']
2	C	Philipp ging ins Schwimmbecken . Nach einem Moment setzte er sich auf sein	Fahrrad	['Fahrrad', 'Rad', 'Bett', 'Knie', 'Becken']
3	A	Dorian wollte die Clowns sehen . Ein paar Minuten später erreichte er den	Zirkus	['Ort', 'Bahnhof', 'Eingang', 'Laden', 'Flughafen']
3	B	Dorian verließ den Barbier . Ein paar Minuten später erreichte er den	Zirkus	['Ort', 'Hafen', 'Bahnhof', 'Laden', 'Strand']
3	C	Dorian betrat den Barbier . Ein paar Minuten später erreichte er den	Zirkus	['Barb', 'Ort', 'Laden', 'Hof', 'Salon']
4	A	Alina sollte Calamari kochen . Nach kurzem Überlegen ging sie zum	Fischverkäufer	['Kochen', 'Abendessen', 'Koch', 'Tisch', 'Herd']
4	B	Alina verließ die Krippe . Nach kurzem Überlegen ging sie zum	Fischverkäufer	['Altar', 'Stall', 'ersten', 'König', 'Haus']
4	C	Alina betrat die Krippe . Nach kurzem Überlegen ging sie zum	Fischverkäufer	['Altar', 'Stall', 'ersten', 'Fenster', 'Vor']
5	A	Thorben zog seine Motorradjacke an . Einen Moment später nahm er sein	Motorrad	['Motorrad', 'Fahrrad', 'Gewehr', 'Auto', 'Handy']
5	B	Thorben war fertig damit , den Zaun zu streichen . Einen Moment später nahm er sein	Motorrad	['Handy', 'Fahrrad', 'Gewehr', 'Messer', 'Auto']
5	C	Thorben begann , den Zaun zu streichen . Einen Moment später nahm er sein	Motorrad	['Fahrrad', 'Gewehr', 'Auto', 'Handy', 'erstes']
6	A	Manuel ging raus ins Gewitter . Bald darauf sah er einen	Blitz	['Mann', 'riesigen', 'großen', 'anderen', 'schwarzen']
6	B	Manuel verließ das Finanzamt . Bald darauf sah er einen	Blitz	['Mann', 'neuen', 'Film', 'anderen', 'weiteren']
6	C	Manuel betrat das Finanzamt . Bald darauf sah er einen	Blitz	['Mann', 'Brief', 'anderen', 'weiteren', 'jungen']
7	A	Mia ging raus in die Kälte . Augenblicklich sah sie einen	Schneemann	['Mann', 'Jungen', 'anderen', 'großen', 'Licht']
7	B	Mia verließ das Schmuckgeschäft . Augenblicklich sah sie einen	Schneemann	['Mann', 'anderen', 'neuen', 'jungen', 'Jungen']
7	C	Mia betrat das Schmuckgeschäft . Augenblicklich sah sie einen	Schneemann	['Mann', 'jungen', 'Jungen', 'anderen', 'schwarzen']
8	A	Lea legte sich an den Pool . Nach einer Weile bekam sie	Sonnenbrand	['einen', 'ein', 'eine', 'das', 'den']
8	B	Lea verließ die Gaststätte . Nach einer Weile bekam sie	Sonnenbrand	['einen', 'eine', 'ein', 'die', 'den']
8	C	Lea betrat die Gaststätte . Nach einer Weile bekam sie	Sonnenbrand	['einen', 'eine', 'ein', 'den', 'von']
9	A	Jörg raste über die Autobahn . Kurz darauf hatte er einen	Unfall	['Unfall', 'schweren', 'Autounfall', 'Herzinfarkt', 'Verkehrsunfall']
9	B	Jörg beendete seinen Mittagsschlaf . Kurz darauf hatte er einen	Unfall	['Herzinfarkt', 'Schlaganfall', 'Unfall', 'weiteren', 'schweren']
9	C	Jörg begann seinen Mittagsschlaf . Kurz darauf hatte er einen	Unfall	['Traum', 'Schlaganfall', 'Herzinfarkt', 'Unfall', 'Schlag']
10	A	Lisa ging in den Garten . Nachdenklich betrachtete sie die	Rosen	['Blumen', 'Früchte', 'Blüten', 'Sonne', 'Pflanzen']
10	B	Lisa öffnete das Rollo . Nachdenklich betrachtete sie die	Rosen	['Welt', 'Augen', 'Schönheit', 'Dinge', 'Hand']
10	C	Lisa schloss das Rollo . Nachdenklich betrachtete sie die	Rosen	['Welt', 'Schönheit', 'Menschen', 'Geschichte', 'Dinge']
11	A	Josef wollte eine SMS schreiben . Umgehend nahm er sein	Handy	['Handy', 'Telefon', 'Fahrrad', 'Mobiltelefon', 'Geld']
11	B	Josef kam aus dem Wasser . Umgehend nahm er sein	Handy	['Schwert', 'Pferd', 'Wasser', 'Messer', 'Rad']
11	C	Josef sprang ins Wasser . Umgehend nahm er sein	Handy	['Seil', 'Wasser', 'Pferd', 'Bad', 'Fahrrad']
12	A	Regina setzte sich in die Bibliothek . Bald öffnete sie ein	Buch	['Zimmer', 'Fenster', 'Buch', 'kleines', 'eigenes']
12	B	Regina kam vom Fußballtraining zurück . Bald öffnete sie ein	Buch	['Geschäft', 'Zimmer', 'Restaurant', 'kleines', 'Sport']
12	C	Regina ging zum Fußballtraining . Bald öffnete sie ein	Buch	['Geschäft', 'Restaurant', 'Zimmer', 'Sport', 'kleines']
13	A	Mario kaufte Mehl und Milch . Bald darauf machte er einen	Kuchen	['Ausflug', 'großen', 'Spaziergang', 'Besuch', 'Fehler']
13	B	Mario stand aus dem Bett auf . Bald darauf machte er einen	Kuchen	['Fehler', 'Rück', 'Anruf', 'großen', 'schrecklichen']
13	C	Mario legte sich ins Bett . Bald darauf machte er einen	Kuchen	['Heirats', 'Rück', 'Fehler', 'Ausflug', 'Spaziergang']
14	A	Lukas kam am Skiort an . Eine Weile spielte er im	Schnee	['Team', 'Verein', '""', 'Jugend', 'Nachwuchs']
14	B	Lukas stieg aus dem Auto aus . Eine Weile spielte er im	Schnee	['Kinder', 'Team', 'Chor', 'Auto', 'Club']
14	C	Lukas stieg ins Auto ein . Eine Weile spielte er im	Schnee	['Team', 'Kinder', 'Club', 'Chor', '""']
15	A	Sophia machte den Gameboy an . Prompt begann sie ein	Spiel	['neues', 'paar', 'Video', 'Fan', 'langes']
15	B	Sophia kam von der Arbeit . Prompt begann sie ein	Spiel	['Praktikum', 'neues', 'paar', 'Foto', 'Studium']
15	C	Sophia ging zur Arbeit . Prompt begann sie ein	Spiel	['Praktikum', 'Studium', 'neues', 'Gespräch', 'Buch']
16	A	Luisa kam von der Post . Schnell öffnete sie ihre	Briefe	['Augen', 'Tasche', 'Tür', 'Brie', 'Wohnung']
16	B	Luisa hörte auf , Klavier zu spielen . Schnell öffnete sie ihre	Briefe	['Augen', 'Stimme', 'eigene', 'Mä', 'Tür']
16	C	Luisa begann , Klavier zu spielen . Schnell öffnete sie ihre	Briefe	['Augen', 'eigenen', 'Stimme', 'eigene', 'musikalische']
17	A	Marina hatte Hunger . Rasch öffnete sie den	Kühlschrank	['Mund', 'Kühlschrank', 'Laden', 'Koffer', 'Beutel']
17	B	Marina kam vom Sozialamt . Rasch öffnete sie den	Kühlschrank	['Brief', 'Laden', 'Koffer', 'Kofferraum', 'Kühlschrank']
17	C	Marina ging zum Sozialamt . Rasch öffnete sie den	Kühlschrank	['Brief', 'Laden', 'Schalter', 'Kindern', 'Kofferraum']
18	A	Klaus erreichte sein Apartment . Sogleich öffnete er die	Wohnungstür	['Tür', 'Wohnung', 'Haustür', 'Eingang', 'Fenster']
18	B	Klaus beendete sein Mittagessen . Sogleich öffnete er die	Wohnungstür	['Tür', 'Augen', 'Küche', 'Flasche', 'Eingang']
18	C	Klaus begann sein Mittagessen . Sogleich öffnete er die	Wohnungstür	['Tür', 'Augen', 'Küche', 'Flasche', 'Eingang']
19	A	Max plante ein Lagerfeuer . Zunächst sammelte er ein paar	Stöcke	['Flaschen', 'Kerzen', 'Vorräte', 'Holz', 'Pilze']
19	B	Max verließ den Wohnwagen . Zunächst sammelte er ein paar	Stöcke	['Flaschen', 'Fotos', 'Tage', 'Bilder', 'Bier']
19	C	Max betrat den Wohnwagen . Zunächst sammelte er ein paar	Stöcke	['Flaschen', 'Bier', 'Sachen', 'Zigaretten', 'Pilze']
20	A	Emilie war unterwegs zum Tennis . Eilig ging sie auf den	Sportplatz	['Rasen', 'Golfplatz', 'Platz', 'Tennis', 'Spielplatz']
20	B	Emilie verließ den Buchladen . Eilig ging sie auf den	Sportplatz	['Markt', 'Friedhof', 'Hof', 'Strich', 'Weg']
20	C	Emilie betrat den Buchladen . Eilig ging sie auf den	Sportplatz	['Laden', 'Markt', 'Bücher', 'Kunden', 'Tisch']
21	A	Martha wollte Steinpilze sammeln . Glücklich ging sie in den	Wald	['Wald', 'Garten', 'Keller', 'Park', 'Laden']
21	B	Martha verließ das Parkhaus . Glücklich ging sie in den	Wald	['Park', 'Keller', 'Laden', 'Supermarkt', 'nächsten']
21	C	Martha betrat das Parkhaus . Glücklich ging sie in den	Wald	['Park', 'ersten', 'Laden', 'Fahrstuhl', 'zweiten']
22	A	Sarah war es zu heiß in der Sonne . Daraufhin setzte sie sich in den	Schatten	['Schatten', 'Garten', 'Wagen', 'Wald', 'Pool']
22	B	Sarah stieg aus ihrem LKW aus . Daraufhin setzte sie sich in den	Schatten	['Bus', 'Wagen', 'Zug', 'Kofferraum', 'Wald']
22	C	Sarah stieg in ihren LKW . Daraufhin setzte sie sich in den	Schatten	['Bus', 'Wagen', 'Zug', 'Kofferraum', 'LKW']
23	A	Paul setzte sich auf die Wiese . Unmittelbar hörte er die	Bienen	['Musik', 'Schüsse', 'Sch', 'Geräusche', 'Stimme']
23	B	Paul verließ das Versicherungsgebäude . Unmittelbar hörte er die	Bienen	['Stimme', 'Schüsse', 'Musik', 'Nachricht', 'Rede']
23	C	Paul betrat das Versicherungsgebäude . Unmittelbar hörte er die	Bienen	['Schüsse', 'Stimme', 'Explosion', 'Sir', 'Sch']
24	A	Annika fuhr mit ihrem Motorrad . Eine Weile stand sie an der	Ampel	['Straße', 'Spitze', 'Kreuzung', 'Seite', 'Ampel']
24	B	Annika verließ die Konditorei . Eine Weile stand sie an der	Ampel	['Kasse', 'Spitze', 'Seite', 'Ecke', 'Tür']

to be continued on the next page

Table A.2.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
24	C	Annika betrat die Konditorei . Eine Weile stand sie an der	Ampel	['Kasse', 'Tür', 'Seite', 'Ecke', 'The']
25	A	Pia erwartete eine Postkarte . Prompt öffnete sie den	Briefkasten	['Brief', 'Umschlag', 'Koffer', 'Deckel', 'Schlüssel']
25	B	Pia kam vom Wandern zurück . Prompt öffnete sie den	Briefkasten	['Kofferraum', 'Koffer', 'Brief', 'Mund', 'Rucksack']
25	C	Pia ging wandern . Prompt öffnete sie den	Briefkasten	['Mund', 'Brief', 'Koffer', 'Eingang', 'Weg']
26	A	Markus betrat die Bank . Eilig öffnete er den	Tresor	['Tresor', 'Safe', 'Schalter', 'Geld', 'Eingang']
26	B	Markus betrat die Wohnung . Eilig öffnete er den	Tresor	['Schlüssel', 'Safe', 'Tresor', 'Koffer', 'Schrank']
26	C	Markus verließ die Wohnung . Eilig öffnete er den	Tresor	['Brief', 'Koffer', 'Schlüssel', 'Kofferraum', 'Safe']
27	A	Ingo musste etwas programmieren . Unverzüglich öffnete er seinen	Laptop	['Computer', 'Rechner', 'Laptop', 'Kopf', 'Schreibtisch']
27	B	Ingo kam aus dem Whirlpool . Unverzüglich öffnete er seinen	Laptop	['Mund', 'Kopf', 'Bauch', 'Penis', 'Fuß']
27	C	Ingo stieg in den Whirlpool . Unverzüglich öffnete er seinen	Laptop	['Mund', 'Kopf', 'Bauch', 'Fuß', 'Penis']
28	A	Jana betrat den Forst . Auf der Stelle stolperte sie über einen	Baumstamm	['Baum', 'Ast', 'kleinen', 'Wald', 'Zaun']
28	B	Jana verließ die Autowerkstatt . Auf der Stelle stolperte sie über einen	Baumstamm	['Baum', 'Zaun', 'Reifen', 'Mann', 'kleinen']
28	C	Jana betrat die Autowerkstatt . Auf der Stelle stolperte sie über einen	Baumstamm	['Zaun', 'Baum', 'Mann', 'Reifen', 'kleinen']
29	A	Nina betrat das Gartencenter . Bald sah sie eine	Blume	['Frau', 'große', 'junge', 'neue', 'riesige']
29	B	Nina verließ das Hallenbad . Bald sah sie eine	Blume	['Frau', 'große', 'neue', 'Reihe', 'Menge']
29	C	Nina betrat das Hallenbad . Bald sah sie eine	Blume	['große', 'Frau', 'junge', 'riesige', 'Gruppe']
30	A	Jenni wusch die Gurken . Dann machte sie einen	Salat	['kleinen', 'Schnitt', 'Spaziergang', 'großen', 'kurzen']
30	B	Jenni stieg aus der Badewanne . Dann machte sie einen	Salat	['kleinen', 'Spaziergang', 'Sprung', 'großen', 'Schritt']
30	C	Jenni setzte sich in die Badewanne . Dann machte sie einen	Salat	['kleinen', 'Spaziergang', 'großen', 'tiefen', 'kurzen']
31	A	Bernd betrat das Klassenzimmer . Unverzüglich reparierte er die	Tafel	['Tür', 'Sch', 'Schul', 'Türen', 'Maschine']
31	B	Bernd betrat das Büro . Unverzüglich reparierte er die	Tafel	['Tür', 'Maschine', 'Wohnung', 'Eingang', 'Sch']
31	C	Bernd verließ das Büro . Unverzüglich reparierte er die	Tafel	['Maschine', 'Tür', 'Wohnung', 'alte', 'Schreib']
32	A	Barbara machte einen Fahrradausflug . Bald erreichte sie den	Badesee	['Campingplatz', 'Strand', 'Ort', 'Flughafen', 'Hafen']
32	B	Barbara verließ das Gebäude . Bald erreichte sie den	Badesee	['Ort', 'Hafen', 'Bahnhof', 'Hof', 'Eingang']
32	C	Barbara betrat das Gebäude . Bald erreichte sie den	Badesee	['Hof', 'Eingang', 'Ort', 'Turm', 'Palast']
33	A	Maria brauchte einen Strauß Rosen . Nach kurzer Überlegung ging sie in den	Blumenladen	['Garten', 'Wald', 'Stall', 'Park', 'Hof']
33	B	Maria verließ das Jugendamt . Nach kurzer Überlegung ging sie in den	Blumenladen	['Kindergarten', 'Ruhestand', 'Jugend', 'Schul', 'Vor']
33	C	Maria betrat das Jugendamt . Nach kurzer Überlegung ging sie in den	Blumenladen	['Kindergarten', 'Jugend', 'Zeugen', 'Keller', 'Laden']
34	A	Jens ging raus in die Natur . Sofort hörte er die	Vögel	['Stimme', 'Geräusche', 'Musik', 'Vögel', 'ersten']
34	B	Jens verließ die Mensa . Sofort hörte er die	Vögel	['Musik', 'Stimme', 'ersten', 'Frage', 'Nachricht']
34	C	Jens betrat die Mensa . Sofort hörte er die	Vögel	['ersten', 'Sir', 'Musik', 'Sch', 'Stimme']
35	A	Helena rief den Notruf . Wenig später sah sie einen	Krankenwagen	['Mann', 'Jungen', 'schwarzen', 'jungen', 'weißen']
35	B	Helena öffnete die Jalousie . Wenig später sah sie einen	Krankenwagen	['Mann', 'anderen', 'jungen', 'großen', 'riesigen']
35	C	Helena schloss die Jalousie . Wenig später sah sie einen	Krankenwagen	['Mann', 'anderen', 'großen', 'jungen', 'neuen']
36	A	Sven ging zur Wattwanderung . Schnell schlüpfte er in seine	Gummistiefel	['Rolle', 'neue', 'alte', 'Jacke', 'Kleidung']
36	B	Sven verließ die Moschee . Schnell schlüpfte er in seine	Gummistiefel	['Rolle', 'neue', 'Kleidung', 'alte', 'Haut']
36	C	Sven betrat die Moschee . Schnell schlüpfte er in seine	Gummistiefel	['Kleidung', 'Rolle', 'neue', 'Uniform', 'Jacke']
37	A	Klara hatte Lust , shoppen zu gehen . In Eile ging sie in die	Boutique	['Stadt', 'Kirche', 'Schule', 'Innenstadt', 'Sauna']
37	B	Klara verließ die Pizzeria . In Eile ging sie in die	Boutique	['Küche', 'Piz', 'Wohnung', 'Bar', 'Kirche']
37	C	Klara betrat die Pizzeria . In Eile ging sie in die	Boutique	['Küche', 'Piz', 'Bar', 'Wohnung', 'Bibliothek']
38	A	Marius erreichte den Supermarkt . Sogleich holte er einen	Einkaufswagen	['Freund', 'Mann', 'kleinen', 'seiner', 'anderen']
38	B	Marius stieg von seinem Moped ab . Sogleich holte er einen	Einkaufswagen	['anderen', 'Freund', 'Mann', 'Bekannten', 'neuen']
38	C	Marius stieg auf sein Moped . Sogleich holte er einen	Einkaufswagen	['Motor', 'kleinen', 'Freund', 'anderen', 'weiteren']
39	A	Michael ging zum Training . Als Erstes machte er ein paar	Dehnübungen	['Übungen', 'Fotos', 'kleine', 'Spr', 'Runden']
39	B	Michael kam vom Kino zurück . Als Erstes machte er ein paar	Dehnübungen	['Fotos', 'Aufnahmen', 'Filme', 'Szenen', 'Sachen']
39	C	Michael ging ins Kino . Als Erstes machte er ein paar	Dehnübungen	['Fotos', 'Aufnahmen', 'Filme', 'Szenen', 'Bilder']
40	A	Andrea hatte Karies . Nach einer Weile ging sie zum	Zahnarzt	['Arzt', 'Zahnarzt', 'Friseur', 'ersten', 'Tierarzt']
40	B	Andrea beendete die Chorprobe . Nach einer Weile ging sie zum	Zahnarzt	['ersten', 'Chor', 'Gesang', 'Proben', 'Konzert']
40	C	Andrea begann die Chorprobe . Nach einer Weile ging sie zum	Zahnarzt	['ersten', 'Proben', 'Chor', 'Gesang', 'Singen']
41	A	Frank ging zum Tierarzt . Gut gelaunt holte er seinen	Hund	['Hund', 'Sohn', 'Vater', 'Bruder', 'kleinen']
41	B	Frank beendete den Pilatesunterricht . Gut gelaunt holte er seinen	Hund	['Vater', 'ersten', 'Trainer', 'Bruder', 'Freund']
41	C	Frank begann den Pilatesunterricht . Gut gelaunt holte er seinen	Hund	['Vater', 'ersten', 'Bruder', 'Trainer', 'Freund']
42	A	Daniel ging zum Yoga . Direkt griff er nach seiner	Yogamatte	['Hand', 'Schulter', 'Tasche', 'Hose', 'Waffe']
42	B	Daniel kam vom Marktplatz zurück . Direkt griff er nach seiner	Yogamatte	['Tasche', 'Jacke', 'Waffe', 'Gitarre', 'Hand']
42	C	Daniel ging auf den Marktplatz . Direkt griff er nach seiner	Yogamatte	['Tasche', 'Waffe', 'Hand', 'Jacke', 'Hose']
43	A	Erik setzte sich vor den Fernseher . Sofort nahm er die	Fernbedienung	['Kamera', 'Herausforderung', 'erste', 'Einladung', 'nächste']
43	B	Erik hörte auf , Gitarre zu spielen . Sofort nahm er die	Fernbedienung	['Gitarre', 'Arbeit', 'Gelegenheit', 'Bass', 'Band']
43	C	Erik begann , Gitarre zu spielen . Sofort nahm er die	Fernbedienung	['Gitarre', 'ersten', 'Musik', 'erste', 'Gelegenheit']
44	A	Melissa hatte im Halteverbot geparkt . Sofort sah sie das	Knöllchen	['Auto', 'Kind', 'Tier', 'Licht', 'Schild']
44	B	Melissa verließ die Universität . Sofort sah sie das	Knöllchen	['Ende', 'Potential', 'Potenzial', 'Licht', 'Problem']
44	C	Melissa betrat die Universität . Sofort sah sie das	Knöllchen	['Potential', 'Ausmaß', 'Licht', 'Potenzial', 'Problem']
45	A	Karla hatte Geburtstag . Kurzerhand machte sie eine	Torte	['Party', 'kleine', 'Reise', 'Geburtstag', 'Überraschung']
45	B	Karla bestand ihre Prüfung . Kurzerhand machte sie eine	Torte	['Ausbildung', 'Lehre', 'Reise', 'Pause', 'Ausnahme']
45	C	Karla fiel durch ihre Prüfung . Kurzerhand machte sie eine	Torte	['Ausbildung', 'Lehre', 'Ausnahme', 'Diät', 'Prüfung']
46	A	Roman wollte Geld abheben . Schnell ging er zum	Bankautomaten	['ersten', 'Geld', 'Telefon', 'nächsten', 'Poker']
46	B	Roman verließ die Klinik . Schnell ging er zum	Bankautomaten	['Studium', 'ersten', 'Militär', 'Arzt', 'Psychiater']
46	C	Roman betrat die Klinik . Schnell ging er zum	Bankautomaten	['Arzt', 'ersten', 'Psychiater', 'Studium', 'Training']
47	A	Emil brauchte Benzin . Prompt erreichte er die	Tankstelle	['Stadt', 'Nachricht', 'Grenze', 'Stelle', 'Polizei']
47	B	Emil verließ das Fitnessstudio . Prompt erreichte er die	Tankstelle	['gleiche', 'Nachricht', 'Nummer', 'Top', '"'"']
47	C	Emil betrat das Fitnessstudio . Prompt erreichte er die	Tankstelle	['Nachricht', 'Nummer', 'gleiche', 'Fitness', 'Stelle']
48	A	Susanne ging los zur Beerdigung . Schnell erreichte sie den	Friedhof	['Ort', 'Friedhof', 'Tod', 'ersten', 'Hof']
48	B	Susanne verließ die Reitschule . Schnell erreichte sie den	Friedhof	['Ruf', 'Status', 'Höhepunkt', 'ersten', 'höchsten']
48	C	Susanne betrat die Reitschule . Schnell erreichte sie den	Friedhof	['Reit', 'ersten', 'Hof', 'Stall', 'Höhepunkt']
49	A	Hans sollte ein Steak kaufen . Umgehend ging er zum	Metzger	['Restaurant', 'Laden', 'Supermarkt', 'Metzger', 'Markt']
49	B	Hans verließ die Eislaufhalle . Umgehend ging er zum	Metzger	['Eis', 'Eishockey', 'Training', 'Studium', 'ersten']
49	C	Hans betrat die Eislaufhalle . Umgehend ging er zum	Metzger	['Eis', 'ersten', 'Training', 'Eingang', 'nächsten']

to be continued on the next page

Table A.2.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
50	A	Lara liebte die Vorweihnachtszeit . Zufrieden ging sie auf den	Weihnachtsmarkt	['Weihnachtsmarkt', 'Markt', 'Weg', 'Weihnachts', 'Heim']
50	B	Lara verließ das Bowlingcenter . Zufrieden ging sie auf den	Weihnachtsmarkt	['Spielplatz', 'Golfplatz', 'Parkplatz', 'Weg', 'Platz']
50	C	Lara betrat das Bowlingcenter . Zufrieden ging sie auf den	Weihnachtsmarkt	['Tisch', 'Parkplatz', 'Platz', 'Spielplatz', 'Ball']
51	A	Lara ging ins Callcenter . Als Erstes machte sie einen	Anruf	['Anruf', 'Job', 'Termin', 'kleinen', 'Fehler']
51	B	Lara hörte auf zu meditieren . Als Erstes machte sie einen	Anruf	['Spaziergang', 'Ausflug', 'kleinen', 'kurzen', 'Kurs']
51	C	Lara begann zu meditieren . Als Erstes machte sie einen	Anruf	['Spaziergang', 'Kurs', 'Ausflug', 'kleinen', 'tiefen']
52	A	Matthias brauchte ein neues Passbild . Umgehend ging er zum	Fotografen	['Friseur', 'ersten', 'nächsten', 'Arzt', 'Kunden']
52	B	Matthias verließ das Schwimmbad . Umgehend ging er zum	Fotografen	['Arzt', 'Flughafen', 'Training', 'Bahnhof', 'Auto']
52	C	Matthias betrat das Schwimmbad . Umgehend ging er zum	Fotografen	['Pool', 'Schwimm', 'Wasser', 'Bad', 'Arzt']
53	A	Sabine kam auf der Badeinsel an . Direkt ging sie an den	Strand	['Strand', 'Start', 'Ball', 'Tisch', 'Steuer']
53	B	Sabine verließ das Fußballstadion . Direkt ging sie an den	Strand	['Start', 'Strand', 'Ball', 'Computer', 'Tisch']
53	C	Sabine betrat das Fußballstadion . Direkt ging sie an den	Strand	['Ball', 'Mann', 'Tisch', 'Start', 'Platz']
54	A	Susi hatte Lust , Poker zu spielen . Schnell ging sie ins	Kasino	['Bett', 'Internet', 'Geld', 'Fitness', 'Spiel']
54	B	Susi verließ das Tierheim . Schnell ging sie ins	Kasino	['Ausland', 'Tierheim', 'Heim', 'Kinder', 'Krankenhaus']
54	C	Susi betrat das Tierheim . Schnell ging sie ins	Kasino	['Tierheim', 'Haus', 'Bett', 'Heim', 'Zimmer']
55	A	Jakob erreichte den Bahnhof . Sofort ging er zum	Gleis	['Bahnhof', 'Arzt', 'Dienst', 'nächsten', 'Post']
55	B	Jakob verließ die Bücherei . Sofort ging er zum	Gleis	['Studium', 'Militär', 'Unterricht', 'Arzt', 'ersten']
55	C	Jakob betrat die Bücherei . Sofort ging er zum	Gleis	['Lesen', 'ersten', 'Unterricht', 'Arzt', 'Fenster']
56	A	Lisa hatte Lust auf ein Bier . Schnellen Schrittes ging sie in die	Kneipe	['Küche', 'Bar', 'Wohnung', 'Kneipe', 'Sauna']
56	B	Lisa verließ den Plattenladen . Schnellen Schrittes ging sie in die	Kneipe	['Wohnung', 'Küche', 'Bar', 'Garage', 'Stadt']
56	C	Lisa betrat den Plattenladen . Schnellen Schrittes ging sie in die	Kneipe	['Küche', 'Wohnung', 'Bar', 'Garder', 'Bibliothek']
57	A	Tim hatte Lust zu rauchen . Nach einem Moment nahm er eine	Zigarette	['Zigarette', 'kleine', 'Zigar', 'Flasche', 'Pistole']
57	B	Tim verließ den Baumarkt . Nach einem Moment nahm er eine	Zigarette	['Stelle', 'neue', 'Auszeit', 'andere', 'Arbeit']
57	C	Tim betrat den Baumarkt . Nach einem Moment nahm er eine	Zigarette	['Pistole', 'Waffe', 'Flasche', 'der', 'kleine']
58	A	Manuela hatte Halsschmerzen . Rasch machte sie einen	Tee	['Spaziergang', 'Arzt', 'An', 'starken', 'kleinen']
58	B	Manuela beendete ihr Workout . Rasch machte sie einen	Tee	['großen', 'guten', 'neuen', 'Sprung', 'riesigen']
58	C	Manuela begann ihr Workout . Rasch machte sie einen	Tee	['großen', 'guten', 'sehr', 'Sprung', 'riesigen']
59	A	Claudia ging schlafen . Rasch stellte sie einen	Wecker	['neuen', 'Zusammenhang', 'Kontakt', 'Antrag', 'Brief']
59	B	Claudia begann , Eier zu kochen . Rasch stellte sie einen	Wecker	['Kontakt', 'großen', 'neuen', 'kleinen', 'Zusammenhang']
59	C	Claudia war fertig damit , Eier zu kochen . Rasch stellte sie einen	Wecker	['neuen', 'Topf', 'großen', 'kleinen', 'Eimer']
60	A	Jonathan wollte einen Urlaub buchen . Zügig ging er ins	Reisebüro	['Hotel', 'Meer', 'Bett', 'Dorf', 'Restaurant']
60	B	Jonathan verließ die Eisdiele . Zügig ging er ins	Reisebüro	['Eis', 'Restaurant', 'Geschäft', 'Kino', 'Café']
60	C	Jonathan betrat die Eisdiele . Zügig ging er ins	Reisebüro	['Eis', 'Bad', 'Restaurant', 'Lokal', 'Geschäft']
61	A	Marie sah , dass sich ein Sturm ankündigte . Nachdenklich betrachtete sie die	Wolken	['Sonne', 'Lage', 'Wolken', 'Tatsache', 'Aussicht']
61	B	Marie verließ die Messehalle . Nachdenklich betrachtete sie die	Wolken	['Welt', 'Atmosphäre', 'Menschen', 'Musik', 'Lage']
61	C	Marie betrat die Messehalle . Nachdenklich betrachtete sie die	Wolken	['Menschen', 'vielen', 'Sonne', 'Augen', 'Welt']
62	A	Dominik musste einen Anruf machen . Eilig suchte er eine	Telefonzelle	['Wohnung', 'Unterkunft', 'neue', 'Stelle', 'Frau']
62	B	Dominik verließ den Hubschrauber . Eilig suchte er eine	Telefonzelle	['Unterkunft', 'Wohnung', 'Stelle', 'neue', 'Toilette']
62	C	Dominik stieg in den Hubschrauber . Eilig suchte er eine	Telefonzelle	['Unterkunft', 'Wohnung', 'Stelle', 'neue', 'Toilette']
63	A	Georg betrat das Wellnesscenter . Nach kurzer Zeit ging er in die	Sauna	['Sauna', 'Küche', 'Bar', 'Lobby', 'Badewanne']
63	B	Georg beendete den Polizeieinsatz . Nach kurzer Zeit ging er in die	Sauna	['USA', 'Politik', 'Türkei', 'Schweiz', 'Lehre']
63	C	Georg begann den Polizeieinsatz . Nach kurzer Zeit ging er in die	Sauna	['USA', 'Türkei', 'Wohnung', 'Politik', 'Lehre']
64	A	Frauke ging auf die Straße . Nachdenklich betrachtete sie den	Verkehr	['Mann', 'Menschen', 'Zustand', 'Blick', 'Anblick']
64	B	Frauke verließ den Zoo . Nachdenklich betrachtete sie den	Verkehr	['Zoo', 'Menschen', 'Tod', 'Anblick', 'Zustand']
64	C	Frauke betrat den Zoo . Nachdenklich betrachtete sie den	Verkehr	['Menschen', 'kleinen', 'Löwen', 'Blick', 'Kopf']
65	A	Sophie wollte eine neue Lampe anbringen . Nachdenklich betrachtete sie die	Decke	['Lampe', 'neue', 'Sonne', 'Welt', 'Uhr']
65	B	Sophie ging in die Bar . Nachdenklich betrachtete sie die	Decke	['Welt', 'Bar', 'Frau', 'Menschen', 'Schönheit']
65	C	Sophie verließ die Bar . Nachdenklich betrachtete sie die	Decke	['Welt', 'Menschen', 'Lage', 'Tatsache', 'Schönheit']
66	A	Herrmann ging zum Trompetenunterricht . Eilig betrat er die	Musikschule	['Bühne', 'Schule', 'Orgel', 'Kapelle', 'Werkstatt']
66	B	Herrmann beendete die Fahrstunde . Eilig betrat er die	Musikschule	['Bühne', 'Bahn', 'Straße', 'Strecke', 'Kabine']
66	C	Herrmann begann die Fahrstunde . Eilig betrat er die	Musikschule	['Bühne', 'Bahn', 'Straße', 'Schule', 'Werkstatt']
67	A	Heidi musste einen Wecker kaufen . Rasch ging sie zum	Uhrmacher	['Radio', 'Telefon', 'Friseur', 'nächsten', 'ersten']
67	B	Heidi verließ das Sozialamt . Rasch ging sie zum	Uhrmacher	['Sozial', 'ersten', 'Arzt', 'Einkaufen', 'Arbeits']
67	C	Heidi betrat das Sozialamt . Rasch ging sie zum	Uhrmacher	['Telefon', 'Sozial', 'Arzt', 'Dienst', 'Arbeits']
68	A	Olaf wollte Brötchen kaufen . Schnell ging er zum	Bäcker	['Bäcker', 'Supermarkt', 'Kiosk', 'ersten', 'Laden']
68	B	Olaf hörte auf , Badminton zu spielen . Schnell ging er zum	Bäcker	['ersten', 'Training', 'Sport', 'Studium', 'Badminton']
68	C	Olaf begann , Badminton zu spielen . Schnell ging er zum	Bäcker	['Studium', 'Sport', 'ersten', 'Training', 'Badminton']
69	A	Johanna verlor ihre Brille . Eilig ging sie zum	Optiker	['Arzt', 'Friseur', 'Bahnhof', 'Telefon', 'Auto']
69	B	Johanna stieg aus dem Bus aus . Eilig ging sie zum	Optiker	['Bus', 'Bahnhof', 'Auto', 'nächsten', 'Flughafen']
69	C	Johanna stieg in den Bus . Eilig ging sie zum	Optiker	['Bahnhof', 'Bus', 'Flughafen', 'nächsten', 'ersten']
70	A	Gregor wollte ein Experiment durchführen . Geschwind ging er zum	Labor	['ersten', 'Mond', 'Mars', 'Labor', 'Doktor']
70	B	Gregor stieg aus der Straßenbahn aus . Geschwind ging er zum	Labor	['Bahnhof', 'ersten', 'Auto', 'nächsten', 'Bus']
70	C	Gregor stieg in die Straßenbahn ein . Geschwind ging er zum	Labor	['Bahnhof', 'ersten', 'Straßenbahn', 'Hauptbahnhof', 'nächsten']
71	A	Martha betrat ihr Wohnzimmer . Einen Augenblick später ging sie zum	Sofa	['Fenster', 'ersten', 'Schlafzimmer', 'Badezimmer', 'Telefon']
71	B	Martha zog ihren Regenmantel aus . Einen Augenblick später ging sie zum	Sofa	['Auto', 'Strand', 'Haus', 'Fenster', 'Telefon']
71	C	Martha zog ihren Regenmantel an . Einen Augenblick später ging sie zum	Sofa	['Auto', 'Strand', 'Fenster', 'Haus', 'Telefon']
72	A	Stefan suchte nach Sternbildern . Eine Weile beobachtete er den	Himmel	['Stern', 'Sternen', 'Himmel', 'Weg', 'Mond']
72	B	Stefan verließ das Rathaus . Eine Weile beobachtete er den	Himmel	['Bürgermeister', 'Bau', 'Markt', 'Verlauf', 'Verkehr']
72	C	Stefan betrat das Rathaus . Eine Weile beobachtete er den	Himmel	['Bürgermeister', 'Markt', 'Verkehr', 'Bau', 'Eingang']
73	A	Greta suchte ihr Make - Up . Rasch ging sie zu ihrem	Schminktisch	['Freund', 'Friseur', 'Vater', 'ersten', 'Mann']
73	B	Greta kam vom Konzert zurück . Rasch ging sie zu ihrem	Schminktisch	['Vater', 'Freund', 'Mann', 'Bruder', 'ersten']
73	C	Greta erreichte das Konzert . Rasch ging sie zu ihrem	Schminktisch	['Vater', 'Freund', 'ersten', 'Mann', 'Bruder']
74	A	Paula bekam einen Strafzettel . Direkt eilte sie zum	Ordnungsamt	['Arzt', 'Auto', 'Flughafen', 'Bahnhof', 'Tierarzt']

to be continued on the next page

Table A.2.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
74	B	Paula verließ den Bauernhof . Direkt eilte sie zum	Ordnungsamt	['Arzt', 'Bahnhof', 'Haus', 'Hof', 'Stall']
74	C	Paula betrat den Bauernhof . Direkt eilte sie zum	Ordnungsamt	['Stall', 'Haus', 'Hof', 'Fenster', 'Bauernhof']
75	A	Lasse bekam eine Grippe . Unverzüglich eilte er zum	Arzt	['Arzt', 'Krankenhaus', 'Tierarzt', 'Flughafen', 'Haus']
75	B	Lasse kam von der Tanzschule . Unverzüglich eilte er zum	Arzt	['Haus', 'Hotel', 'Auto', 'Ausgang', 'Arzt']
75	C	Lasse ging zur Tanzschule . Unverzüglich eilte er zum	Arzt	['Tanz', 'Hotel', 'Haus', 'Bahnhof', 'Unterricht']
76	A	Albert betrat das Hochhaus . Schnellen Schrittes ging er zum	Aufzug	['Eingang', 'ersten', 'Aufzug', 'Fahrstuhl', 'Fenster']
76	B	Albert betrat das Einrichtungshaus . Schnellen Schrittes ging er zum	Aufzug	['Laden', 'Eingang', 'ersten', 'Geschäft', 'Kunden']
76	C	Albert verließ das Einrichtungshaus . Schnellen Schrittes ging er zum	Aufzug	['Ein', 'ersten', '.', 'neuen', 'Kauf']
77	A	Tobias erreichte den Spielplatz . Als Erstes lief er zur	Rutsche	['Schule', 'Seite', 'Toilette', 'Tür', 'Straße']
77	B	Tobias verließ den Comicladen . Als Erstes lief er zur	Rutsche	['Schule', 'Seite', 'Comic', 'Arbeit', 'Polizei']
77	C	Tobias betrat den Comicladen . Als Erstes lief er zur	Rutsche	['Seite', 'Tür', 'Arbeit', 'Schule', 'Toilette']
78	A	Anna wollte einen Cappuccino kaufen . Auf der Stelle ging sie ins	Café	['Kino', 'Bett', 'Bad', 'Büro', 'Badezimmer']
78	B	Anna verließ die Videothek . Auf der Stelle ging sie ins	Café	['Kino', 'Bett', 'Badezimmer', 'Büro', 'Bad']
78	C	Anna betrat die Videothek . Auf der Stelle ging sie ins	Café	['Badezimmer', 'Bad', 'Wohnzimmer', 'Bett', 'Kino']
79	A	Ben röstete die Bohnen . Nach einer Weile machte er einen	Kaffee	['kleinen', 'Schnitt', 'großen', 'Spaziergang', 'Teil']
79	B	Ben betrat das Haus . Nach einer Weile machte er einen	Kaffee	['Spaziergang', 'Fehler', 'sehr', 'kleinen', 'Rück']
79	C	Ben verließ das Haus . Nach einer Weile machte er einen	Kaffee	['Ausflug', 'Spaziergang', 'Fehler', 'kurzen', 'kleinen']
80	A	Julia hörte den Streik . Bald darauf sah sie die	Demonstranten	['Polizei', 'Straße', 'Frau', 'ersten', 'erste']
80	B	Julia verließ den Hort . Bald darauf sah sie die	Demonstranten	['Chance', 'Möglichkeit', 'Gelegenheit', 'Mutter', 'Welt']
80	C	Julia betrat den Hort . Bald darauf sah sie die	Demonstranten	['beiden', 'Kinder', 'Mutter', 'Leiche', 'Polizei']
81	A	Oliver feierte Silvester . Direkt sah er das	Feuerwerk	['Feuerwerk', 'erste', 'Licht', 'Foto', 'Finale']
81	B	Oliver verließ das Geschäft . Direkt sah er das	Feuerwerk	['Ende', 'Unternehmen', 'erste', 'Problem', 'Geschäft']
81	C	Oliver betrat das Geschäft . Direkt sah er das	Feuerwerk	['Bild', 'Foto', 'erste', 'Auto', 'Schild']
82	A	Elena fuhr zum Flughafen . Einen Moment später sah sie ein	Flugzeug	['Auto', 'Flugzeug', 'Taxi', 'Mädchen', 'Foto']
82	B	Elena verließ die Boutique . Einen Moment später sah sie ein	Flugzeug	['Foto', 'Mädchen', 'Bild', ',', 'Auto']
82	C	Elena betrat die Boutique . Einen Moment später sah sie ein	Flugzeug	['Foto', 'Mädchen', 'Bild', 'Auto', 'kleines']
83	A	Sebastian wollte ins Bett gehen . Eine Weile suchte er seinen	Schlafanzug	['Vater', 'Bruder', 'Freund', 'Platz', 'Arzt']
83	B	Sebastian hörte auf zu duschen . Eine Weile suchte er seinen	Schlafanzug	['Vater', 'Freund', 'Bruder', 'Hund', 'Platz']
83	C	Sebastian begann zu duschen . Eine Weile suchte er seinen	Schlafanzug	['Vater', 'Freund', 'Bruder', 'Hund', 'eigenen']
84	A	Simon wollte lesen . Unmittelbar nahm er seine	Lesebrille	['erste', 'Bücher', 'ersten', 'Arbeit', 'Mutter']
84	B	Simon kam vom Joggen zurück . Unmittelbar nahm er seine	Lesebrille	['Sachen', 'Brille', 'Medikamente', 'Waffe', 'erste']
84	C	Simon begann zu joggen . Unmittelbar nahm er seine	Lesebrille	['erste', 'ersten', 'eigene', 'Lauf', 'Frau']
85	A	Nathalie wollte eine Komödie ansehen . Fröhlich ging sie zum	Kino	['Film', 'Casting', 'Kino', 'ersten', 'Theater']
85	B	Nathalie verließ den Kosmetiksalon . Fröhlich ging sie zum	Kino	['Friseur', 'Kosmetik', 'Arzt', 'Frise', 'Zahnarzt']
85	C	Nathalie betrat den Kosmetiksalon . Fröhlich ging sie zum	Kino	['Friseur', 'Schalter', 'Badezimmer', 'Bad', 'Spiegel']
86	A	Hannes schaltete den Motor aus . Kurz darauf öffnete er die	Autotür	['Tür', 'Heck', 'Türen', 'Motor', 'Ven']
86	B	Hannes verließ den Drogeriemarkt . Kurz darauf öffnete er die	Autotür	['Filiale', 'Dro', 'Firma', 'Laden', 'Bäckerei']
86	C	Hannes betrat den Drogeriemarkt . Kurz darauf öffnete er die	Autotür	['Tür', 'Kasse', 'Haustür', 'Eingang', 'Wohnung']
87	A	Caroline wollte die Nachrichten lesen . Kurz darauf öffnete sie die	Zeitung	['Tür', 'Wohnung', 'Haustür', 'Augen', 'Türen']
87	B	Caroline war fertig damit , ihre Haare zu waschen . Kurz darauf öffnete sie die	Zeitung	['Tür', 'Augen', 'Wohnung', 'T', 'Haustür']
87	C	Caroline begann , ihre Haare zu waschen . Kurz darauf öffnete sie die	Zeitung	['Tür', 'Augen', 'Wohnung', 'Brust', 'T']
88	A	Richard musste Lebensmittel kaufen . Sogleich ging er zum	Supermarkt	['Bahnhof', 'nächsten', 'Arzt', 'Markt', 'ersten']
88	B	Richard beendete die Hausarbeit . Sogleich ging er zum	Supermarkt	['Studium', 'Schreiben', 'Haus', 'ersten', 'Unterricht']
88	C	Richard begann die Hausarbeit . Sogleich ging er zum	Supermarkt	['Studium', 'Schreiben', 'Unterricht', 'Haus', 'ersten']
89	A	Christina brauchte einen Haarschnitt . Umgehend ging sie zum	Friseur	['Friseur', 'Arzt', 'Zahnarzt', 'Casting', 'Fris']
89	B	Christina verließ die Kletterhalle . Umgehend ging sie zum	Friseur	['Kletter', 'Training', 'ersten', 'Berg', 'Sport']
89	C	Christina betrat die Kletterhalle . Umgehend ging sie zum	Friseur	['ersten', 'Lift', 'Kletter', 'Eingang', 'Sprung']
90	A	Christoph überquerte die Brücke . Eine Zeit lang betrachtete er den	Fluss	['Fluss', 'Weg', 'Ort', 'Turm', 'Berg']
90	B	Christoph verließ die Lagerhalle . Eine Zeit lang betrachtete er den	Fluss	['Krieg', 'Tod', 'Lager', 'Ort', 'Kampf']
90	C	Christoph betrat die Lagerhalle . Eine Zeit lang betrachtete er den	Fluss	['Lager', 'Laden', 'Ort', 'Tod', 'Kampf']

Table A.3.: **ADSBC21** GPT-2 top 5 predictions for the target word position.

I	Con	Target Sentence	Target	Predictions
1	A	Gestern schärfte der Holzfäller , bevor er das Holz stapelte , die	Axt	['Arbeit', 'Holz', 'Hände', 'Hand', 'Wege']
1	B	Gestern schärfte der Holzfäller , bevor er den Film schaute , die	Axt	['Sinne', 'Augen', 'Aufmerksamkeit', 'Hand', 'Grundlagen']
1	C	Gestern aß der Holzfäller , bevor er das Holz stapelte , die	Axt	['Holz', 'Rinde', 'Kartoffeln', 'Früchte', 'ganze']
1	D	Gestern aß der Holzfäller , bevor er den Film schaute , die	Axt	['Kinder', 'Sonne', '"', 'ganze', 'Holz']
2	A	Nachdenklich schärft der Barbier , nachdem er den Rasierschaum aufgetragen hat , das	Messer	['Gesicht', 'Auge', 'Haar', 'Messer', 'Mädchen']
2	B	Nachdenklich schärft der Barbier , nachdem er die Topfpflanze gegossen hat , das	Messer	['Gesicht', 'Auge', 'Haar', 'Herz', 'Wasser']
2	C	Nachdenklich trinkt der Barbier , nachdem er den Rasierschaum aufgetragen hat , das	Messer	['Wasser', 'Blut', 'Bier', 'Glas', 'ganze']
2	D	Nachdenklich trinkt der Barbier , nachdem er die Topfpflanze gegossen hat , das	Messer	['Wasser', 'ganze', 'Blut', 'Bier', 'Glas'
3	A	Sogleich süßt der Kellner , nachdem er die Bestellung aufgenommen hat , den	Kaffee	['Wein', 'Kaffee', 'Teller', 'Champagner', 'Fisch']
3	B	Sogleich süßt der Kellner , nachdem er das Radio angemacht hat , den	Kaffee	['Kaffee', 'Wein', 'Teller', 'Tee', 'Topf']
3	C	Sogleich putzt der Kellner , nachdem er die Bestellung aufgenommen hat , den	Kaffee	['Tisch', 'Boden', 'Kühlschrank', 'Teller', 'Raum']
3	D	Sogleich putzt der Kellner , nachdem er das Radio angemacht hat , den	Kaffee	['Tisch', 'Boden', 'Kühlschrank', 'Raum', 'ganzen']
4	A	Vorsichtig erhitzt der Bäckerlehrling , nachdem er die Brötchen geformt hat , den	Ofen	['Teig', 'Boden', 'Ofen', 'Brot', 'Backofen']
4	B	Vorsichtig erhitzt der Bäckerlehrling , nachdem er die Fenster gekippt hat , den	Ofen	['Teig', 'Ofen', 'Boden', 'Stein', 'Brot']
4	C	Vorsichtig schlürft der Bäckerlehrling , nachdem er die Brötchen geformt hat , den	Ofen	['Teig', 'Kaffee', 'Kuchen', 'ersten', 'Saft']
4	D	Vorsichtig schlürft der Bäckerlehrling , nachdem er die Fenster gekippt hat , den	Ofen	['Teig', 'Kaffee', 'Kuchen', 'Apfel', 'Saft']
5	A	Zufrieden entkorkt der Winzer , der die Reben geschnitten hat , die	Weinflasche	['Trauben', 'Reben', 'Wein', 'die', 'mit']
5	B	Zufrieden entkorkt der Winzer , der die Lohnabrechnungen beendet hat , die	Weinflasche	['Wein', 'Trauben', 'er', 'Weine', 'Winzer']
5	C	Zufrieden glättet der Winzer , der die Reben geschnitten hat , die	Weinflasche	['Trauben', 'Farbe', 'Wein', 'Blätter', 'Weine']
5	D	Zufrieden glättet der Winzer , der die Lohnabrechnungen beendet hat , die	Weinflasche	['Wein', 'Rechnungen', 'Trauben', 'Arbeit', 'Weine']
6	A	Umgehend löst der Fahrgast , der den Automaten bedient , das	Ticket	['Ticket', 'Geld', 'Problem', 'Schloss', 'Licht']
6	B	Umgehend löst der Fahrgast , der die Einkaufstasche abstellt , das	Ticket	['Ticket', 'Problem', 'Auto', 'Schloss', 'Handy']
6	C	Umgehend knotet der Fahrgast , der den Automaten bedient , das	Ticket	['Ticket', 'Gerät', 'Geld', 'Glas', 'Handy']
6	D	Umgehend knotet der Fahrgast , der die Einkaufstasche abstellt , das	Ticket	['Gepäck', 'Ticket', 'Auto', 'Handy', 'Fahrrad']
7	A	Am Abend feiert der Gewinner , der den Pokal erhält , den	Sieg	['Geburtstag', 'Tag', 'Abend', 'Sieg', 'Abschluss']
7	B	Am Abend feiert der Gewinner , der das Eis isst , den	Sieg	['Geburtstag', 'Tag', 'Abschluss', 'Abend', 'Sieg']
7	C	Am Abend knabbert der Gewinner , der den Pokal erhält , den	Sieg	['Pokal', 'Preis', 'Ball', 'Schlüssel', 'Kuchen']
7	D	Am Abend knabbert der Gewinner , der das Eis isst , den	Sieg	['Kuchen', 'Preis', 'Eis', 'ganzen', 'Becher']
8	A	Endlich fängt der Angler , der den Köder auswarf , den	Fisch	['Fisch', 'kleinen', 'Vogel', 'Hai', 'Hund']
8	B	Endlich fängt der Angler , der die Zeitung las , den	Fisch	['Fisch', 'kleinen', 'ersten', 'Hund', 'Vogel']
8	C	Endlich baut der Angler , der den Köder auswarf , den	Fisch	['Fisch', 'K', 'Angel', 'Haken', 'kleinen']
8	D	Endlich baut der Angler , der die Zeitung las , den	Fisch	['Fisch', 'ersten', 'Zaun', 'kleinen', 'Teich']
9	A	Am Nachmittag jätete der Schrebergärtner , der das Beet pflegte , das	Unkraut	['Gemüse', 'die', 'Unkraut', 'Beet', 'erste']
9	B	Am Nachmittag jätete der Schrebergärtner , der die Sonne genoss , das	Unkraut	['Gelände', 'Haus', 'Grün', 'Grundstück', 'Feld']
9	C	Am Nachmittag polierte der Schrebergärtner , der das Beet pflegte , das	Unkraut	['Beet', 'Grün', 'Gemüse', 'Obst', 'Haus']
9	D	Am Nachmittag polierte der Schrebergärtner , der die Sonne genoss , das	Unkraut	['Grün', 'Grundstück', 'Gelände', 'Haus', 'Dach']
10	A	Eine Weile lüftet der Lehrer , bevor er die Tafel beschreibt , das	Klassenzimmer	['Geheimnis', 'Rätsel', 'Bild', 'Wort', 'Buch']
10	B	Eine Weile lüftet der Lehrer , bevor er den Mantel aufhängt , das	Klassenzimmer	['Geheimnis', 'Gesicht', 'Licht', 'Zimmer', 'Bett']
10	C	Eine Weile bastelt der Lehrer , bevor er die Tafel beschreibt , das	Klassenzimmer	['ist', 'Bild', 'Buch', 'Wort', 'heißt']
10	D	Eine Weile bastelt der Lehrer , bevor er den Mantel aufhängt , das	Klassenzimmer	['ist', 'Bild', 'erste', 'war', 'heißt']
11	A	Schnell reibt der Chefkoch , während er die Nudeln kocht , den	Käse	['Teig', 'Reis', 'Deckel', 'Boden', 'Fisch']
11	B	Schnell reibt der Chefkoch , während er die Nachrichten hört , den	Käse	['Teig', 'Fisch', 'Teller', 'Inhalt', 'Mund']
11	C	Schnell flickt der Chefkoch , während er die Nudeln kocht , den	Käse	['Teig', 'Topf', 'Teller', 'Reis', 'Boden']
11	D	Schnell flickt der Chefkoch , während er die Nachrichten hört , den	Käse	['Kühlschrank', 'Tisch', 'Chef', 'Koch', 'Teig']
12	A	Vorgestern probierte der Braumeister , nachdem er den Kessel ausgeschaltet hatte , das	Bier	['Bier', 'erste', 'neue', 'Lager', '"']
12	B	Vorgestern probierte der Braumeister , nachdem er den Boden gewischt hatte , das	Bier	['Bier', 'erste', 'neue', '"', 'Lager']
12	C	Vorgestern bemalte der Braumeister , nachdem er den Kessel ausgeschaltet hatte , das	Bier	['Holz', 'Bild', 'Bier', 'Innere', 'Glas']
12	D	Vorgestern bemalte der Braumeister , nachdem er den Boden gewischt hatte , das	Bier	['Holz', 'Bier', 'Bild', 'Glas', 'alte']
13	A	Eine Zeit lang mähte der Gartenhelfer , nachdem er den Garten umgegraben hatte , den	Rasen	['Rasen', 'Boden', 'Garten', 'Baum', 'Weg']
13	B	Eine Zeit lang mähte der Gartenhelfer , nachdem er die Garage aufgeräumt hatte , den	Rasen	['Rasen', 'Garten', 'Boden', 'Platz', 'Zaun']
13	C	Eine Zeit lang räucherte der Gartenhelfer , nachdem er den Garten umgegraben hatte , den	Rasen	['Garten', 'Rasen', 'alten', 'Baum', 'Geist']
13	D	Eine Zeit lang räucherte der Gartenhelfer , nachdem er die Garage aufgeräumt hatte , den	Rasen	['Garten', 'Rasen', 'Müll', 'alten', 'kleinen']
14	A	Achtsam pflasterte der Bauarbeiter , der die Absperrungen umging , die	Straße	['Straße', 'Fahrbahn', 'Straßen', 'sich', 'die']
14	B	Achtsam pflasterte der Bauarbeiter , der die Steine schleppte , die	Straße	['Straße', 'Straßen', 'Wände', 'mit', 'Steine']
14	C	Achtsam salzte der Bauarbeiter , der die Absperrungen umging , die	Straße	['sich', 'die', 'den', 'mit', 'in']
14	D	Achtsam salzte der Bauarbeiter , der die Steine schleppte , die	Straße	['Steine', 'die', 'mit', 'Kinder', 'Arbeiter']
15	A	Gestern läutete der Priester , nachdem er den Kirchturm erklommen hatte , die	Glocken	['Glocke', 'Glocken', 'Tür', 'Pfor', 'Kirchen']
15	B	Gestern läutete der Priester , nachdem er die Armbanduhr ausgezogen hatte , die	Glocken	['Glocke', 'Tür', 'Glocken', 'Uhr', 'Türen']
15	C	Gestern faltete der Priester , nachdem er den Kirchturm erklommen hatte , die	Glocken	['Glocken', 'Hände', 'Glocke', 'Steine', 'Mauer']
15	D	Gestern faltete der Priester , nachdem er die Armbanduhr ausgezogen hatte , die	Glocken	['Uhr', 'Hand', 'Hände', 'Armb', 'Kette']
16	A	Rasch näht der Notarzt , der den Verletzten betreut , die	Wunde	['Wunden', 'Wunde', 'Beine', 'Kleidung', 'Arme']
16	B	Rasch näht der Notarzt , der den Ohrring trägt , die	Wunde	['Haare', 'Ohr', 'Hand', 'Ärmel', 'Augen']
16	C	Rasch versteckt der Notarzt , der den Verletzten betreut , die	Wunde	['Verletzten', 'Leiche', 'beiden', 'verletzte', 'Patienten']
16	D	Rasch versteckt der Notarzt , der den Ohrring trägt , die	Wunde	['Leiche', 'Tasche', 'beiden', 'Frau', 'Waffe']
17	A	Eilig bügelt der Geschäftsmann , nachdem er das Bügelbrett aufgestellt hat , das	Hemd	['Bügel', 'sich', 'in', 'er', 'auf']
17	B	Eilig bügelt der Geschäftsmann , nachdem er den Eistee ausgetrunken hat , das	Hemd	['Geschäft', 'Glas', 'Geschirr', 'Bett', 'Auto']
17	C	Eilig salzt der Geschäftsmann , nachdem er das Bügelbrett aufgestellt hat , das	Hemd	['er', 'sich', 'die', 'in', 'Bügel']
17	D	Eilig salzt der Geschäftsmann , nachdem er den Eistee ausgetrunken hat , das	Hemd	['Bier', 'Essen', 'Wasser', 'Eis', 'ganze']
18	A	Schnell repariert der KFZ - Mechatroniker , der den Motor ausgetauscht hat , das	Auto	['Auto', 'Getriebe', 'Fahrzeug', 'Motorrad', 'Rad']
18	B	Schnell repariert der KFZ - Mechatroniker , der den Urlaub gebucht hat , das	Auto	['Auto', 'Fahrzeug', 'Motorrad', 'Fahrrad', 'Haus']
18	C	Schnell brät der KFZ - Mechatroniker , der den Motor ausgetauscht hat , das	Auto	['Auto', 'Motorrad', 'Fahrzeug', 'Fahrrad', 'Rad']
18	D	Schnell brät der KFZ - Mechatroniker , der den Urlaub gebucht hat , das	Auto	['Auto', 'Fahrrad', 'Essen', 'Fleisch', 'neue']

to be continued on the next page

Table A.3.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
19	A	Am Morgen knetet die Bäckerin , die den Kuchen backt , den	Teig	['Kuchen', 'Teig', 'Kopf', 'Ofen', 'Mund']
19	B	Am Morgen knetet die Bäckerin , die den Stammgast begrüßt , den	Teig	['Stamm', 'Teig', 'ersten', 'Baum', 'Kopf']
19	C	Am Morgen hört die Bäckerin , die den Kuchen backt , den	Teig	['Mann', 'Hund', 'Herrn', 'ersten', 'Lärm']
19	D	Am Morgen hört die Bäckerin , die den Stammgast begrüßt , den	Teig	['ersten', 'Gesang', 'Hund', 'Herrn', 'Vogel']
20	A	Am Nachmittag pfeffert der Grillmeister , bevor er die Kohle anzündet , das	Steak	['Feuer', 'Holz', 'erste', 'Gas', 'letzte']
20	B	Am Nachmittag pfeffert der Grillmeister , bevor er die Mücke verscheucht , das	Steak	['Fleisch', 'Wild', 'Grill', 'Feuer', 'Essen']
20	C	Am Nachmittag druckt der Grillmeister , bevor er die Kohle anzündet , das	Steak	['Feuer', 'Bild', 'erste', 'Holz', 'Grill']
20	D	Am Nachmittag druckt der Grillmeister , bevor er die Mücke verscheucht , das	Steak	['Fleisch', 'Grill', '"', 'erste', 'Buch']
21	A	Schnell entfacht der Brandstifter , der das Streichholz fallengelassen hat , das	Feuer	['Feuer', 'Chaos', 'Haus', 'ganze', 'Licht']
21	B	Schnell entfacht der Brandstifter , der die Treppe hochgeklettert ist , das	Feuer	['Feuer', 'Chaos', 'Haus', 'ganze', 'gesamte']
21	C	Schnell entstaubt der Brandstifter , der das Streichholz fallengelassen hat , das	Feuer	['Feuer', 'die', 'Haus', 'sich', 'den']
21	D	Schnell entstaubt der Brandstifter , der die Treppe hochgeklettert ist , das	Feuer	['Haus', 'Feuer', 'Gebäude', 'Dach', 'Auto']
22	A	Nun kämmt der Friseur , nachdem er die Kopfhaut massiert hat , die	Haare	['Haare', 'Haut', 'Augen', 'Kopf', 'Lippen']
22	B	Nun kämmt der Friseur , nachdem er den Cappuccino ausgetrunken hat , die	Haare	['Haare', 'Augen', 'Haut', 'Hände', 'Lippen']
22	C	Nun speichert der Friseur , nachdem er die Kopfhaut massiert hat , die	Haare	['Haare', 'Haut', 'Kopf', 'Haar', 'Farbe']
22	D	Nun speichert der Friseur , nachdem er den Cappuccino ausgetrunken hat , die	Haare	['Haare', 'Zeit', 'Farbe', 'ganze', 'Ergebnisse']
23	A	Flink spitzte der Zeichner , der die Karikatur entwarf , den	Bleistift	['Text', 'Rahmen', 'Bogen', 'Kopf', 'Inhalt']
23	B	Flink spitzte der Zeichner , der den Atlas öffnete , den	Bleistift	['Bogen', 'Kopf', 'Text', 'Blick', 'Rahmen']
23	C	Flink knotete der Zeichner , der die Karikatur entwarf , den	Bleistift	['Text', 'Rahmen', 'Bogen', 'Karton', 'Kopf']
23	D	Flink knotete der Zeichner , der den Atlas öffnete , den	Bleistift	['Bogen', 'Text', 'Rahmen', 'Kopf', 'ersten']
24	A	Heute impft der Arzt , nachdem er die Spritze befüllt hat , den	Patienten	['Patienten', 'Blut', 'Patient', 'Arzt', 'Hund']
24	B	Heute impft der Arzt , nachdem er die Schublade geschlossen hat , den	Patienten	['Patienten', 'Kindern', 'Hund', 'Mann', 'Menschen']
24	C	Heute feuert der Arzt , nachdem er die Spritze befüllt hat , den	Patienten	['Patienten', 'Mann', 'Arzt', 'Krankenwagen', 'Jungen']
24	D	Heute feuert der Arzt , nachdem er die Schublade geschlossen hat , den	Patienten	['Patienten', 'Mann', 'Schlüssel', 'Arzt', 'Rollstuhl']
25	A	Unverzüglich obduziert der Pathologe , nachdem er die Mordakte gelesen hat , die	Leiche	['Leiche', 'Tat', 'Todesursache', 'Mord', 'er']
25	B	Unverzüglich obduziert der Pathologe , nachdem er das Dokument unterschrieben hat , die	Leiche	['Leiche', 'Todesursache', 'beiden', 'Person', 'Blut']
25	C	Unverzüglich pfeffert der Pathologe , nachdem er die Mordakte gelesen hat , die	Leiche	['Leiche', 'Leichen', 'Tat', 'Blut', 'Mord']
25	D	Unverzüglich pfeffert der Pathologe , nachdem er das Dokument unterschrieben hat , die	Leiche	['Blut', 'Leiche', 'Wunde', 'ster', 'Probe']
26	A	Am Nachmittag signiert der Autor , der die Geschichte erfunden hat , das	Buch	['Buch', 'Werk', 'neue', 'Bild', 'erste']
26	B	Am Nachmittag signiert der Autor , der den Vorhang zugezogen hat , das	Buch	['Buch', 'neue', 'Werk', 'Bild', 'Manuskript']
26	C	Am Nachmittag schluckt der Autor , der die Geschichte erfunden hat , das	Buch	['Buch', 'erste', 'Wort', 'Thema', '"']
26	D	Am Nachmittag schluckt der Autor , der den Vorhang zugezogen hat , das	Buch	['Buch', 'erste', 'Thema', 'Gedicht', 'Video']
27	A	Sofort begrüßte der Hotelier , der die Koffer stapelte , die	Gäste	['Gäste', 'Passagiere', 'Besucher', 'vielen', 'Reisenden']
27	B	Sofort begrüßte der Hotelier , der die Fische räucherte , die	Gäste	['Gäste', 'Besucher', 'Teilnehmer', 'zahlreichen', 'neuen']
27	C	Sofort feuerte der Hotelier , der die Koffer stapelte , die	Gäste	['Polizei', 'Passagiere', 'Bombe', 'Gäste', 'er']
27	D	Sofort feuerte der Hotelier , der die Fische räucherte , die	Gäste	['Fische', 'Flammen', 'Fischer', 'Gäste', 'Polizei']
28	A	Vorsichtig fällt der Holzarbeiter , der die Säge gestartet hat , den	Baum	['Boden', 'Baum', 'Holz', 'Weg', 'Stein']
28	B	Vorsichtig fällt der Holzarbeiter , der den Transporter geparkt hat , den	Baum	['Weg', 'Boden', 'Baum', 'Schlüssel', 'Stein']
28	C	Vorsichtig durchwühlt der Holzarbeiter , der die Säge gestartet hat , den	Baum	['Boden', 'Wald', 'Holz', 'Baum', 'Weg']
28	D	Vorsichtig durchwühlt der Holzarbeiter , der den Transporter geparkt hat , den	Baum	['Boden', 'Wagen', 'Container', 'Zaun', 'Weg']
29	A	Langsam paffte der Kubaner , der den Rauchring blies , die	Zigarre	['ersten', 'erste', 'Zigar', 'Luft', 'Welt']
29	B	Langsam paffte der Kubaner , der die Straßenbahn verpasste , die	Zigarre	['Fahrgäste', 'ersten', 'Kinder', 'Menschen', 'Strecke']
29	C	Langsam öffnete der Kubaner , der den Rauchring blies , die	Zigarre	['Tür', 'Augen', 'Türen', 'Luft', 'Wohnung']
29	D	Langsam öffnete der Kubaner , der die Straßenbahn verpasste , die	Zigarre	['Tür', 'Türen', 'Augen', 'Straße', 'Eingang']
30	A	Fröhlich paniert der Koch , nachdem er die Bratkartoffeln geschält hat , das	Schnitzel	['Fleisch', 'Gemüse', 'Brot', 'Gericht', 'Ei']
30	B	Fröhlich paniert der Koch , nachdem er die Ärmel hochgekrempelt hat , das	Schnitzel	['Fleisch', 'Brot', 'Gemüse', 'Gericht', 'Essen']
30	C	Fröhlich baut der Koch , nachdem er die Bratkartoffeln geschält hat , das	Schnitzel	['Brot', 'Gemüse', 'Fleisch', 'Gericht', 'Essen']
30	D	Fröhlich baut der Koch , nachdem er die Ärmel hochgekrempelt hat , das	Schnitzel	['Essen', 'Brot', 'Frühstück', 'Geschirr', 'erste']
31	A	Umsichtig glasiert der Konditor , der die Backform eingefettet hat , den	Kuchen	['Teig', 'Kuchen', 'Zucker', 'Inhalt', 'Boden']
31	B	Umsichtig glasiert der Konditor , der den Tisch abgewischt hat , den	Kuchen	['Teig', 'Kuchen', 'Boden', 'Ofen', 'Tisch']
31	C	Umsichtig durchsucht der Konditor , der die Backform eingefettet hat , den	Kuchen	['Teig', 'Kuchen', 'Inhalt', 'Raum', 'Back']
31	D	Umsichtig durchsucht der Konditor , der den Tisch abgewischt hat , den	Kuchen	['Tisch', 'Raum', 'Boden', 'Teller', 'Teig']
32	A	Gestern ölte der Fahrradfahrer , bevor er die Bremse einstellte , die	Kette	['Brem', 'Bremsen', 'Räder', 'Straße', 'Rad']
32	B	Gestern ölte der Fahrradfahrer , bevor er die Limonade trank , die	Kette	['er', 'ihm', 'ihn', 'sich', 'in']
32	C	Gestern beheizte der Fahrradfahrer , bevor er die Bremse einstellte , die	Kette	['Brem', 'Rad', 'Räder', 'Straße', 'Vorder']
32	D	Gestern beheizte der Fahrradfahrer , bevor die Limonade trank , die	Kette	['Frau', 'sich', 'Kinder', 'mit', 'die']
33	A	Hastig frankiert der Postbeamte , der das Postamt aufgeschlossen hat , den	Brief	['Brief', 'Post', 'Umschlag', 'Namen', 'Inhalt']
33	B	Hastig frankiert der Postbeamte , der das Mittagessen bestellt hat , den	Brief	['Brief', 'Post', 'nächsten', 'ganzen', 'Kaffee']
33	C	Hastig umgeht der Postbeamte , der das Postamt aufgeschlossen hat , den	Brief	['Post', 'Beamten', 'Brief', 'Laden', 'Mann']
33	D	Hastig umgeht der Postbeamte , der das Mittagessen bestellt hat , den	Brief	['Post', 'Brief', 'Mann', 'Laden', 'Chef']
34	A	Fröhlich erklimmt der Wanderer , nachdem er das Tal durchquert hat , den	Berg	['Gipfel', 'höchsten', 'Berg', 'Weg', 'Aussicht']
34	B	Fröhlich erklimmt der Wanderer , nachdem er die Stoppuhr gestartet hat , den	Berg	['Gipfel', 'Berg', 'höchsten', 'Weg', 'oberen']
34	C	Fröhlich kauft der Wanderer , nachdem er das Tal durchquert hat , den	Berg	['Weg', 'Berg', 'Stein', 'Wald', '"']
34	D	Fröhlich kauft der Wanderer , nachdem er die Stoppuhr gestartet hat , den	Berg	['Rucksack', 'Berg', 'Stein', 'Weg', 'Schlüssel']
35	A	Bedächtig erteilt der General , der die Rekruten kommandiert , den	Befehl	['Befehl', 'Auftrag', 'Anweisungen', 'Soldaten', 'Offizieren']
35	B	Bedächtig erteilt der General , der die Getränke einschenkt , den	Befehl	['Befehl', 'Gästen', 'Anwesenden', 'Anweisungen', 'Blick']
35	C	Bedächtig biegt der General , der die Rekruten kommandiert , den	Befehl	['Weg', 'rechten', 'Bogen', 'Hügel', 'Pfad']
35	D	Bedächtig biegt der General , der die Getränke einschenkt , den	Befehl	['Weg', 'Arm', 'rechten', 'Wagen', 'linken']
36	A	Sofort streicht der Maler , nachdem er die Tapete angebracht hat , die	Wand	['Wände', 'Tap', 'Farbe', 'Wand', 'Bilder']

to be continued on the next page

Table A.3.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
36	B	Sofort streicht der Maler , nachdem er den Hund gefüttert hat , die	Wand	['Haare', 'Augen', 'Haut', 'Beine', 'Zähne']
36	C	Sofort lenkt der Maler , nachdem er die Tapete angebracht hat , die	Wand	['Aufmerksamkeit', 'Augen', 'Blicke', 'Pinsel', 'Kamera']
36	D	Sofort lenkt der Maler , nachdem er den Hund gefüttert hat , die	Wand	['Aufmerksamkeit', 'Augen', 'Kamera', 'Blicke', 'Katze']
37	A	Direkt schwingt der Schmied , der den Amboss aufgestellt hat , den	Hammer	['Hammer', 'Amb', 'Stab', 'Eisen', 'Kopf']
37	B	Direkt schwingt der Schmied , der den Stuhl weggestellt hat , den	Hammer	['Kopf', 'Stuhl', 'Hammer', 'Arm', 'Rücken']
37	C	Direkt speichert der Schmied , der den Amboss aufgestellt hat , den	Hammer	['Inhalt', 'Hammer', 'Rest', 'Eisen', 'fertigen']
37	D	Direkt speichert der Schmied , der den Stuhl weggestellt hat , den	Hammer	['Inhalt', 'Rest', 'Gegenstand', 'Stuhl', 'Stein']
38	A	Vor langer Zeit regierte der König , der die Reise unternahm , das	Land	['Land', 'Königreich', 'Reich', 'ganze', 'Volk']
38	B	Vor langer Zeit regierte der König , der den Gesang vernahm , das	Land	['Land', 'Königreich', 'Reich', 'Volk', 'ganze']
38	C	Vor langer Zeit briet der König , der die Reise unternahm , das	Land	['erste', 'Brot', 'Land', 'Buch', 'Kind']
38	D	Vor langer Zeit briet der König , der den Gesang vernahm , das	Land	['Lied', 'Buch', 'Brot', 'Kind', 'Fleisch']
39	A	Zügig hob der Dirigent , der das Tempo wechselte , den	Stab	['Takt', 'Bogen', 'ersten', 'Satz', 'Dirigenten']
39	B	Zügig hob der Dirigent , der das Gähnen unterdrückte , den	Stab	['Takt', 'Stab', 'Bogen', 'Finger', 'Dirigenten']
39	C	Zügig verdrückte der Dirigent , der das Tempo wechselte , den	Stab	['Takt', 'ersten', 'Blick', 'Klang', 'Satz']
39	D	Zügig verdrückte der Dirigent , der das Gähnen unterdrückte , den	Stab	['Blick', 'Takt', 'ersten', 'ganzen', 'Klang']
40	A	Direkt trimmt der Herrenfriseur , nachdem er die Klinge ausgepackt hat , den	Bart	['Hals', 'Ras', 'Körper', 'Griff', 'Mund']
40	B	Direkt trimmt der Herrenfriseur , nachdem er das Wasserglas abgestellt hat , den	Bart	['Körper', 'Hals', 'Ras', 'Mund', 'Mann']
40	C	Direkt hört der Herrenfriseur , nachdem er die Klinge ausgepackt hat , den	Bart	['Herrn', 'ersten', 'Hund', 'Ton', '"'"]
40	D	Direkt hört der Herrenfriseur , nachdem er das Wasserglas abgestellt hat , den	Bart	['Herrn', 'Hund', 'Ton', 'ersten', 'Gesang']
41	A	Gestern spülte der Tellerwäscher , bevor er den Tisch wischte , das	Geschirr	['Geschirr', 'Wasser', 'Geld', 'ganze', 'Glas']
41	B	Gestern spülte der Tellerwäscher , bevor er das Taxi rief , das	Geschirr	['ihn', 'er', 'die', 'Taxi', 'in']
41	C	Gestern glättete der Tellerwäscher , bevor er den Tisch wischte , das	Geschirr	['Geschirr', 'Glas', 'Gesicht', 'Fleisch', 'Brot']
41	D	Gestern glättete der Tellerwäscher , bevor er das Taxi rief , das	Geschirr	['ihn', 'er', 'Taxi', 'die', 'sich']
42	A	Im Handumdrehen knackt der Einbrecher , der den Zahlencode erraten hat , das	Schloss	['Schloss', 'Passwort', 'Geld', 'Internet', 'Haus']
42	B	Im Handumdrehen knackt der Einbrecher , der die Akte durchblättert hat , das	Schloss	['Schloss', 'Internet', 'Schließ', 'Gesetz', 'Haus']
42	C	Im Handumdrehen kopiert der Einbrecher , der den Zahlencode erraten hat , das	Schloss	['Bild', 'Video', 'Geld', 'Opfer', 'gesamte']
42	D	Im Handumdrehen kopiert der Einbrecher , der die Akte durchblättert hat , das	Schloss	['Bild', 'Foto', 'gesamte', 'Video', 'Buch']
43	A	Routiniert hackte der Küchengehilfe , bevor er die Suppe umrührte , die	Zwiebeln	['er', 'Suppe', 'sich', 'mit', 'die']
43	B	Routiniert hackte der Küchengehilfe , bevor er das Radio einschaltete , die	Zwiebeln	['Tür', 'Küchen', 'Tastatur', 'Kaffeemaschine', 'Tasten']
43	C	Routiniert schraubte der Küchengehilfe , bevor er die Suppe umrührte , die	Zwiebeln	['er', 'Suppe', 'mit', 'in', 'Kartoffeln']
43	D	Routiniert schraubte der Küchengehilfe , bevor er das Radio einschaltete , die	Zwiebeln	['Musik', 'Lautsprecher', 'Uhr', 'Tasten', 'Tische']
44	A	Umsichtig schmolz der Kerzenzieher , nachdem er den Docht gekürzt hatte , das	Wachs	['Holz', 'Wachs', 'Papier', 'Seil', 'Glas']
44	B	Umsichtig schmolz der Kerzenzieher , nachdem er das Nähkästchen zugeklappt hatte , das	Wachs	['sich', 'er', 'die', 'mit', 'in']
44	C	Umsichtig würzte der Kerzenzieher , nachdem er den Docht gekürzt hatte , das	Wachs	['Holz', 'Feuer', 'Wachs', 'Glas', 'Tuch']
44	D	Umsichtig würzte der Kerzenzieher , nachdem er das Nähkästchen zugeklappt hatte , das	Wachs	['mit', 'Tuch', 'in', 'sich', 'Stück']
45	A	Gestern mixte der Barkeeper , der den Wodka geöffnet hatte , den	Cocktail	['Wod', 'Drink', 'Wein', 'Rest', 'Schnaps']
45	B	Gestern mixte der Barkeeper , der den Boden gefegt hatte , den	Cocktail	['Wein', 'Kaffee', 'Kuchen', 'ganzen', 'Drink']
45	C	Gestern gründete der Barkeeper , der den Wodka geöffnet hatte , den	Cocktail	['Verein', 'Club', '"'", 'Klub', 'ersten']
45	D	Gestern gründete der Barkeeper , der den Boden gefegt hatte , den	Cocktail	['Verein', 'Club', '"'", 'Klub', 'ersten']
46	A	Am Abend dirigiert der Kapellmeister , der das Konzert eröffnet hat , das	Orchester	['Orchester', 'Publikum', 'Konzert', 'erste', '"'"]
46	B	Am Abend dirigiert der Kapellmeister , der die Unterschriften gesammelt hat , das	Orchester	['Orchester', 'Konzert', 'erste', 'Publikum', '"'"]
46	C	Am Abend vernichtet der Kapellmeister , der das Konzert eröffnet hat , das	Orchester	['Orchester', 'Konzert', 'Publikum', 'Haus', 'gesamte']
46	D	Am Abend vernichtet der Kapellmeister , der die Unterschriften gesammelt hat , das	Orchester	['Schiff', 'Haus', 'Orchester', 'Lied', 'Manuskript']
47	A	Direkt buchstabiert der Sprachschüler , nachdem er das Buch aufgeschlagen hat , das	Wort	['Wort', 'Buch', 'in', 'Thema', 'er']
47	B	Direkt buchstabiert der Sprachschüler , nachdem er den Pausenhof verlassen hat , das	Wort	['Wort', 'Ergebnis', 'richtige', ',', 'erste']
47	C	Direkt dekoriert der Sprachschüler , nachdem er das Buch aufgeschlagen hat , das	Wort	['er', 'Buch', 'ihm', 'in', 'mit']
47	D	Direkt dekoriert der Sprachschüler , nachdem er den Pausenhof verlassen hat , das	Wort	['Gebäude', 'Zimmer', 'Schulgebäude', 'Klassenzimmer', 'Haus']
48	A	Vorsichtig umzäunt der Schäfer , der die Schafe gehütet hat , die	Weide	['Schafe', 'Herde', 'Weide', 'Tiere', 'St']
48	B	Vorsichtig umzäunt der Schäfer , der das Mittagessen vergessen hat , die	Weide	['Schafe', 'Herde', 'Tiere', 'Hunde', 'Schweine']
48	C	Vorsichtig fälscht der Schäfer , der die Schafe gehütet hat , die	Weide	['Schafe', 'Herde', 'Wolle', 'Weide', 'von']
48	D	Vorsichtig fälscht der Schäfer , der das Mittagessen vergessen hat , die	Weide	['Schafe', 'Hunde', 'Nachricht', 'Eier', 'Herde']
49	A	Nachdenklich hält der Pfarrer , nachdem er die Bibel aufgeschlagen hat , die	Predigt	['Bibel', 'Worte', 'Predigt', 'Kirche', 'Hand']
49	B	Nachdenklich hält der Pfarrer , nachdem er die Brille aufgesetzt hat , die	Predigt	['er', 'sich', 'ihm', 'Brille', 'Hände']
49	C	Nachdenklich leert der Pfarrer , nachdem er die Bibel aufgeschlagen hat , die	Predigt	['Bibel', 'Tür', 'Kirchen', 'Bücher', 'Hände']
49	D	Nachdenklich leert der Pfarrer , nachdem er die Brille aufgesetzt hat , die	Predigt	['er', 'Brille', 'sich', 'Augen', 'in']
50	A	Unverzüglich möblierte der Vermieter , der den Mieter aufgenommen hatte , die	Wohnung	['Wohnung', 'Unterkunft', 'Ferienwohnung', 'Kaution', 'Zimmer']
50	B	Unverzüglich möblierte der Vermieter , der die Überweisung erhalten hatte , die	Wohnung	['Wohnung', 'die', 'den', 'sich', 'in']
50	C	Unverzüglich erntete der Vermieter , der den Mieter aufgenommen hatte , die	Wohnung	['Zustimmung', 'ersten', 'Kritik', 'Aufmerksamkeit', '"'"]
50	D	Unverzüglich erntete der Vermieter , der die Überweisung erhalten hatte , die	Wohnung	['Früchte', 'Aufmerksamkeit', 'Zustimmung', 'Kritik', '"'"]
51	A	Sofort betätigt der Killer , der das Ziel anvisiert hat , den	Abzug	['Schalter', 'Hebel', 'Knopf', 'Wagen', 'Schlüssel']
51	B	Sofort betätigt der Killer , der den Tee abgestellt hat , den	Abzug	['Tee', 'Schalter', 'Knopf', 'Boden', 'er']
51	C	Sofort knickt der Killer , der das Ziel anvisiert hat , den	Abzug	['Kopf', 'Weg', 'Boden', 'rechten', 'Hals']
51	D	Sofort knickt der Killer , der den Tee abgestellt hat , den	Abzug	['Kopf', 'Boden', 'Tisch', 'Weg', 'Mund']
52	A	Heute überbringt der Bote , als er den Empfänger angetroffen hat , das	Paket	['Wort', 'Geschenk', 'Zeichen', 'Signal', 'Versprechen']
52	B	Heute überbringt der Bote , als er das Foyer durchquert hat , das	Paket	['Bild', 'Wort', 'von', 'Evangelium', 'die']
52	C	Heute erntet der Bote , als er den Empfänger angetroffen hat , das	Paket	['gleiche', 'Wort', 'Zeichen', 'Echo', 'erste']
52	D	Heute erntet der Bote , als er das Foyer durchquert hat , das	Paket	['Wort', 'Bild', 'von', 'gleiche', 'erste']

to be continued on the next page

Table A.3.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
53	A	Schnell schoss der Fußballspieler , der das Fußballfeld überquert hatte , das	Tor	['Tor', 'Spiel', 'erste', 'Foul', 'Tier']
53	B	Schnell schoss der Fußballspieler , der die Jacke ausgezogen hatte , das	Tor	['Tor', 'Spiel', 'erste', 'Team', 'Opfer']
53	C	Schnell spürte der Fußballspieler , der das Fußballfeld überquert hatte , das	Tor	['Feuer', 'Un', 'Unglück', 'Ausmaß', 'Auf']
53	D	Schnell spürte der Fußballspieler , der die Jacke ausgezogen hatte , das	Tor	['Feuer', 'Fehlen', 'Un', 'Problem', 'Ausmaß']
54	A	Konzentriert steuerte der Pilot , der die Stewardess gerufen hatte , das	Flugzeug	['Flugzeug', 'Schiff', 'Fahrzeug', 'Cockpit', 'Ziel']
54	B	Konzentriert steuerte der Pilot , der den Krimi gelesen hatte , das	Flugzeug	['Schiff', 'Flugzeug', 'Fahrzeug', 'Ziel', 'Auto']
54	C	Konzentriert faltete der Pilot , der die Stewardess gerufen hatte , das	Flugzeug	['Flugzeug', 'Gepäck', 'Seil', 'Schiff', 'Steuer']
54	D	Konzentriert faltete der Pilot , der den Krimi gelesen hatte , das	Flugzeug	['Buch', 'Bild', 'Manuskript', 'Material', 'Papier']
55	A	Rasch entwarf der Architekt , bevor er den Bauplan erstellte , die	Skizze	['Bau', 'Pläne', 'erste', '"', 'Innen']
55	B	Rasch entwarf der Architekt , bevor er den Vortrag schrieb , die	Skizze	['Idee', 'Entwürfe', 'erste', '"', 'Architektur']
55	C	Rasch schüttelte der Architekt , bevor er den Bauplan erstellte , die	Skizze	['Hände', 'ersten', 'Bau', 'Worte', 'Hand']
55	D	Rasch schüttelte der Architekt , bevor er den Vortrag schrieb , die	Skizze	['Hände', 'Worte', 'Gedanken', 'Hand', 'Köpfe']
56	A	Geschwind kehrt der Schornsteinfeger , der den Rauch einatmet , den	Kamin	['Schornstein', 'Rauch', 'Ofen', 'Weg', 'Kamin']
56	B	Geschwind kehrt der Schornsteinfeger , der die Katze verscheucht , den	Kamin	['Schornstein', 'Weg', 'Turm', 'Ofen', 'Berg']
56	C	Geschwind wirft der Schornsteinfeger , der den Rauch einatmet , den	Kamin	['Schornstein', 'Rauch', 'Staub', 'Ofen', 'Ruß']
56	D	Geschwind wirft der Schornsteinfeger , der die Katze verscheucht , den	Kamin	['Schornstein', 'letzten', 'ersten', 'Rauch', 'Sack']
57	A	Vorgestern gewann der Politiker , der die Stimmen zählte , die	Wahl	['Wahl', 'Wahlen', 'Präsidentschafts', 'Abstimmung', 'die']
57	B	Vorgestern gewann der Politiker , der die Hände schüttelte , die	Wahl	['Wahl', 'Wahlen', 'erste', 'Stichwahl', 'Präsidentschafts']
57	C	Vorgestern reparierte der Politiker , der die Stimmen zählte , die	Wahl	['Brücke', 'Mauer', 'Straßen', 'die', 'in']
57	D	Vorgestern reparierte der Politiker , der die Hände schüttelte , die	Wahl	['Straßen', 'Brücke', 'Straße', 'alte', 'Mauer']
58	A	Sofort schoss der Reporter , der die Kamera bediente , das	Foto	['Foto', 'Bild', 'Video', 'erste', 'Team']
58	B	Sofort schoss der Reporter , der die Zigarette rauchte , das	Foto	['Foto', 'Bild', 'erste', 'Feuer', 'Video']
58	C	Sofort verspeiste der Reporter , der die Kamera bediente , das	Foto	['Fleisch', 'Bild', 'erste', '"', 'Frühstück']
58	D	Sofort verspeiste der Reporter , der die Zigarette rauchte , das	Foto	['Essen', 'erste', 'Frühstück', 'Bier', 'Fleisch']
59	A	Unverzüglich entwickelte der Informatiker , der den Computer hochfuhr , das	Programm	['System', 'Programm', 'erste', 'Problem', 'Konzept']
59	B	Unverzüglich entwickelte der Informatiker , der den Energydrink trank , das	Programm	['Konzept', 'Bewusstsein', 'Produkt', 'Wissen', 'Problem']
59	C	Unverzüglich trocknete der Informatiker , der den Computer hochfuhr , das	Programm	['Programm', 'System', 'Bild', 'Problem', 'Projekt']
59	D	Unverzüglich trocknete der Informatiker , der den Energydrink trank , das	Programm	['Gesicht', 'Produkt', 'Buch', 'erste', 'Bild']
60	A	Nun verkündete der Richter , nachdem er den Angeklagten hereingerufen hatte , das	Urteil	['Urteil', 'Gericht', 'er', 'Geständnis', 'Verfahren']
60	B	Nun verkündete der Richter , nachdem er die Lesebrille aufgesetzt hatte , das	Urteil	['Urteil', 'Gericht', 'Verfahren', 'Ergebnis', 'Opfer']
60	C	Nun roch der Richter , nachdem er den Angeklagten hereingerufen hatte , das	Urteil	['Gericht', 'Urteil', 'Blut', 'Gesicht', 'ganze']
60	D	Nun roch der Richter , nachdem er die Lesebrille aufgesetzt hatte , das	Urteil	['Gesicht', 'erste', 'Blut', 'ganze', 'Haar']
61	A	Zunächst bestand der Student , der das Studium wiederaufnahm , die	Prüfung	['Prüfung', 'Prüfungen', 'Aufnahme', 'Abschlus', 'Möglichkeit']
61	B	Zunächst bestand der Student , der das Jackett auszog , die	Prüfung	['Prüfung', 'Aufgabe', 'Möglichkeit', 'erste', 'aus']
61	C	Zunächst kochte der Student , der das Studium wiederaufnahm , die	Prüfung	['Zutaten', 'Suppe', 'Speisen', 'Rezepte', 'ersten']
61	D	Zunächst kochte der Student , der das Jackett auszog , die	Prüfung	['Suppe', 'Zutaten', '"', 'Speisen', 'ersten']
62	A	Konzentriert sang der Sänger , der das Mikrofon umklammerte , das	Lied	['Lied', 'Publikum', '"', 'Stück', 'von']
62	B	Konzentriert sang der Sänger , der den Hut abnahm , das	Lied	['Lied', '"', 'Stück', 'erste', ',']
62	C	Konzentriert reparierte der Sänger , der das Mikrofon umklammerte , das	Lied	['Publikum', 'Mikrofon', 'Lied', 'Schlagzeug', 'Klavier']
62	D	Konzentriert reparierte der Sänger , der den Hut abnahm , das	Lied	['Lied', 'Album', 'Mikrofon', 'Publikum', 'Auto']
63	A	Zufrieden mischte der Anstreicher , der die Pinsel gewaschen hatte , die	Farbe	['Farbe', 'Farben', 'Pinsel', 'mit', 'Blumen']
63	B	Zufrieden mischte der Anstreicher , der die Plätzchen gegessen hatte , die	Farbe	['Kinder', 'Gäste', 'mit', 'Kerzen', 'Pl']
63	C	Zufrieden würzte der Anstreicher , der die Pinsel gewaschen hatte , die	Farbe	['Farbe', 'Farben', 'Pinsel', 'mit', 'Wände']
63	D	Zufrieden würzte der Anstreicher , der die Plätzchen gegessen hatte , die	Farbe	['Pl', 'Gäste', 'Kinder', 'mit', 'Weihnachts']
64	A	Einen Moment lang gießt der Gärtner , der den Dünger verteilt hat , die	Blumen	['Erde', 'Blumen', 'Pflanzen', 'Blätter', 'Äpfel']
64	B	Einen Moment lang gießt der Gärtner , der die Opernmusik angestellt hat , die	Blumen	['Blumen', 'Luft', 'ganze', 'Blätter', 'Noten']
64	C	Einen Moment lang ölt der Gärtner , der den Dünger verteilt hat , die	Blumen	['Blätter', 'Erde', 'Pflanzen', 'Sonne', 'Pflanze']
64	D	Einen Moment lang ölt der Gärtner , der die Opernmusik angestellt hat , die	Blumen	['Musik', 'Luft', 'ganze', 'Stimmung', 'Hände']
65	A	Gestern Nachmittag stutzte der Landschaftsgärtner , der die Gartenschere geschärft hatte , die	Hecke	['Äste', 'Blätter', 'Blumen', 'Pflanzen', 'Bäume']
65	B	Gestern Nachmittag stutzte der Landschaftsgärtner , der die Sonnenbrille aufgesetzt hatte , die	Hecke	['Bäume', 'Blätter', 'Sonne', 'Blumen', 'Sonnen']
65	C	Gestern Nachmittag bezahlte der Landschaftsgärtner , der die Gartenschere geschärft hatte , die	Hecke	['Arbeiten', 'Arbeit', 'Garten', 'Kosten', 'ersten']
65	D	Gestern Nachmittag bezahlte der Landschaftsgärtner , der die Sonnenbrille aufgesetzt hatte , die	Hecke	['Sonnen', 'Sonne', 'Arbeit', 'ersten', 'Kosten']
66	A	Früher beackerte der Landwirt , der den Traktor gestartet hatte , das	Feld	['Feld', 'Gebiet', 'Gelände', 'Land', 'Acker']
66	B	Früher beackerte der Landwirt , der den Sonnenhut aufgesetzt hatte , das	Feld	['Feld', 'Land', 'Gebiet', 'Acker', 'Getreide']
66	C	Früher kopierte der Landwirt , der den Traktor gestartet hatte , das	Feld	['Bild', 'Gerät', 'Fahrzeug', 'Rad', 'Tier']
66	D	Früher kopierte der Landwirt , der den Sonnenhut aufgesetzt hatte , das	Feld	['Bild', 'Buch', 'Foto', 'Wort', 'Tier']
67	A	Heute Morgen flickte der Schuster , der den Absatz ausgetauscht hatte , den	Schuh	['Schuh', 'Sack', 'Wagen', 'Laden', 'Zaun']
67	B	Heute Morgen flickte der Schuster , der die Sirene gehört hatte , den	Schuh	['Schuh', 'Teppich', 'Weg', 'Zaun', 'Sack']
67	C	Heute Morgen schmeckte der Schuster , der den Absatz ausgetauscht hatte , den	Schuh	['Kunden', 'ganzen', 'die', 'ich', 'Gästen']
67	D	Heute Morgen schmeckte der Schuster , der die Sirene gehört hatte , den	Schuh	['ganzen', 'Kuchen', 'Gästen', 'ersten', 'Geschmack']
68	A	Schnell packte der Urlauber , als er die Klamotten gefaltet hatte , den	Koffer	['Rucksack', 'Koffer', 'Sack', 'Schlüssel', 'Stoff']
68	B	Schnell packte der Urlauber , als er die Gardinen aufgezogen hatte , den	Koffer	['Koffer', 'Rucksack', 'Schlüssel', 'Wagen', 'Müll']
68	C	Schnell rührte der Urlauber , als er die Klamotten gefaltet hatte , den	Koffer	['Rucksack', 'Finger', 'Beutel', 'Laden', 'Boden']
68	D	Schnell rührte der Urlauber , als er die Gardinen aufgezogen hatte , den	Koffer	['Teppich', 'Tisch', 'Boden', 'ersten', 'Staub']
69	A	Sorgfältig erntet der Imker , nachdem er den Bienenstock erreicht hat , den	Honig	['Honig', 'Saft', 'ersten', 'Bienen', 'Pollen']
69	B	Sorgfältig erntet der Imker , nachdem er das Haus verlassen hat , den	Honig	['Honig', 'Saft', 'ersten', 'Samen', 'Bienen']

to be continued on the next page

Table A.3.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
69	C	Sorgfältig näht der Imker , nachdem er den Bienenstock erreicht hat , den	Honig	['Honig', 'Bienen', 'Bien', 'Stoff', 'Pollen']
69	D	Sorgfältig näht der Imker , nachdem er das Haus verlassen hat , den	Honig	['Honig', 'Bienen', 'Stoff', 'Baum', 'Faden']
70	A	Vorsichtig lädt der Jäger , der die Munition verschossen hat , das	Gewehr	['Gewehr', 'Tier', 'Geschoss', 'Feuer', 'Wild']
70	B	Vorsichtig lädt der Jäger , der den Campingplatz verlassen hat , das	Gewehr	['Tier', 'Auto', 'Feuer', 'Gewehr', 'Wild']
70	C	Vorsichtig schneidet der Jäger , der die Munition verschossen hat , das	Gewehr	['Fleisch', 'Holz', 'Gewehr', 'Wild', 'Tier']
70	D	Vorsichtig schneidet der Jäger , der den Campingplatz verlassen hat , das	Gewehr	['Gras', 'Holz', 'Fleisch', 'Wild', 'Tier']
71	A	Vorsichtig mischte der Maurer , der die Backsteine aufeinanderlegte , den	Mörtel	['Stein', 'Mör', 'Boden', 'Sand', 'Ziegel']
71	B	Vorsichtig mischte der Maurer , der die Nachbarn belauschte , den	Mörtel	['Boden', 'ganzen', 'Staub', 'Rasen', 'Dreck']
71	C	Vorsichtig schrieb der Maurer , der die Backsteine aufeinanderlegte , den	Mörtel	['Stein', 'Bau', 'Arbeitern', 'Rest', 'Namen']
71	D	Vorsichtig schrieb der Maurer , der die Nachbarn belauschte , den	Mörtel	['Brief', 'Namen', 'Kindern', 'beiden', 'Nachbarn']
72	A	Vorgestern mahlte der Müller , der die Mühle betrieb , das	Mehl	['Korn', 'Getreide', 'Mehl', 'Brot', 'Wasser']
72	B	Vorgestern mahlte der Müller , der das Lied pfiff , das	Mehl	['Lied', 'Brot', 'war', 'erste', 'Gedicht']
72	C	Vorgestern verzierte der Müller , der die Mühle betrieb , das	Mehl	['Haus', 'Gebäude', 'Mühl', 'Tor', 'Mühlen']
72	D	Vorgestern verzierte der Müller , der das Lied pfiff , das	Mehl	['Haus', 'Lied', 'Bild', 'Denkmal', 'Dorf']
73	A	Vorsichtig melkte der Viehzüchter , nachdem er den Stall ausgemistet hatte , die	Kuh	['Pferde', 'Tiere', 'Kühe', 'Schafe', 'Kuh']
73	B	Vorsichtig melkte der Viehzüchter , nachdem er das Gemüse sortiert hatte , die	Kuh	['Tiere', 'Schweine', 'Schafe', 'Kühe', 'Pflanzen']
73	C	Vorsichtig kleidete der Viehzüchter , nachdem er den Stall ausgemistet hatte , die	Kuh	['Tiere', 'Pferde', 'Schafe', 'Kuh', 'Kühe']
73	D	Vorsichtig kleidete der Viehzüchter , nachdem er das Gemüse sortiert hatte , die	Kuh	['Tiere', 'Schafe', 'Schweine', 'Kuh', 'Kühe']
74	A	Zuerst bezog der Hotelangestellte , nachdem er den Bezug entfaltet hatte , das	Bett	['Zimmer', 'Hotel', 'Geld', 'Frühstück', 'Wort']
74	B	Zuerst bezog der Hotelangestellte , nachdem er den Hamburger verspeist hatte , das	Bett	['Essen', 'Geld', 'Abendessen', 'Fleisch', 'Frühstück']
74	C	Zuerst fälschte der Hotelangestellte , nachdem er den Bezug entfaltet hatte , das	Bett	['Zimmer', 'Bild', 'Geld', 'Hotel', 'Fenster']
74	D	Zuerst fälschte der Hotelangestellte , nachdem er den Hamburger verspeist hatte , das	Bett	['Geld', 'Ticket', 'Bild', 'Hotel', 'Foto']
75	A	Augenblicklich erreicht der Sprinter , der das Rennen gewinnt , das	Ziel	['Ziel', 'Podium', 'Finale', 'Feld', 'Maximum']
75	B	Augenblicklich erreicht der Sprinter , der den Wind genießt , das	Ziel	['Ziel', 'Meer', 'Tempo', 'Maximum', 'Feld']
75	C	Augenblicklich reinigt der Sprinter , der das Rennen gewinnt , das	Ziel	['Auto', 'Cockpit', 'Gesicht', 'Fahrzeug', 'gesamte']
75	D	Augenblicklich reinigt der Sprinter , der den Wind genießt , das	Ziel	['Auto', 'Cockpit', 'Dach', 'Fahrzeug', 'Getriebe']
76	A	Direkt surft der Surfer , der die Badehose angezogen hat , die	Welle	['sich', 'Wellen', 'ganze', 'auf', 'er']
76	B	Direkt surft der Surfer , der den Tagesausflug unternommen hat , die	Welle	['Wellen', 'ganze', 'ersten', 'Sonne', 'Insel']
76	C	Direkt reserviert der Surfer , der die Badehose angezogen hat , die	Welle	['Wellen', 'in', 'sich', 'Dusche', 'Sonnen']
76	D	Direkt reserviert der Surfer , der den Tagesausflug unternommen hat , die	Welle	['Möglichkeit', 'Wellen', 'Insel', 'Zeit', 'Gelegenheit']
77	A	Gut gelaunt kontrollierte der Schaffner , der das Zugabteil betreten hatte , die	Fahrkarten	['Reisenden', 'Fahrgäste', 'Passagiere', 'beiden', 'Türen']
77	B	Gut gelaunt kontrollierte der Schaffner , der die Mittagspause beendet hatte , die	Fahrkarten	['Fahrgäste', 'Passagiere', 'beiden', 'Züge', 'Reisenden']
77	C	Gut gelaunt zitierte der Schaffner , der das Zugabteil betreten hatte , die	Fahrkarten	['Worte', 'Frau', 'Fahrgäste', 'ersten', 'beiden']
77	D	Gut gelaunt zitierte der Schaffner , der die Mittagspause beendet hatte , die	Fahrkarten	['Gäste', 'Worte', 'Frau', 'Fahrgäste', 'beiden']
78	A	Am Mittag verlegte der Elektriker , nachdem er den Strom abgeschaltet hatte , die	Leitung	['Stromversorgung', 'Strom', 'Leitung', 'Kabel', 'Arbeiten']
78	B	Am Mittag verlegte der Elektriker , nachdem er den Anwohner informiert hatte , die	Leitung	['Stromversorgung', 'Strom', 'Leitung', 'Arbeiten', 'Kabel']
78	C	Am Mittag verfasste der Elektriker , nachdem er den Strom abgeschaltet hatte , die	Leitung	['erste', 'Nachricht', 'ersten', 'letzte', 'Strom']
78	D	Am Mittag verfasste der Elektriker , nachdem er den Anwohner informiert hatte , die	Leitung	['erste', 'ersten', 'folgende', 'Nachricht', 'letzte']
79	A	Langsam hisst der Matrose , nachdem er die Taue gekappt hat , die	Segel	['Segel', 'Flagge', 'Anker', 'Fahne', 'Leinen']
79	B	Langsam hisst der Matrose , nachdem er die Handschuhe angezogen hat , die	Segel	['Segel', 'Fahne', 'Flagge', 'er', 'Hand']
79	C	Langsam verzehrt der Matrose , nachdem er die Taue gekappt hat , die	Segel	['Tau', 'Segel', 'Fische', 'er', 'sich']
79	D	Langsam verzehrt der Matrose , nachdem er die Handschuhe angezogen hat , die	Segel	['er', 'Fische', 'ihn', 'sich', 'ihm']
80	A	Direkt entspiegelt der Optiker , nachdem er den Sehtest durchgeführt hat , die	Gläser	['Augen', 'Ergebnisse', 'Bilder', 'Farben', 'Wirkung']
80	B	Direkt entspiegelt der Optiker , nachdem er die Jahresabrechnung aufgestellt hat , die	Gläser	['die', 'er', 'Ergebnisse', 'Jahres', 'Zahlen']
80	C	Direkt verschließt der Optiker , nachdem er den Sehtest durchgeführt hat , die	Gläser	['Augen', 'Brillen', 'Brille', 'Tür', 'Glas']
80	D	Direkt verschließt der Optiker , nachdem er die Jahresabrechnung aufgestellt hat , die	Gläser	['Tür', 'Augen', 'Brillen', 'Türen', 'Glas']
81	A	Auf der Stelle wirft der Basketballer , der den Ball gedribbelt hat , den	Korb	['Ball', 'Korb', 'P', 'ersten', 'nächsten']
81	B	Auf der Stelle wirft der Basketballer , der die Sitzbank verlassen hat , den	Korb	['Ball', 'Kopf', 'Blick', 'Korb', 'Basketball']
81	C	Auf der Stelle bucht der Basketballer , der den Ball gedribbelt hat , den	Korb	['Ball', 'nächsten', 'ersten', 'Weg', 'Pass']
81	D	Auf der Stelle bucht der Basketballer , der die Sitzbank verlassen hat , den	Korb	['nächsten', 'Flug', 'Platz', 'Weg', 'Bus']
82	A	Ohne zu zögern knackt der Perlentaucher , der den Meeresgrund erreicht hat , die	Muschel	['Tiefen', 'Wasser', 'Eis', 'Wellen', 'Schwelle']
82	B	Ohne zu zögern knackt der Perlentaucher , der die Taschenlampe angemacht hat , die	Muschel	['Tür', 'Taschen', 'Wand', 'Wasser', 'Taucher']
82	C	Ohne zu zögern verknotet der Perlentaucher , die den Meeresgrund erreicht hat , die	Muschel	['Fische', 'Felsen', 'Wasser', 'Insel', 'Eier']
82	D	Ohne zu zögern verknotet der Perlentaucher , der die Taschenlampe angemacht hat , die	Muschel	['Taschen', 'Fische', 'Taucher', 'Tasche', 'Augen']
83	A	Umsichtig spannt der Schütze , der den Pfeil gespitzt hat , den	Bogen	['Bogen', 'Pfeil', 'Abzug', 'Ball', 'Gegenstand']
83	B	Umsichtig spannt der Schütze , der das Laub zusammengekehrt hat , den	Bogen	['Bogen', 'Abzug', 'Arm', 'Haken', 'Gegenstand']
83	C	Umsichtig tippt der Schütze , der den Pfeil gespitzt hat , den	Bogen	['Pfeil', 'Bogen', 'Punkt', 'Ziel', 'Winkel']
83	D	Umsichtig tippt der Schütze , der das Laub zusammengekehrt hat , den	Bogen	['Pfeil', 'Winkel', 'Bogen', 'Abstand', 'Punkt']
84	A	Schnell verlegte der Dachdecker , der das Dach abdeckte , die	Ziegel	['Arbeiten', 'Dach', 'Leiter', 'Rohre', 'Last']
84	B	Schnell verlegte der Dachdecker , der den Ehering auszog , die	Ziegel	['Arbeit', 'Leiter', 'Frau', 'Arbeiten', 'E']
84	C	Schnell naschte der Dachdecker , der das Dach abdeckte , die	Ziegel	['Dach', 'Holz', 'Äste', 'ersten', 'Reste']
84	D	Schnell naschte der Dachdecker , der den Ehering auszog , die	Ziegel	['Hand', 'Hände', 'Geschenke', 'Kinder', 'Blumen']
85	A	Zuerst stapelt der Umzugshelfer , der den Umzugswagen gefahren hat , die	Kartons	['Äpfel', 'Kisten', 'Ja', 'leeren', 'einzelnen']
85	B	Zuerst stapelt der Umzugshelfer , der die Firma gegründet hat , die	Kartons	['Kisten', 'Äpfel', 'alten', 'Holz', 'Kiste']
85	C	Zuerst serviert der Umzugshelfer , der den Umzugswagen gefahren hat , die	Kartons	['Speisen', 'Gäste', 'Getränke', 'Geschenke', '""']
85	D	Zuerst serviert der Umzugshelfer , der die Firma gegründet hat , die	Kartons	['Speisen', 'Gäste', '""', 'Süßigkeiten', 'leck']
86	A	Heute früh schwänzte der Schüler , der die Klasse wiederholt hatte , den	Unterricht	['Unterricht', 'Klassen', 'Schul', 'obligatorischen', 'Schult']
86	B	Heute früh schwänzte der Schüler , der das Handy verloren hatte , den	Unterricht	['Unterricht', 'Schul', 'Bus', 'Weg', 'Klassen']

to be continued on the next page

Table A.3.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
86	C	Heute früh dekorierte der Schüler , der die Klasse wiederholt hatte , den	Unterricht	['Raum', 'Eingang', 'Tisch', 'Boden', 'Schul']
86	D	Heute früh dekorierte der Schüler , der das Handy verloren hatte , den	Unterricht	['Raum', 'Weg', 'Eingang', 'Parkplatz', 'Platz']
87	A	Am Abend kontrollierte der Türsteher , der den Geldbeutel durchsuchte , den	Ausweis	['Mann', 'Geld', 'Fahrer', 'Täter', 'jungen']
87	B	Am Abend kontrollierte der Türsteher , der die Nachtluft genoss , den	Ausweis	['Mann', 'Laden', 'Platz', 'jungen', 'Eingang']
87	C	Am Abend probierte der Türsteher , der den Geldbeutel durchsuchte , den	Ausweis	['Geld', 'Mann', 'anderen', 'Tresor', 'Beutel']
87	D	Am Abend probierte der Türsteher , der die Nachtluft genoss , den	Ausweis	['Schlüssel', 'anderen', 'ganzen', 'Gang', 'Rest']
88	A	Umsichtig wechselte der Babysitter , der das Baby geweckt hatte , die	Windel	['Worte', 'Seiten', 'Nummer', 'Straßenseite', 'Situation']
88	B	Umsichtig wechselte der Babysitter , der den Obstsalat zubereitet hatte , die	Windel	['Zutaten', 'Worte', 'Rollen', 'Hand', 'Ware']
88	C	Umsichtig bastelte der Babysitter , der das Baby geweckt hatte , die	Windel	['kleinen', 'kleine', 'Puppe', 'ersten', 'Bilder']
88	D	Umsichtig bastelte der Babysitter , der den Obstsalat zubereitet hatte , die	Windel	['Äpfel', 'kleinen', 'Zutaten', 'Eier', 'Figuren']
89	A	Langsam stemmte der Bodybuilder , der das Fitnessstudio betreten hatte , die	Gewichte	['Arme', 'Tür', 'Beine', 'Schulter', 'Schultern']
89	B	Langsam stemmte der Bodybuilder , der das Licht angemacht hatte , die	Gewichte	['Arme', 'Beine', 'Schultern', 'Schulter', 'Box']
89	C	Langsam verdaute der Bodybuilder , der das Fitnessstudio betreten hatte , die	Gewichte	['Aufregung', 'Welt', 'Stimmung', 'Situation', 'Aufmerksamkeit']
89	D	Langsam verdaute der Bodybuilder , der das Licht angemacht hatte , die	Gewichte	['Aufregung', 'Welt', 'Stimmung', 'Situation', 'Schmerzen']
90	A	Umsichtig repariert der Klempner , der das Wasser ausgestellt hat , das	Rohr	['Wasser', 'Rohr', 'Haus', 'Dach', 'in']
90	B	Umsichtig repariert der Klempner , der den Kirchenchor geleitet hat , das	Rohr	['Dach', 'Instrument', 'alte', 'Kirchen', 'Gel']
90	C	Umsichtig nascht der Klempner , der das Wasser ausgestellt hat , das	Rohr	['Wasser', 'in', 'Glas', 'mit', 'aus']
90	D	Umsichtig nascht der Klempner , der den Kirchenchor geleitet hat , das	Rohr	['Lied', 'Blut', 'ganze', 'Essen', 'Brot']
91	A	Schnell schreibt der Journalist , der die Recherche durchgeführt hat , den	Artikel	['Artikel', 'Bericht', 'Text', 'Leser', 'Namen']
91	B	Schnell schreibt der Journalist , der das Portemonnaie gefunden hat , den	Artikel	['Brief', 'Artikel', 'Namen', 'Text', 'Bericht']
91	C	Schnell lenkt der Journalist , der die Recherche durchgeführt hat , den	Artikel	['Blick', 'Fokus', 'Bogen', 'Verdacht', 'Kopf']
91	D	Schnell lenkt der Journalist , der das Portemonnaie gefunden hat , den	Artikel	['Blick', 'Fokus', 'Kopf', 'Finger', 'Wagen']
92	A	Heute Morgen flickte der Radler , der die Luftpumpe benutzt hatte , den	Reifen	['Arm', 'Helm', 'Weg', 'Sattel', 'Boden']
92	B	Heute Morgen flickte der Radler , der den Schrank abgeschlossen hatte , den	Reifen	['Schlüssel', 'Boden', 'Rucksack', 'Schrank', 'Deckel']
92	C	Heute Morgen trank der Radler , der die Luftpumpe benutzt hatte , den	Reifen	['Kaffee', 'Wein', 'ganzen', 'Champagner', 'Alkohol']
92	D	Heute Morgen trank der Radler , der den Schrank abgeschlossen hatte , den	Reifen	['Kaffee', 'Champagner', 'Wein', 'Sekt', 'ganzen']
93	A	Umgehend entleert der Postbote , der das Postauto geparkt hat , den	Briefkasten	['Brief', 'Post', 'Wagen', 'Container', 'Zug']
93	B	Umgehend entleert der Postbote , der den Kugelschreiber verloren hat , den	Briefkasten	['Brief', 'Umschlag', 'Karton', 'Koffer', 'Post']
93	C	Umgehend verdrückt der Postbote , der das Postauto geparkt hat , den	Briefkasten	['Brief', 'Zettel', 'Post', 'Inhalt', 'Schlüssel']
93	D	Umgehend verdrückt der Postbote , der den Kugelschreiber verloren hat , den	Briefkasten	['Brief', 'Zettel', 'Umschlag', 'Kugel', 'Satz']
94	A	Hektisch kauft der Junkie , der den Dealer angerufen hat , die	Drogen	['Ware', 'Drogen', 'Zigaretten', 'Wohnung', 'Waren']
94	B	Hektisch kauft der Junkie , der die Kapuze aufgezogen hat , die	Drogen	['Ware', 'Zigaretten', 'Taschen', 'Kap', 'Pistole']
94	C	Hektisch näht der Junkie , der den Dealer angerufen hat , die	Drogen	['er', 'Tür', 'Ware', 'Schuhe', 'sich']
94	D	Hektisch näht der Junkie , der die Kapuze aufgezogen hat , die	Drogen	['Schuhe', 'Jacke', 'Haare', 'Beine', 'Hosen']
95	A	Zügig leerte der Müllmann , der den Müllwagen geparkt hatte , die	Tonne	['Mü', 'leeren', 'Müll', 'Container', 'Taschen']
95	B	Zügig leerte der Müllmann , der den Hausbesitzer gegrüßt hatte , die	Tonne	['Mü', 'leeren', 'Müll', 'Flaschen', 'Taschen']
95	C	Zügig bastelte der Müllmann , der den Müllwagen geparkt hatte , die	Tonne	['Figuren', 'ersten', 'Holz', 'Idee', 'neuen']
95	D	Zügig bastelte der Müllmann , der den Hausbesitzer gegrüßt hatte , die	Tonne	['ersten', 'Figuren', 'Holz', 'kleinen', 'neuen']
96	A	Hektisch knallte der Reiter , der das Pferd ritt , die	Peitsche	['Pferde', 'Bahn', 'Hindernisse', 'Beine', 'Tür']
96	B	Hektisch knallte der Reiter , der die Landschaft durchquerte , die	Peitsche	['Pferde', 'Beine', 'Ab', 'Reiter', 'Kurven']
96	C	Hektisch verfasste der Reiter , der das Pferd ritt , die	Peitsche	['Geschichte', 'erste', 'Befehle', 'Nachricht', 'ersten']
96	D	Hektisch verfasste der Reiter , der die Landschaft durchquerte , die	Peitsche	['ersten', 'erste', 'Geschichte', '"', 'Schilder']
97	A	Nach kurzem Überlegen zückte der Gangster , der die Bank überfiel , die	Waffe	['Geld', 'Waffe', 'Pistole', 'Hand', 'Karte']
97	B	Nach kurzem Überlegen zückte der Gangster , der den Bus betrat , die	Waffe	['Waffe', 'Geld', 'Pistole', 'Hand', 'Taschen']
97	C	Nach kurzem Überlegen gründete der Gangster , der die Bank überfiel , die	Waffe	['Bank', '"', 'Firma', 'Bande', 'Band']
97	D	Nach kurzem Überlegen gründete der Gangster , der den Bus betrat , die	Waffe	['Bande', '"', 'Gang', 'Firma', 'Gruppe']
98	A	Zufrieden kaperte der Pirat , der die Flagge hisste , das	Schiff	['Schiff', 'Boot', 'Feuer', 'Segel', 'Meer']
98	B	Zufrieden kaperte der Pirat , der die Fehde austrug , das	Schiff	['Schiff', 'Boot', 'Dorf', 'Land', 'Haus']
98	C	Zufrieden verrührte der Pirat , der die Flagge hisste , das	Schiff	['Schiff', 'Boot', 'Banner', 'Feuer', 'Meer']
98	D	Zufrieden verrührte der Pirat , der die Fehde austrug , das	Schiff	['Schiff', 'Land', 'Boot', 'Dorf', 'Ziel']
99	A	Sofort kauft der Börsenmakler , der die Börse betreten hat , die	Aktie	['Aktien', 'Papiere', 'Ware', 'Wertpapiere', 'Börse']
99	B	Sofort kauft der Börsenmakler , der den Whiskey eingeschenkt hat , die	Aktie	['Ware', 'Flasche', 'Aktien', 'ersten', 'Flaschen']
99	C	Sofort knabbert der Börsenmakler , der die Börse betreten hat , die	Aktie	['Börse', 'Aktien', 'Augen', 'Hände', 'Ware']
99	D	Sofort knabbert der Börsenmakler , der den Whiskey eingeschenkt hat , die	Aktie	['Augen', 'Ware', 'Flaschen', 'Zigaretten', 'Flasche']
100	A	Gestern bohrte der Handwerker , der die Bohrmaschine hielt , das	Loch	['Loch', 'Bohr', 'erste', 'Holz', 'Haus']
100	B	Gestern bohrte der Handwerker , der das Kaugummi kaute , das	Loch	['Loch', 'Haus', 'in', 'erste', 'Kau']
100	C	Gestern warf der Handwerker , der die Bohrmaschine hielt , das	Loch	['Werkzeug', 'Holz', 'Haus', 'Bohr', 'Messer']
100	D	Gestern warf der Handwerker , der das Kaugummi kaute , das	Loch	['Auto', 'Kau', 'Haus', 'Fenster', 'Geschirr']
101	A	Am Morgen misst der Arzthelfer , der das Messgerät hält , den	Blutdruck	['Blutdruck', 'Puls', 'Blut', 'Körper', 'Umfang']
101	B	Am Morgen misst der Arzthelfer , der den Bildschirm anstellt , den	Blutdruck	['Blutdruck', 'Bildschirm', 'Umfang', 'Blut', 'Puls']
101	C	Am Morgen kauft der Arzthelfer , der das Messgerät hält , den	Blutdruck	['Patienten', 'Kaffee', 'Koffer', 'Karton', 'Schlüssel']
101	D	Am Morgen kauft der Arzthelfer , der den Bildschirm anstellt , den	Blutdruck	['Bildschirm', 'Computer', 'Fernseher', 'Laptop', 'Patienten']
102	A	Konzentriert schwang der Torero , der den Stier verwundet hatte , das	Tuch	['Schwert', 'Pferd', 'Tier', 'Gewicht', 'Gesicht']
102	B	Konzentriert schwang der Torero , der das Gebet aufgesagt hatte , das	Tuch	['Gesicht', 'ganze', 'Kreuz', 'Schwert', 'Lied']
102	C	Konzentriert befüllte der Torero , der den Stier verwundet hatte , das	Tuch	['Tor', 'Pferd', 'Tier', 'Haus', 'Fleisch']
102	D	Konzentriert befüllte der Torero , der das Gebet aufgesagt hatte , das	Tuch	['Kreuz', 'Feuer', 'Herz', 'Gebet', 'Fleisch']
103	A	Angespannt zückte der Ritter , nachdem er den Kampf begonnen hatte , das	Schwert	['Schwert', 'Messer', 'Gewehr', 'Schild', 'Kreuz']

to be continued on the next page

Table A.3.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
103	B	Angespannt zückte der Ritter , nachdem er die Brücke überquert hatte , das	Schwert	['Schwert', 'Messer', 'Schild', 'Gewehr', 'Schloss']
103	C	Angespannt kaute der Ritter , nachdem er den Kampf begonnen hatte , das	Schwert	['Schwert', 'Pferd', 'Blut', 'erste', 'Fleisch']
103	D	Angespannt kaute der Ritter , nachdem er die Brücke überquert hatte , das	Schwert	['Schwert', 'Schiff', 'erste', 'Pferd', 'Fleisch']
104	A	Wachsam steuert der Sanitäter , der das Blaulicht eingeschaltet hat , den	Rettungswagen	['Rettungswagen', 'Krankenwagen', 'Einsatz', 'Wagen', 'Patienten']
104	B	Wachsam steuert der Sanitäter , der den Notizblock weggelegt hat , den	Rettungswagen	['Patienten', 'Krankenwagen', 'Rettungswagen', 'Wagen', 'Einsatz']
104	C	Wachsam kopiert der Sanitäter , der das Blaulicht eingeschaltet hat , den	Rettungswagen	['Krankenwagen', 'Wagen', 'Mann', 'Patienten', 'Rettungswagen']
104	D	Wachsam kopiert der Sanitäter , der den Notizblock weggelegt hat , den	Rettungswagen	['Zettel', 'Ordner', 'Inhalt', 'Notiz', 'Brief']
105	A	Ohne zu zögern verweigerte der Zeuge , der den Angeklagten wiedererkannte , die	Aussage	['Aussage', 'Herausgabe', 'Aussagen', 'Tat', 'weitere']
105	B	Ohne zu zögern verweigerte der Zeuge , der den Gehstock umklammerte , die	Aussage	['Aussage', 'Herausgabe', 'Hand', 'Tat', 'Festnahme']
105	C	Ohne zu zögern bestellte der Zeuge , der den Angeklagten wiedererkannte , die	Aussage	['Zeugen', 'Aussage', 'Waffe', 'Anklage', 'Polizei']
105	D	Ohne zu zögern bestellte der Zeuge , der den Gehstock umklammerte , die	Aussage	['Waffe', 'Polizei', 'Pistole', 'Leiche', 'Zeugen']
106	A	Sogleich schnürt der Eiskunstläufer , der die Eishalle erreicht hat , die	Schlittschuhe	['Schl', 'Arme', 'Schuhe', 'Luft', 'Kehle']
106	B	Sogleich schnürt der Eiskunstläufer , der das Bonbon gelutscht hat , die	Schlittschuhe	['Kehle', 'Luft', 'Arme', 'Schuhe', 'Schl']
106	C	Sogleich zerbricht der Eiskunstläufer , der die Eishalle erreicht hat , die	Schlittschuhe	['Eis', 'Tür', 'Arme', 'Augen', 'Hand']
106	D	Sogleich zerbricht der Eiskunstläufer , der das Bonbon gelutscht hat , die	Schlittschuhe	['Augen', 'Arme', 'ganze', 'Eis', 'Welt']
107	A	Gut gelaunt schwingt der Cowboy , nachdem er das Pferd gesattelt hatte , das	Lasso	['sich', 'er', 'auf', 'in', 'mit']
107	B	Gut gelaunt schwingt der Cowboy , nachdem er den Kameraden gerufen hat , das	Lasso	['Tanz', 'Kommando', 'Wort', 'Band', 'Schlacht']
107	C	Gut gelaunt zerhackt der Cowboy , nachdem er das Pferd gesattelt hatte , das	Lasso	['Pferd', 'er', 'sich', 'auf', 'mit']
107	D	Gut gelaunt zerhackt der Cowboy , nachdem er den Kameraden gerufen hat , das	Lasso	['Pferd', 'Schloss', 'Schlacht', 'Haus', 'Schwein']
108	A	Entspannt schrieb der Regisseur , der den Plot konstruiert hatte , das	Drehbuch	['Drehbuch', 'Buch', 'Skript', 'Stück', 'Ende']
108	B	Entspannt schrieb der Regisseur , der den Kamin angemacht hatte , das	Drehbuch	['Drehbuch', 'Stück', 'Buch', 'Bild', 'Video']
108	C	Entspannt verzehrte der Regisseur , der den Plot konstruiert hatte , das	Drehbuch	['Drehbuch', 'Publikum', 'Buch', 'Thema', 'Stück']
108	D	Entspannt verzehrte der Regisseur , der den Kamin angemacht hatte , das	Drehbuch	['Essen', 'Publikum', 'Stück', 'Beste', 'Feuer']
109	A	Sofort entdeckt der Astronom , der das Teleskop aufgestellt hat , den	Stern	['Stern', 'Planeten', 'Ursprung', 'Himmel', 'Fehler']
109	B	Sofort entdeckt der Astronom , der die Uhrzeit notiert hat , den	Stern	['Stern', 'Tag', 'Himmel', 'Mond', 'Fehler']
109	C	Sofort serviert der Astronom , der das Teleskop aufgestellt hat , den	Stern	['Gästen', 'Sternen', 'ersten', 'Planeten', 'Besuchern']
109	D	Sofort serviert der Astronom , der die Uhrzeit notiert hat , den	Stern	['Sonnenaufgang', 'Tag', 'Stern', 'Planeten', 'Mond']
110	A	Gestern lichtete der Seemann , der das Boot kommandierte , den	Anker	['Anker', 'Hafen', 'Kanal', 'Strand', 'Rumpf']
110	B	Gestern lichtete der Seemann , der den Atem anhielt , den	Anker	['Anker', 'Hafen', 'Strand', 'Himmel', 'Wind']
110	C	Gestern kochte der Seemann , der das Boot kommandierte , den	Anker	['Fisch', 'ganzen', 'Brei', '""', 'ersten']
110	D	Gestern kochte der Seemann , der den Atem anhielt , den	Anker	['ganzen', 'Fisch', 'Tee', 'Kaffee', 'letzten']
111	A	Am Vormittag knotete der Bergsteiger , nachdem er den Haken festgebohrt hatte , das	Seil	['Seil', 'erste', 'Haken', 'Loch', 'Kabel']
111	B	Am Vormittag knotete der Bergsteiger , nachdem er die Krähe verscheucht hatte , das	Seil	['Seil', 'Dach', 'erste', 'Tier', 'Kreuz']
111	C	Am Vormittag roch der Bergsteiger , nachdem er den Haken festgebohrt hatte , das	Seil	['erste', 'Licht', 'Seil', 'Wasser', 'Feuer']
111	D	Am Vormittag roch der Bergsteiger , nachdem er die Krähe verscheucht hatte , das	Seil	['erste', 'Licht', 'Feuer', 'Wasser', 'Haar']
112	A	Entspannt dreht der Croupier , der den Spieltisch vorbereitet hat , das	Rouletterad	['Geld', 'Spiel', 'Cro', 'Casino', 'Rad']
112	B	Entspannt dreht der Croupier , der die Frau beeindruckt hat , das	Rouletterad	['Geld', 'Spiel', 'Becken', 'Rad', 'Glas']
112	C	Entspannt schlürft der Croupier , der den Spieltisch vorbereitet hat , das	Rouletterad	['Bier', 'Glas', 'erste', 'Wasser', 'Spiel']
112	D	Entspannt schlürft der Croupier , der die Frau beeindruckt hat , das	Rouletterad	['Bier', 'Glas', 'erste', 'Wasser', 'Getränk']
113	A	Neulich feilte die Kosmetikerin , während sie die Maniküre durchführte , die	Nägel	['Haare', 'sie', 'Haut', 'sich', 'Hände']
113	B	Neulich feilte die Kosmetikerin , während sie das Gespräch belauschte , die	Nägel	['Augen', 'Haare', 'Hände', 'Lippen', 'Hand']
113	C	Neulich baute die Kosmetikerin , während sie die Maniküre durchführte , die	Nägel	['Haare', 'Man', 'Haut', 'T', 'Wände']
113	D	Neulich baute die Kosmetikerin , während sie das Gespräch belauschte , die	Nägel	['Augen', 'Haare', 'Kosmetik', 'Wohnung', 'Gläser']
114	A	Sogleich prophezeite der Wahrsager , der die Kristallkugel befragt hatte , die	Zukunft	['Ankunft', 'Geburt', 'Rückkehr', 'Zukunft', 'Welt']
114	B	Sogleich prophezeite der Wahrsager , der die Kerze angezündet hatte , die	Zukunft	['Ankunft', 'Rückkehr', 'Geburt', 'Zukunft', 'Zerstörung']
114	C	Sogleich würzte der Wahrsager , der die Kristallkugel befragt hatte , die	Zukunft	['Worte', 'Geschichte', 'Frage', 'Speisen', 'Fragen']
114	D	Sogleich würzte der Wahrsager , der die Kerze angezündet hatte , die	Zukunft	['Worte', 'Ker', 'Geschichte', 'Augen', 'Kerzen']
115	A	Heute beginnt der Archäologe , der die Schaufel hervorgeholt hat , die	Ausgrabung	['die', 'in', 'den', 'von', 'das']
115	B	Heute beginnt der Archäologe , der das Taxi genommen hat , die	Ausgrabung	['Ausgrabungen', 'Ausgra', 'archäologische', 'Straße', 'archäologischen']
115	C	Heute raspelt der Archäologe , der die Schaufel hervorgeholt hat , die	Ausgrabung	['Erde', 'Steine', 'die', 'in', 'von']
115	D	Heute raspelt der Archäologe , der das Taxi genommen hat , die	Ausgrabung	['Gräber', 'Ruinen', 'Geschichte', 'Stadt', 'Überreste']
116	A	Zügig bindet der Florist , der die Rosen gezüchtet hat , den	Strauß	['Flor', 'Rosen', 'Baum', 'Boden', 'Blumen']
116	B	Zügig bindet der Florist , der die Fenster geschlossen hat , den	Strauß	['Rahmen', 'Teppich', 'Stoff', 'Boden', 'Flor']
116	C	Zügig ölt der Florist , der die Rosen gezüchtet hat , den	Strauß	['Flor', 'Rosen', 'Baum', 'Boden', 'Blumen']
116	D	Zügig ölt der Florist , der die Fenster geschlossen hat , den	Strauß	['Boden', 'Flor', 'Baum', 'Teppich', 'Rahmen']
117	A	Umgehend spielt der Schauspieler , der das Skript auswendig gelernt hat , die	Rolle	['Rolle', 'Hauptrolle', 'Geschichte', 'Rollen', 'Musik']
117	B	Umgehend spielt der Schauspieler , der die Torte gebacken hat , die	Rolle	['Tor', 'Geschichte', 'Rolle', 'Szene', 'Melodie']
117	C	Umgehend schnibbelt der Schauspieler , der das Skript auswendig gelernt hat , die	Rolle	['Dialoge', 'Worte', 'Herzen', 'Lippen', 'Sätze']
117	D	Umgehend schnibbelt der Schauspieler , der die Torte gebacken hat , die	Rolle	['sich', 'Tor', 'in', 'mit', 'die']
118	A	Am Abend moderiert der Moderator , der die Zuschauer unterhält , die	Sendung	['Sendung', 'Show', 'erste', '""', 'Musik']
118	B	Am Abend moderiert der Moderator , der die Weltmeere bereist , die	Sendung	['""', 'Sendung', 'Show', 'Live', 'Gäste']
118	C	Am Abend wiegt der Moderator , der die Zuschauer unterhält , die	Sendung	['Arme', 'Hände', 'Hand', 'Worte', 'Zuschauer']
118	D	Am Abend wiegt der Moderator , der die Weltmeere bereist , die	Sendung	['Worte', '""', 'Gedanken', 'Welt', 'Bilder']

to be continued on the next page

Table A.3.: (continued from previous page)

I	Con	Target Sentence	Target	Predictions
119	A	Ohne zu zögern komponierte der Musiker , der die Noten aufschrieb , das	Stück	['Lied', 'Stück', 'Werk', 'erste', '"']]
119	B	Ohne zu zögern komponierte der Musiker , der die Fragen beantwortete , das	Stück	['Lied', 'Stück', '"', 'Werk', 'erste']
119	C	Ohne zu zögern grillte der Musiker , der die Noten aufschrieb , das	Stück	['Essen', 'Lied', 'Publikum', 'Ergebnis', 'erste']
119	D	Ohne zu zögern grillte der Musiker , der die Fragen beantwortete , das	Stück	['Essen', 'Bier', 'Publikum', 'Fleisch', '"']]
120	A	Fröhlich malte der Künstler , der die Leinwand aufgespannt hatte , das	Bild	['Bild', 'Gemälde', 'Motiv', 'Gesicht', '"']]
120	B	Fröhlich malte der Künstler , der den Nachtisch zubereitet hatte , das	Bild	['Bild', 'Gemälde', '"', 'Motiv', 'Gedicht']
120	C	Fröhlich bewohnte der Künstler , der die Leinwand aufgespannt hatte , das	Bild	['Haus', 'Atelier', 'Gebäude', 'Zimmer', 'Wohnzimmer']
120	D	Fröhlich bewohnte der Künstler , der den Nachtisch zubereitet hatte , das	Bild	['Haus', 'Zimmer', 'Gebäude', 'Atelier', 'Anwesen']

Table A.4.: **ADBC23** GPT-2 top 5 predictions for the target word position.

I	Con	Target Sentence	Target	Distractor	Predictions
1	A	Dann verabschiedete die Dame den	Touristen	Koffer	['Koffer', 'Touristen', 'Reisenden', 'Passagier', 'kleinen']
1	B	Dann wog die Dame den	Touristen	Koffer	['Koffer', 'Rucksack', 'ganzen', 'gesamten', 'kleinen']
1	C	Dann unterschrieb die Dame den	Touristen	Koffer	['Koffer', 'Vertrag', 'Kauf', 'Reise', 'Flug']
2	A	Dann kaufte der Lehrer die	Weltkarte	Verkäuferin	['Karte', 'Welt', 'Kiste', 'Karten', 'Tafel']
2	B	Dann küsste der Lehrer die	Weltkarte	Verkäuferin	['Verkä', 'Tür', 'Lehrerin', 'Verkäufer', 'Kiste']
2	C	Dann füllte der Lehrer die	Weltkarte	Verkäuferin	['Karte', 'Kiste', 'Vit', 'Tafel', 'Flasche']
3	A	Dann stempelte der Busfahrer die	Streifenkarte	Redakteurin	['Streifen', 'Karte', 'Karten', 'Zeitung', 'Zeichnung']
3	B	Dann beschimpfte der Busfahrer die	Streifenkarte	Redakteurin	['Redakteur', 'Reporter', 'Redaktion', 'Frau', 'Chefredakteur']
3	C	Dann aß der Busfahrer die	Streifenkarte	Redakteurin	['Streifen', 'Karte', 'Zeitung', 'Karten', 'ganze']
4	A	Dann nahm der Lehrling den	Hammer	Schreiner	['Hammer', 'Stock', 'Nagel', 'anderen', 'ganzen']
4	B	Dann belächelte der Lehrling den	Hammer	Schreiner	['Hammer', 'Schreiner', 'Meister', 'Lehr', 'Gesellen']
4	C	Dann aß der Lehrling den	Hammer	Schreiner	['Hammer', 'ganzen', 'Stock', 'Stein', 'Nagel']
5	A	Daraufhin verpackte der Konditor den	Apfelkuchen	Opa	['Apfel', 'Kuchen', 'Obst', 'Nuss', 'A']
5	B	Daraufhin belächelte der Konditor den	Apfelkuchen	Opa	['Opa', 'Apfel', 'Kuchen', 'Obst', 'Großvater']
5	C	Daraufhin spülte der Konditor den	Apfelkuchen	Opa	['Apfel', 'Kuchen', 'Obst', 'Teig', 'A']
6	A	Nichtsahnend nahm der Kunde die	Frühlingsrolle	Lieferbotin	['Frühlings', 'Bestellung', 'Oster', 'neue', 'bestellte']
6	B	Nichtsahnend begrüßte der Kunde die	Frühlingsrolle	Lieferbotin	['Liefer', 'Bestellung', 'Lieferung', 'neue', 'Paket']
6	C	Nichtsahnend reparierte der Kunde die	Frühlingsrolle	Lieferbotin	['Frühlings', 'alte', 'Liefer', 'Oster', 'Kühl']
7	A	Dann durchschnitt der Metzger die	Fleischwurst	Vegetarierin	['Flei', 'Fleisch', 'Wurst', 'Kehle', 'Brü']
7	B	Dann belächelte der Metzger die	Fleischwurst	Vegetarierin	['Veget', 'Flei', 'Fleisch', 'Wurst', 'veget']
7	C	Dann mietete der Metzger die	Fleischwurst	Vegetarierin	['Küche', 'Gaststätte', 'Wurst', 'Flei', 'Fleisch']
8	A	Dann bedrohte der Tierliebhaber den	Kutscher	Gaul	['Gaul', 'Kut', 'Hund', 'Tier', 'Pferde']
8	B	Dann streichelte der Tierliebhaber den	Kutscher	Gaul	['Gaul', 'Hund', 'Hals', 'Kopf', 'anderen']
8	C	Dann füllte der Tierliebhaber den	Kutscher	Gaul	['Gaul', 'Kopf', 'Sattel', 'Kut', 'Krug']
9	A	Dann bestieg der Kapitän das	Segelboot	Pärchen	['Boot', 'Segel', 'Schiff', 'P', 'Ruder']
9	B	Dann rettete der Kapitän das	Segelboot	Pärchen	['P', 'Segel', 'Boot', 'Paar', 'kleine']
9	C	Dann verschloss der Kapitän das	Segelboot	Pärchen	['Segel', 'Boot', 'P', 'Ruder', 'Schiff']
10	A	Daraufhin lobte die Hausfrau den	Handwerker	Wasserhahn	['Handwerker', 'neuen', 'Haus', 'Wasser', 'Meister']
10	B	Daraufhin ersetzte die Hausfrau den	Handwerker	Wasserhahn	['Wasser', 'Hahn', 'alten', 'Handwerker', 'repar']
10	C	Daraufhin knickte die Hausfrau den	Handwerker	Wasserhahn	['Wasser', 'Hahn', 'Rohr', 'hahn', 'Abfluss']
11	A	Nach der Führung faltete der Urlauber den	Flyer	Guide	['Flyer', 'Zettel', 'Text', 'Vertrag', 'Brief']
11	B	Nach der Führung lobte der Urlauber den	Flyer	Guide	['Guide', 'Fremden', 'Führer', 'Reise', 'Reisenden']
11	C	Nach der Führung kochte der Urlauber den	Flyer	Guide	['Flyer', 'ganzen', 'Führer', 'Gästen', 'Rei']
12	A	Daraufhin bedrohte der Paparazzi die	Schauspielerin	Kamera	['Schauspielerin', 'Kamera', 'Polizei', 'Schauspieler', 'Reporter']
12	B	Daraufhin schulterte der Paparazzi die	Schauspielerin	Kamera	['Schauspielerin', 'Kamera', 'Frau', 'Schauspieler', 'Kamer']
12	C	Daraufhin schulterte der Paparazzi die	Schauspielerin	Kamera	['Schauspielerin', 'Kamera', 'Frau', 'Schauspieler', 'Kamer']
13	A	Daraufhin lobte der Schneider die	Assistentin	Schaufensterpuppe	['Schaufenster', 'Puppe', 'neue', 'junge', 'Assistentin']
13	B	Daraufhin bewunderte der Schneider die	Assistentin	Schaufensterpuppe	['Schaufenster', 'Puppe', 'neue', 'Idee', 'Statue']
13	C	Daraufhin schnitt der Schneider die	Assistentin	Schaufensterpuppe	['Schaufenster', 'Puppe', 'Tür', 'Fenster', 'Pu']
14	A	Danach musterte das Mädchen den	Schwimmer	Sprung	['Schwimmer', 'Jungen', 'Schwimm', 'Sprung', 'Mann']
14	B	Danach bewertete das Mädchen den	Schwimmer	Sprung	['Sprung', 'Schwimm', 'ersten', 'Schwimmer', 'Spr']
14	C	Danach salzte das Mädchen den	Schwimmer	Sprung	['Schwimmer', 'Sprung', 'Schwimm', 'Becken', 'Jungen']
15	A	Dann verabschiedete der Chefarzt die	Sekretärin	Diktiermaschine	['Patientin', 'Sekretärin', 'D', 'Ärztin', 'neue']
15	B	Dann enthüllte der Chefarzt die	Sekretärin	Diktiermaschine	['D', 'neue', 'Maschine', 'neuen', 'alte']
15	C	Dann leerte der Chefarzt die	Sekretärin	Diktiermaschine	['D', 'Maschine', 'Schreib', 'alte', 'Zeitung']
16	A	Daraufhin verabschiedete die Reporterin den	Mitarbeiter	Bauern	['Bauern', 'Bauer', 'Reporter', 'Farmer', 'Journalisten']
16	B	Daraufhin befragte die Reporterin den	Mitarbeiter	Bauern	['Bauern', 'Bauer', 'Farmer', 'Landwirt', 'Besitzer']
16	C	Daraufhin ordnete die Reporterin den	Mitarbeiter	Bauern	['Stall', 'Besuch', 'Bauern', 'Umzug', 'Bau']
17	A	Danach parkte die Tochter den	Rasenmäher	Gärtner	['Rasen', 'neuen', 'Wagen', 'Ras', 'Garten']
17	B	Danach verabschiedete die Tochter den	Rasenmäher	Gärtner	['Gärtner', 'Rasen', 'Garten', 'jungen', 'Mann']
17	C	Danach halbierte die Tochter den	Rasenmäher	Gärtner	['Rasen', 'Platz', 'Gras', 'Kopf', 'Mä']
18	A	Entzückt entlohnte die Dame den	Juwelier	Edelstein	['Juwel', 'Edel', 'jungen', 'Diamanten', 'Schmuck']
18	B	Entzückt bestaunte die Dame den	Juwelier	Edelstein	['Edel', 'Schmuck', 'Juwel', 'Stein', 'Diamanten']
18	C	Entzückt würzte die Dame den	Juwelier	Edelstein	['Edel', 'Stein', 'Schmuck', 'Juwel', 'Diamanten']
19	A	Verblüfft nahm die Nichte den	Schraubenzieher	Mechaniker	['Schrauben', 'Schrau', 'Zauber', 'Hammer', 'Schlüssel']
19	B	Verblüfft bewunderte die Nichte den	Schraubenzieher	Mechaniker	['Zauber', 'Trick', 'Mut', 'Anblick', 'Umstand']
19	C	Verblüfft kochte die Nichte den	Schraubenzieher	Mechaniker	['Schrauben', 'Zauber', 'ganzen', 'Schrau', 'Schlüssel']
20	A	Daraufhin bestach der Mopedfahrer den	Polizisten	Autofahrer	['Polizisten', 'Beamten', 'M', 'Polizeibeamten', 'Polizei']
20	B	Daraufhin entschädigte der Mopedfahrer den	Polizisten	Autofahrer	['Unfall', 'Schaden', 'Fahrer', 'Autofahrer', 'Geschäd']
20	C	Daraufhin sortierte der Mopedfahrer den	Polizisten	Autofahrer	['Schaden', 'Wagen', 'M', 'beschädigten', 'Motor']
21	A	Daraufhin ermahnte die Freundin den	Segler	Strick	['Seg', 'Jungen', 'Mann', 'Boots', 'Segel']
21	B	Daraufhin schnappte die Freundin den	Segler	Strick	['Strick', 'Knoten', 'Seg', 'Jungen', 'Faden']
21	C	Daraufhin verschraubte die Freundin den	Segler	Strick	['Knoten', 'Strick', 'Seg', 'Faden', 'Gurt']
22	A	Dann versklavten die Piraten die	Einheimischen	Goldschätze	['Einheimischen', 'einheimischen', 'Gold', 'Männer', 'Piraten']
22	B	Dann raubten die Piraten die	Einheimischen	Goldschätze	['Gold', 'Piraten', 'Schätze', 'Schatz', 'Beute']
22	C	Dann wechselten die Piraten die	Einheimischen	Goldschätze	['Seiten', 'Fähr', 'Lager', 'Seite', 'Spur']
23	A	Sofort suchte der Junge den	Korb	Apfel	['Apfel', 'Korb', 'Vater', 'Weg', 'Jungen']
23	B	Sofort zerschnitt der Junge den	Korb	Apfel	['Korb', 'Apfel', 'Baum', 'Obst', 'Eimer']
23	C	Sofort schlug der Junge den	Korb	Apfel	['Apfel', 'Korb', 'Jungen', 'Kopf', 'Baum']
24	A	Danach beglückwünschte der Juror den	Sportler	Vater	['Vater', 'jungen', 'Sieger', 'Sohn', 'Athleten']
24	B	Danach entdeckte der Juror den	Sportler	Vater	['jungen', 'Vater', 'Talent', 'Jungen', 'Sport']
24	C	Danach öffnete der Juror den	Sportler	Vater	['Eltern', 'jungen', 'Jur', 'Jungen', 'Athleten']
25	A	Daraufhin umarmte der Mann die	Geschäftsfrau	Bar	['Frau', 'Geschäfts', 'junge', 'Dame', 'Bar']
25	B	Daraufhin bestaunte der Mann die	Geschäftsfrau	Bar	['Bar', 'schöne', 'Schönheit', 'neue', 'schönen']
25	C	Daraufhin sortierte der Mann die	Geschäftsfrau	Bar	['Bar', 'Möbel', 'Reste', 'Waren', 'Ware']
26	A	Lachend holte der Rentner das	Zelt	Kind	['Kind', 'kleine', 'alte', 'Kinder', 'Klein']
26	B	Lachend tätschelte der Rentner das	Zelt	Kind	['Kind', 'kleine', 'junge', 'Tier', 'alte']

to be continued on the next page

Table A.4.: (continued from previous page)

I	Con	Target Sentence	Target	Distractor	Predictions
26	C	Lachend aß der Rentner das	Zelt	Kind	['Kind', 'kleine', 'Tier', 'erste', 'neue']
27	A	Zurück in der Kantine kaufte der Lehrer die	Limonade	Tasche	['Tasche', 'Lim', 'Flasche', 'leere', 'Taschen']
27	B	Zurück in der Kantine fand der Lehrer die	Limonade	Tasche	['Tasche', 'leere', 'Lim', 'Kass', 'Geld']
27	C	Zurück in der Kantine unterrichtete der Lehrer die	Limonade	Tasche	['Schüler', 'Kinder', 'Klasse', 'anderen', 'Klassen']
28	A	Im Krankenhaus tröstete die Mutter die	Bergsteigerin	Spitzhacke	['Mutter', 'Bergsteiger', 'junge', 'Tochter', 'kleine']
28	B	Im Krankenhaus verwünschte die Mutter die	Bergsteigerin	Spitzhacke	['Spitz', 'Kletter', 'Rettung', 'Hilfe', 'Mutter']
28	C	Im Krankenhaus stapelte die Mutter die	Bergsteigerin	Spitzhacke	['Spitz', 'Tasche', 'ganze', 'Leiche', 'Mutter']
29	A	Dann umarmte die Praktikantin den	Förster	Hirsch	['Hirsch', 'Jäger', 'Hir', 'Hund', 'Wolf']
29	B	Dann beobachtete die Praktikantin den	Förster	Hirsch	['Hirsch', 'Jäger', 'Hir', 'Wald', 'Wolf']
29	C	Dann sammelte die Praktikantin den	Förster	Hirsch	['Hirsch', 'Rest', 'Stein', 'Jäger', 'ganzen']
30	A	Am Ende konsultierte der Richter den	Staatsanwalt	Angeklagten	['Staatsanwalt', 'Angeklagten', 'Zeugen', 'Richter', 'Anwalt']
30	B	Am Ende befragte der Richter den	Staatsanwalt	Angeklagten	['Angeklagten', 'Staatsanwalt', 'Zeugen', 'Beschuldigten', 'Mann']
30	C	Am Ende kopierte der Richter den	Staatsanwalt	Angeklagten	['Antrag', 'Staatsanwalt', 'Prozess', 'Angeklagten', 'Bericht']
31	A	Glücklich küsste die Braut den	Bräutigam	Ring	['Ring', 'Bräutig', 'Braut', 'Trau', 'Altar']
31	B	Glücklich bewunderte die Braut den	Bräutigam	Ring	['Ring', 'Bräutig', 'Schmuck', 'Braut', 'goldenen']
31	C	Glücklich vereinfachte die Braut den	Bräutigam	Ring	['Ablauf', 'Blick', 'Hochzeits', 'Abschied', 'Augenblick']
32	A	Daraufhin faltete das Burgfräulein den	Umhang	Ritter	['Um', 'Mantel', 'um', 'Ring', 'Schleier']
32	B	Daraufhin verspottete das Burgfräulein den	Umhang	Ritter	['Ritter', 'Burg', 'König', 'Turnier', 'Um']
32	C	Daraufhin entleerte das Burgfräulein den	Umhang	Ritter	['Um', 'Mantel', 'Ring', 'um', 'Gürtel']
33	A	Daraufhin begleitete die Aufsichtsperson die	Besucherin	Eintrittskarte	['Besuch', 'Besucher', 'Person', 'Gruppe', 'Frau']
33	B	Daraufhin studierte die Aufsichtsperson die	Besucherin	Eintrittskarte	['Weg', 'Karte', 'Karten', 'Route', 'Umgebung']
33	C	Daraufhin erfand die Aufsichtsperson die	Besucherin	Eintrittskarte	['Weg', 'Route', 'Karte', 'Abkürzung', 'Stelle']
34	A	Dann grüßte der Händler die	Sklaven	Gewürzen	['Sklaven', 'Skla', 'Händler', 'beiden', 'Frau']
34	B	Dann kaufte der Händler die	Sklaven	Gewürzen	['Gewürze', 'Gew', 'Ware', 'Sklaven', 'Gewür']
34	C	Dann versiegelte der Händler die	Sklaven	Gewürzen	['Ware', 'Gew', 'Tasche', 'Tür', 'Kleider']
35	A	Dann drückte das Kind den	Teddybär	Ladenbesitzer	['Teddy', 'Bären', 'Laden', 'Jungen', 'Knopf']
35	B	Dann entlohnte das Kind den	Teddybär	Ladenbesitzer	['Laden', 'Teddy', 'Verkäufer', 'Jungen', 'Besitzer']
35	C	Dann bastelte das Kind den	Teddybär	Ladenbesitzer	['Teddy', 'Bären', 'kleinen', 'Jungen', 'ersten']
36	A	Daraufhin entlohnte die Hundeliebhaberin den	Nachbarn	Welpe	['Hund', 'Welpen', 'Nachbarn', 'Besitzer', 'Wel']
36	B	Daraufhin drückte die Hundeliebhaberin den	Nachbarn	Welpe	['Hund', 'Welpen', 'Nachbarn', 'Nachbar', 'Wel']
36	C	Daraufhin sortierte die Hundeliebhaberin den	Nachbarn	Welpe	['Welpen', 'Hund', 'Wel', 'Vier', 'kleinen']
37	A	Dann bemitleidete der Schuhverkäufer den	Kunden	Dieb	['Dieb', 'Kunden', 'Schuh', 'Laden', 'Verkäufer']
37	B	Dann verfolgte der Schuhverkäufer den	Kunden	Dieb	['Dieb', 'Täter', 'Räuber', 'Einbrecher', 'Kunden']
37	C	Dann hinterlegte der Schuhverkäufer den	Kunden	Dieb	['Geld', 'Schuh', 'Schlüssel', 'Beutel', 'Dieb']
38	A	Dann ermahnte der Eskimo die	Freundin	Robbe	['Rob', 'Freundin', 'Frau', 'Braut', 'junge']
38	B	Dann erschoss der Eskimo die	Freundin	Robbe	['Rob', 'Frau', 'junge', 'beiden', 'Robben']
38	C	Dann verpackte der Eskimo die	Freundin	Robbe	['Rob', 'Frau', 'Robben', 'Tiere', 'Hunde']
39	A	Dann vergütete die Tänzerin die	Babysitterin	Tochter	['Babys', 'Kinder', 'Nacht', 'Geburt', 'Wartezeit']
39	B	Dann umarmte die Tänzerin die	Babysitterin	Tochter	['Babys', 'Kleine', 'Kinder', 'beiden', 'Tochter']
39	C	Dann stapelte die Tänzerin die	Babysitterin	Tochter	['Babys', 'Baby', 'Kinder', 'Kleider', 'Wäsche']
40	A	Dann verhandelte der Minister mit dem	Präsidenten	Berater	['Präsidenten', 'Botschafter', 'Minister', 'Außenminister', 'Ministerpräsidenten']
40	B	Dann feierte der Minister mit dem	Präsidenten	Berater	['Präsidenten', 'Ministerpräsidenten', 'Premierminister', 'Minister', 'Premier']
40	C	Dann schminkte sich der Minister mit dem	Präsidenten	Berater	['Präsidenten', 'Minister', 'Premierminister', 'Gesicht', 'Innenminister']
41	A	Im Hotel polierte die Trainerin den	Pokal	Athlet	['Pokal', 'Sieger', 'Pokals', 'Preis', 'Ball']
41	B	Im Hotel bejubelte die Trainerin den	Pokal	Athlet	['Pokal', 'Sieg', 'Sieger', 'Erfolg', 'ersten']
41	C	Im Hotel verspeiste die Trainerin den	Pokal	Athlet	['Pokal', 'Sieger', 'Preis', 'Kuchen', 'frisch']
42	A	Daraufhin nahm der Autor den	Ball	Hund	['Hund', 'Ball', 'Vogel', 'Jungen', 'nächsten']
42	B	Daraufhin tätschelte der Autor den	Ball	Hund	['Hund', 'Ball', 'Hasen', 'Hunde', 'Welpen']
42	C	Daraufhin zitierte der Autor den	Ball	Hund	['Hund', 'Autor', 'Satz', 'Titel', 'Ball']
43	A	Daraufhin fütterte die Oma das	Kaninchen	Kleinkind	['Kaninchen', 'Klein', 'kleine', 'Tier', 'Katzen']
43	B	Daraufhin streichelte die Oma das	Kaninchen	Kleinkind	['Kaninchen', 'Klein', 'kleine', 'Tier', 'Hasen']
43	C	Daraufhin strickte die Oma das	Kaninchen	Kleinkind	['Kaninchen', 'Tier', 'kleine', 'Hasen', 'Fell']
44	A	Letztendlich suspendierte der Geschäftsführer den	Torwart	Coach	['Torwart', 'Trainer', 'Torhüter', 'Spieler', 'Kapitän']
44	B	Letztendlich umarmte der Geschäftsführer den	Torwart	Coach	['Torwart', 'Torhüter', 'Trainer', 'Keeper', 'Spieler']
44	C	Letztendlich reparierte der Geschäftsführer den	Torwart	Coach	['Torwart', 'Fehler', 'Schaden', 'Ball', 'Tor']
45	A	Amüsiert kaufte der Referendar die	Lose	Schüler	['L', 'ersten', 'Los', 'Lo', 'Tom']
45	B	Amüsiert musterte der Referendar die	Lose	Schüler	['Schüler', 'Kinder', 'Anwesenden', 'anwesenden', 'Mitschüler']
45	C	Amüsiert betrat der Referendar die	Lose	Schüler	['Bühne', 'L', 'Markt', 'Halle', 'Räumlichkeiten']
46	A	Dann nahm der Doktor die	Spritzen	Zwillinge	['Sprit', 'beiden', 'Kinder', 'drei', 'zwei']
46	B	Dann verwechselte der Doktor die	Spritzen	Zwillinge	['Sprit', 'beiden', 'zwei', 'drei', 'Zwillinge']
46	C	Dann bastelte der Doktor die	Spritzen	Zwillinge	['Sprit', 'ersten', 'kleinen', 'beiden', 'T']
47	A	Schnell zückte der Soldat die	Waffe	Zivilistin	['Pistole', 'Waffe', 'Flin', 'Hand', 'Fern']
47	B	Schnell versteckte der Soldat die	Waffe	Zivilistin	['Waffe', 'Zivil', 'Pistole', 'Soldaten', 'Waffen']
47	C	Schnell durchkämmte der Soldat die	Waffe	Zivilistin	['Straßen', 'Straße', 'Häuser', 'Gegend', 'Umgebung']
48	A	An diesem Abend spielte das Orchester das	Stück	Publikum	"['Stück', 'Werk', 'erste', 'Lied', '""']"
48	B	An diesem Abend verzauberte das Orchester das	Stück	Publikum	['Publikum', 'Orchester', 'gesamte', 'Haus', 'ganze']
48	C	An diesem Abend engagierte das Orchester das	Stück	Publikum	['Orchester', 'Ensemble', 'junge', 'Publikum', 'Duo']
49	A	Dort begrüßte der Doktorand die	Betreuerin	Arbeit	['Betreuer', 'Prüfer', 'beiden', 'neue', 'anderen']
49	B	Dort verteidigte der Doktorand die	Betreuerin	Arbeit	['Arbeit', 'Arbeiten', 'Doktor', 'Dissertation', 'Prüfer']
49	C	Dort reparierte der Doktorand die	Betreuerin	Arbeit	['Maschine', 'Geräte', 'Schreib', 'Lampe', 'Tür']

to be continued on the next page

Table A.4.: (continued from previous page)

I	Con	Target Sentence	Target	Distractor	Predictions
50	A	Daraufhin segnete der Pfarrer die	Protestantin	Reise	['Reise', 'Ankunft', 'Abreise', 'Protest', 'Abfahrt']
50	B	Daraufhin bewilligte der Pfarrer die	Protestantin	Reise	['Reise', 'Ausreise', 'Abreise', 'Über', 'Einreise']
50	C	Daraufhin las der Pfarrer die	Protestantin	Reise	['Bibel', 'Zeitung', 'Briefe', 'Liste', 'Protest']
51	A	In der Praxis erfragte der Therapeut die	Symptome	Erzieherin	['Symptome', 'Ursache', 'Diagnose', 'Krankheits', 'Kinder']
51	B	In der Praxis behandelte der Therapeut die	Symptome	Erzieherin	['Kinder', 'Symptome', 'Patienten', 'Patientin', 'Betroffenen']
51	C	In der Praxis tauschte der Therapeut die	Symptome	Erzieherin	['Medikamente', 'Diagnose', 'Patienten', 'Symptome', 'Therapie']
52	A	Dann malte die Designerin die	Eisbären	Geschichten	['Eisbären', 'Bilder', 'Figuren', 'Geschichten', 'Tiere']
52	B	Dann veranschaulichte die Designerin die	Eisbären	Geschichten	['Geschichte', 'Entstehung', 'Geschichten', 'Hintergründe', 'Eisbären']
52	C	Dann leerte die Designerin die	Eisbären	Geschichten	['Bücher', 'Kart', 'Taschen', 'Seiten', 'Bilder']
53	A	Daraufhin schrieb der Architekt die	Rede	Turnhalle	['Rede', 'Idee', 'Namen', 'Turnhalle', 'erste']
53	B	Daraufhin entwarf der Architekt die	Rede	Turnhalle	['Turnhalle', 'Halle', 'neue', 'Sporthalle', 'Pläne']
53	C	Daraufhin rief der Architekt die	Rede	Turnhalle	['Stadt', 'Gemeinderäte', 'Stadtverwaltung', 'Stadträte', 'Bevölkerung']
54	A	Dann verabschiedete der Gitarrist die	Agentin	Sängerin	['Sängerin', 'Agen', 'Band', 'Agentur', 'beiden']
54	B	Dann traf der Gitarrist die	Agentin	Sängerin	['Sängerin', 'Agen', 'erste', 'richtige', 'junge']
54	C	Dann kaufte der Gitarrist die	Agentin	Sängerin	['Gitarre', 'Instrumente', 'Sängerin', 'Aufnahme', 'Geige']
55	A	Danach umarmte der Kurator die	Galeristin	Skulptur	['Skulptur', 'Künstlerin', 'Skulpturen', 'Statue', 'Kunst']
55	B	Danach betrachtete der Kurator die	Galeristin	Skulptur	['Skulptur', 'Skulpturen', 'Plastik', 'Statue', 'Arbeit']
55	C	Danach sammelte der Kurator die	Galeristin	Skulptur	['Skulptur', 'Skulpturen', 'Kunstwerke', 'beiden', 'einzelnen']
56	A	Dann betrat die Kommilitonin die	Toilette	Wimperntusche	['Toilette', 'Toiletten', 'W', 'Wohnung', 'Dusch']
56	B	Dann benutzte die Kommilitonin die	Toilette	Wimperntusche	['W', 'Dusch', 'Toilette', 'Komm', 'Hand']
56	C	Dann las die Kommilitonin die	Toilette	Wimperntusche	['W', 'Geschichte', 'Toilette', 'Zeitung', 'Anzeige']
57	A	Dann streichelte der Verbrecher den	Schäferhund	Ermittler	['Schäfer', 'Hund', 'Schaf', 'Kopf', 'Schädel']
57	B	Dann erschoss der Verbrecher den	Schäferhund	Ermittler	['Schäfer', 'Hund', 'Schaf', 'Polizisten', 'Mann']
57	C	Dann faltete der Verbrecher den	Schäferhund	Ermittler	['Schäfer', 'Hund', 'Kopf', 'Schaf', 'Hals']
58	A	Daraufhin verteidigte die Anwältin den	Beschuldigten	Kläger	['Angeklagten', 'Beschuldigten', 'Zeugen', 'Kläger', 'Mann']
58	B	Daraufhin entließ die Anwältin den	Beschuldigten	Kläger	['Angeklagten', 'Kläger', 'Zeugen', 'Beschuldigten', 'Anwalt']
58	C	Daraufhin schwenkte die Anwältin den	Beschuldigten	Kläger	['Kopf', 'Arm', 'Tisch', 'Blick', 'Stuhl']
59	A	Daraufhin angelte der Kumpel den	Flussbarsch	Jungen	['Fluss', 'Fisch', 'Bar', 'K', 'Bach']
59	B	Daraufhin beobachtete der Kumpel den	Flussbarsch	Jungen	['Fluss', 'Fisch', 'Jungen', 'K', 'Bar']
59	C	Daraufhin trocknete der Kumpel den	Flussbarsch	Jungen	['K', 'Fluss', 'Fisch', 'Bar', 'Teich']
60	A	Daraufhin verstellte die Gynäkologin die	Liege	Schwangere	['Liege', 'Beine', 'Stuhl', 'Blick', 'Sicht']
60	B	Daraufhin untersuchte die Gynäkologin die	Liege	Schwangere	['Patientin', 'Schwanger', 'Frau', 'Mutter', 'Liege']
60	C	Daraufhin verordnete die Gynäkologin die	Liege	Schwangere	['Aufnahme', 'sofortige', 'Abt', 'Geburt', 'Ent']

Appendix B.

Additional rERP Analyses

B.1. DBC19

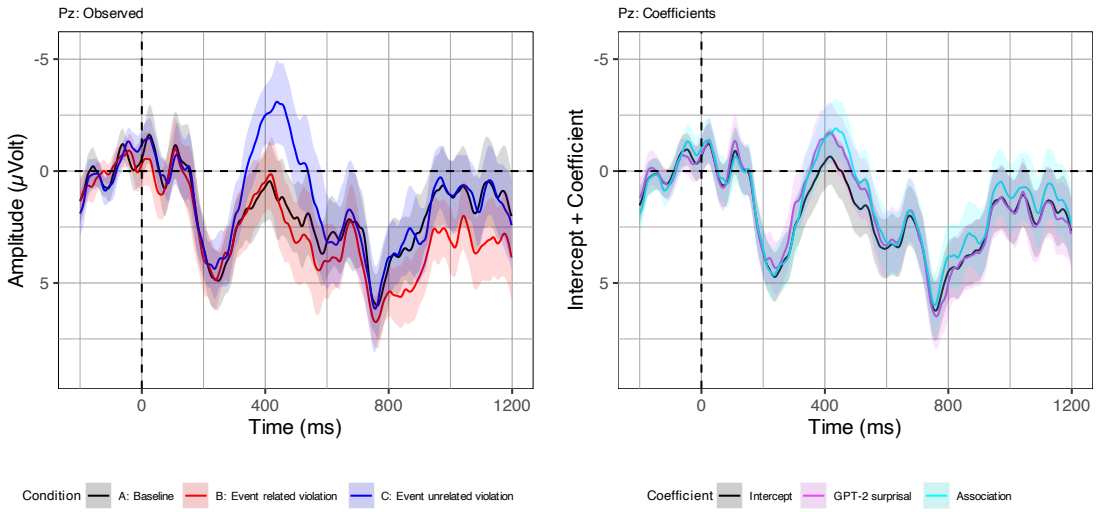

Figure B.1.: **DBC19**, **GPT-2 & Association**: observed voltages per condition (left) and the surprisal coefficient over time (right).

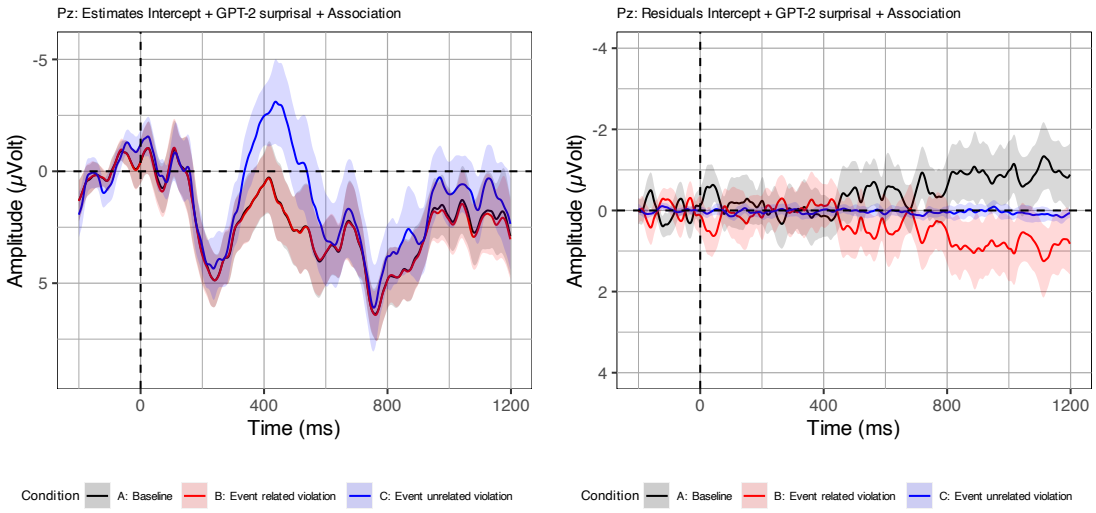

Figure B.2.: **DBC19**, **GPT-2 & Association**: estimated voltages (left) and residuals (right) per condition.

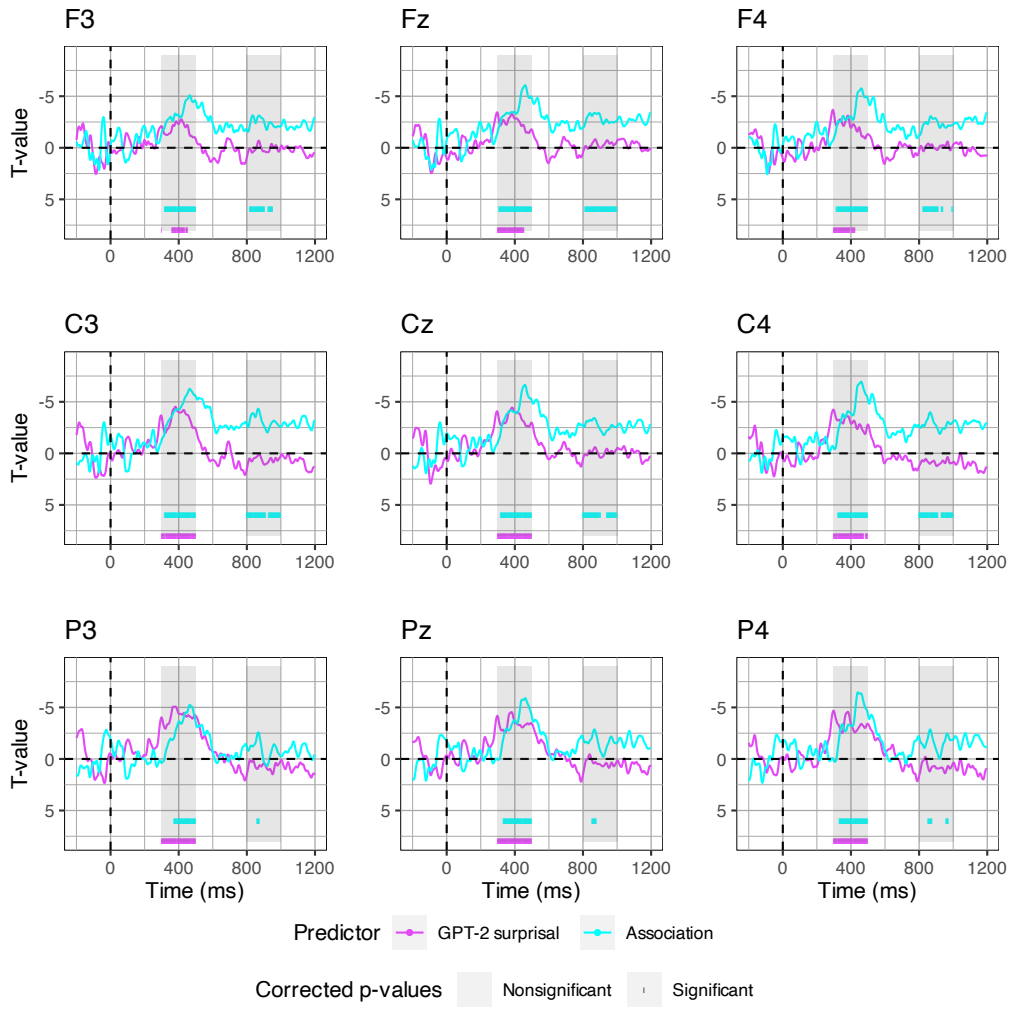

Figure B.3.: **DBC19**, **GPT-2 & Association**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

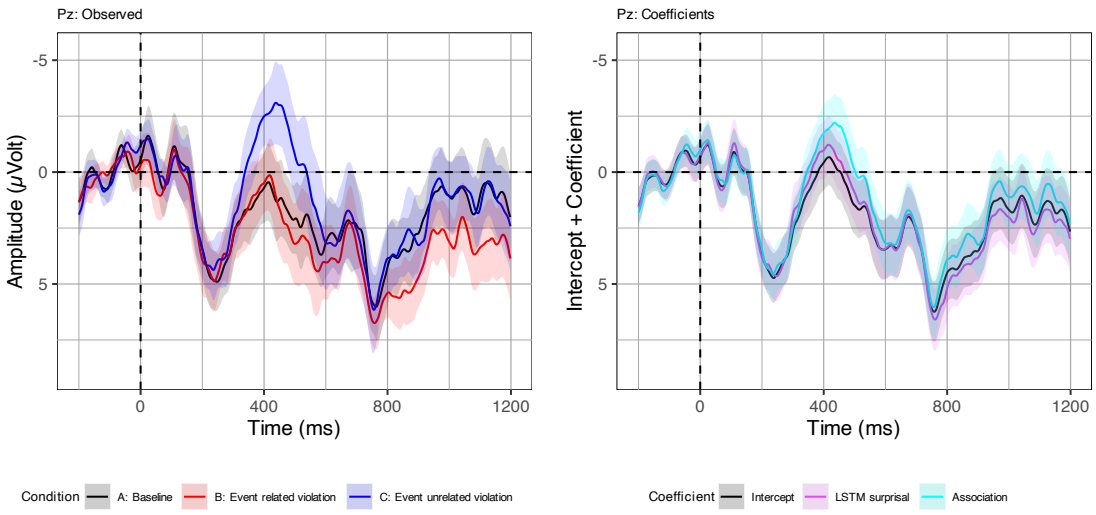

Figure B.4.: **DBC19**, **LSTM & Association**: observed voltages per condition (left) and the surprisal coefficient over time (right).

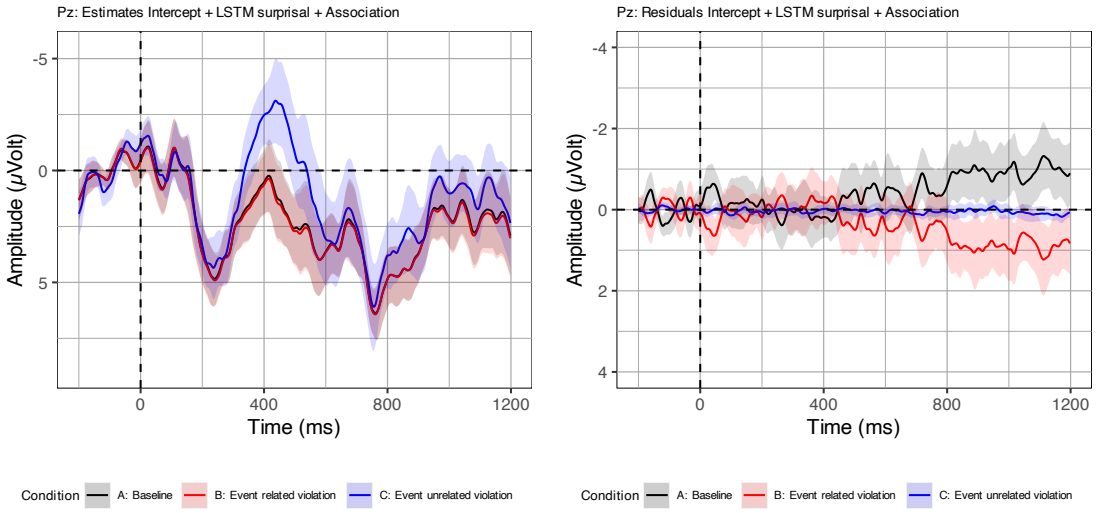

Figure B.5.: **DBC19**, **LSTM & Association**: estimated voltages (left) and residuals (right) per condition.

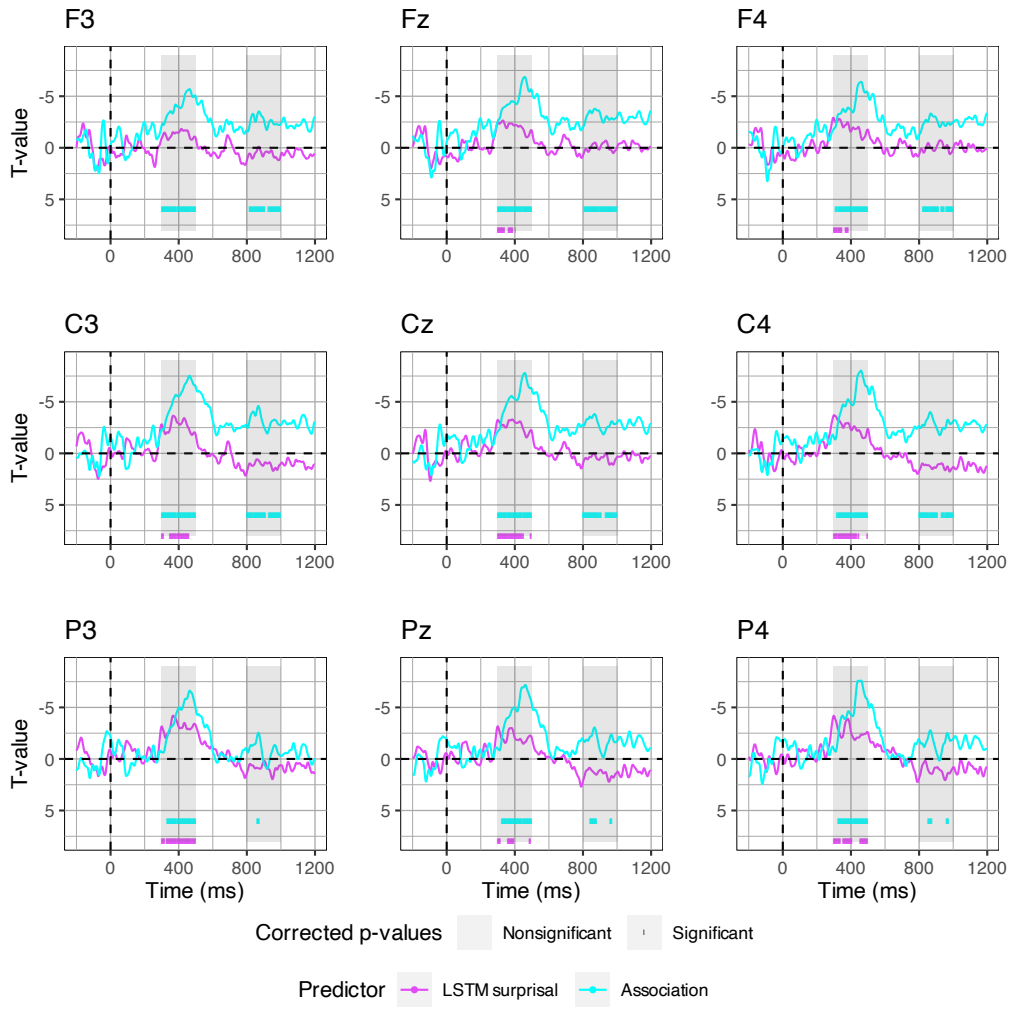

Figure B.6.: **DBC19**, **LSTM & Association**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

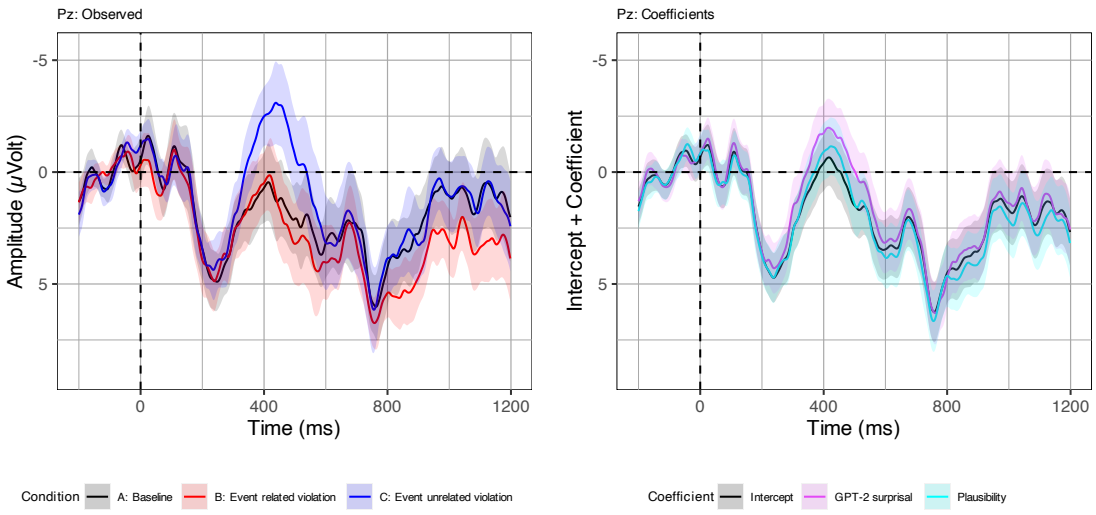

Figure B.7.: **DBC19**, **GPT-2 & Plausibility**: observed voltages per condition (left) and the surprisal coefficient over time (right).

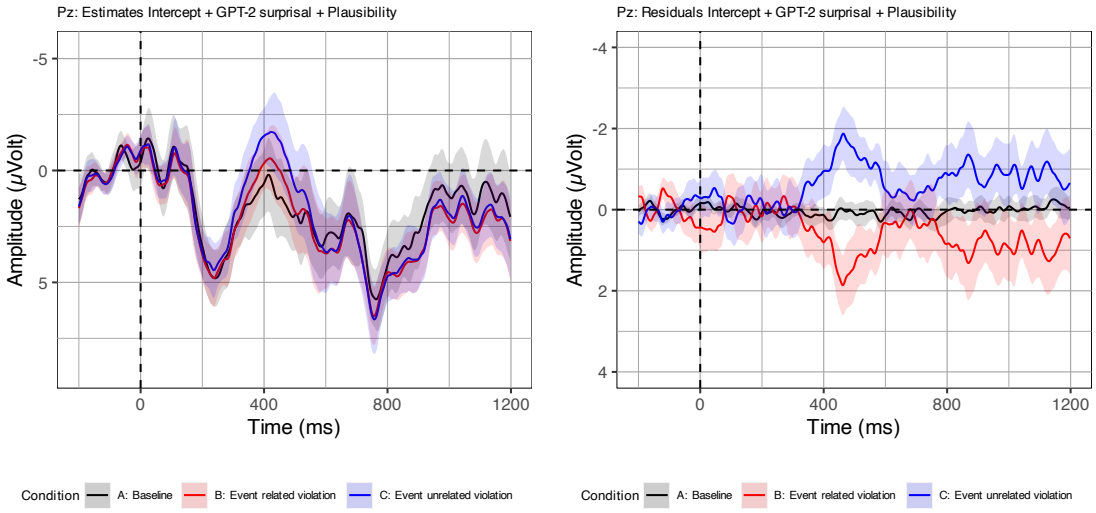

Figure B.8.: **DBC19**, **GPT-2 & Plausibility**: estimated voltages (left) and residuals (right) per condition.

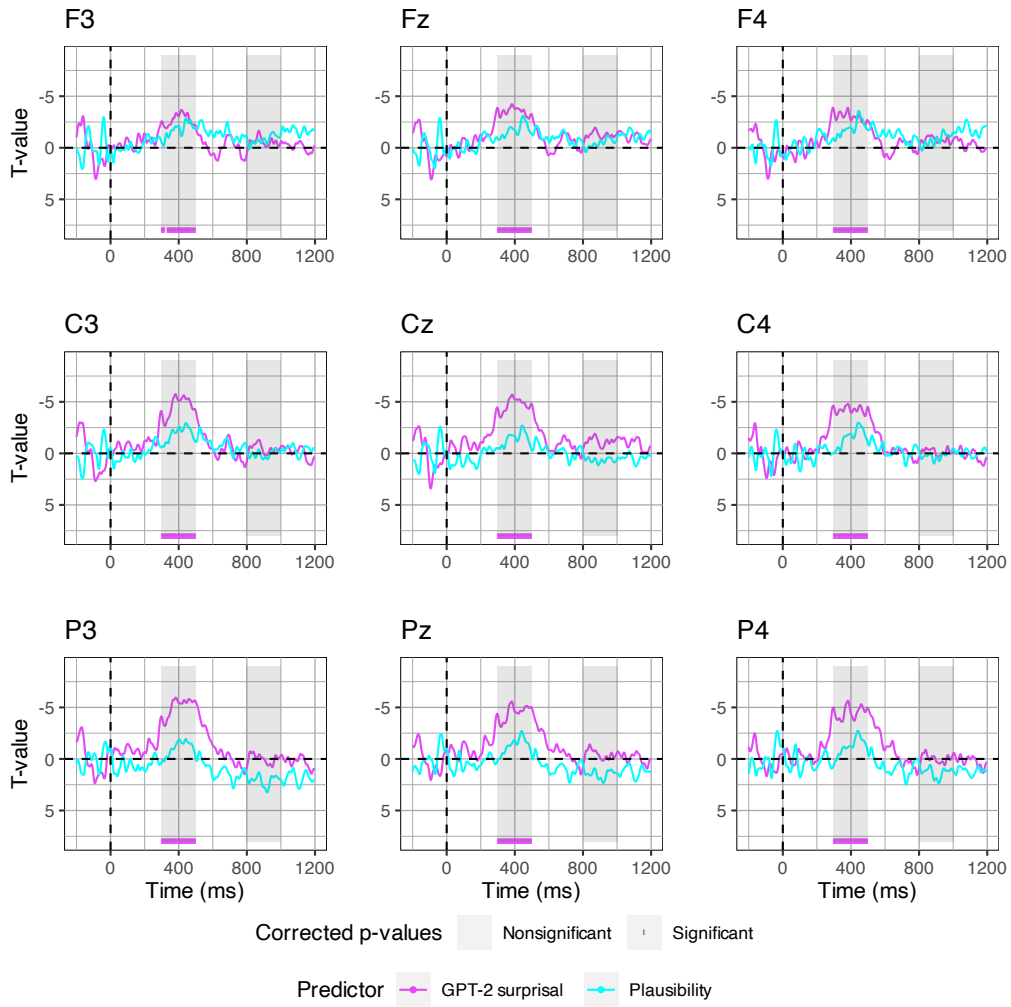

Figure B.9.: **DBC19**, **GPT-2 & Plausibility**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

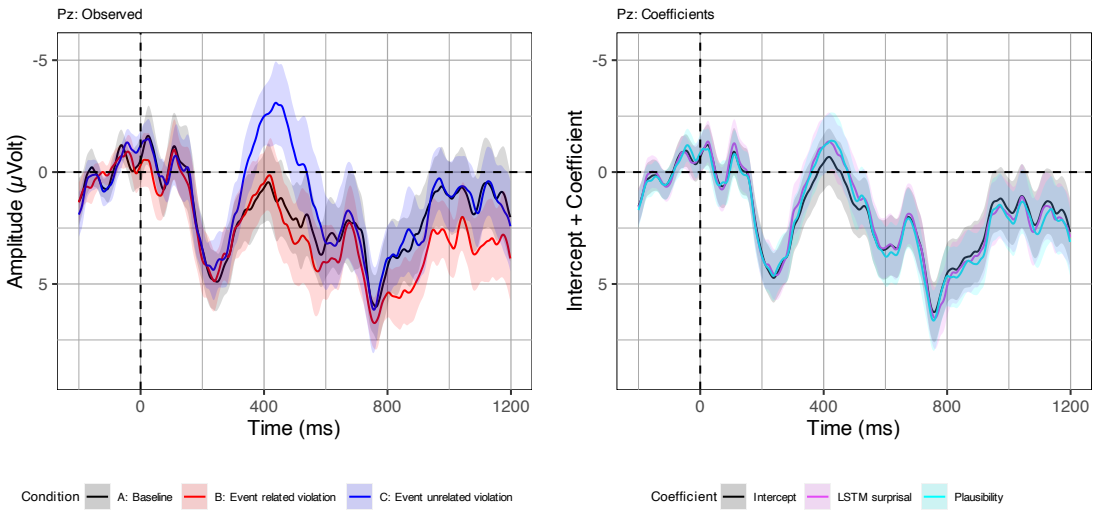

Figure B.10.: **DBC19**, **LSTM & Plausibility**: observed voltages per condition (left) and the surprisal coefficient over time (right).

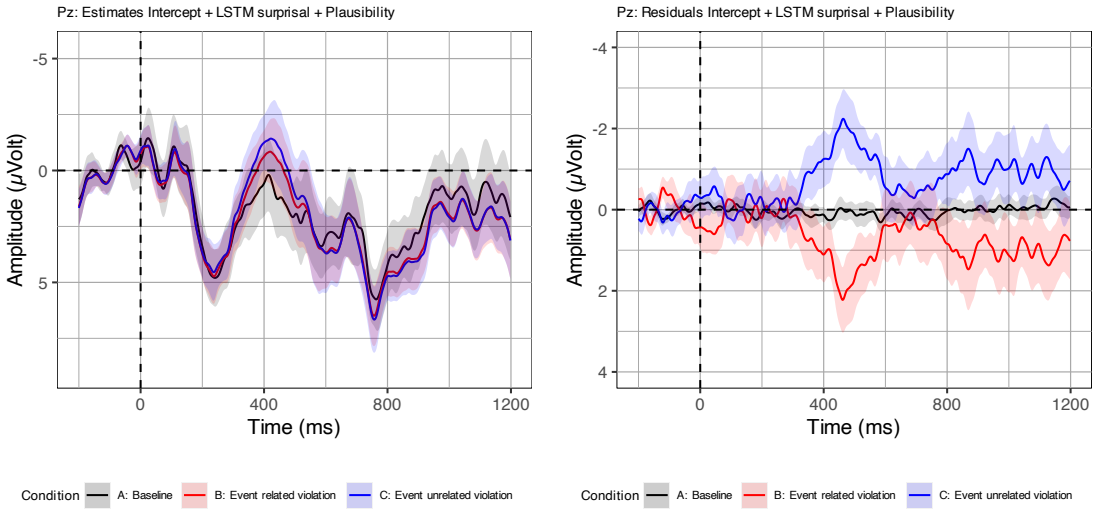

Figure B.11.: **DBC19**, **LSTM & Plausibility**: estimated voltages (left) and residuals (right) per condition.

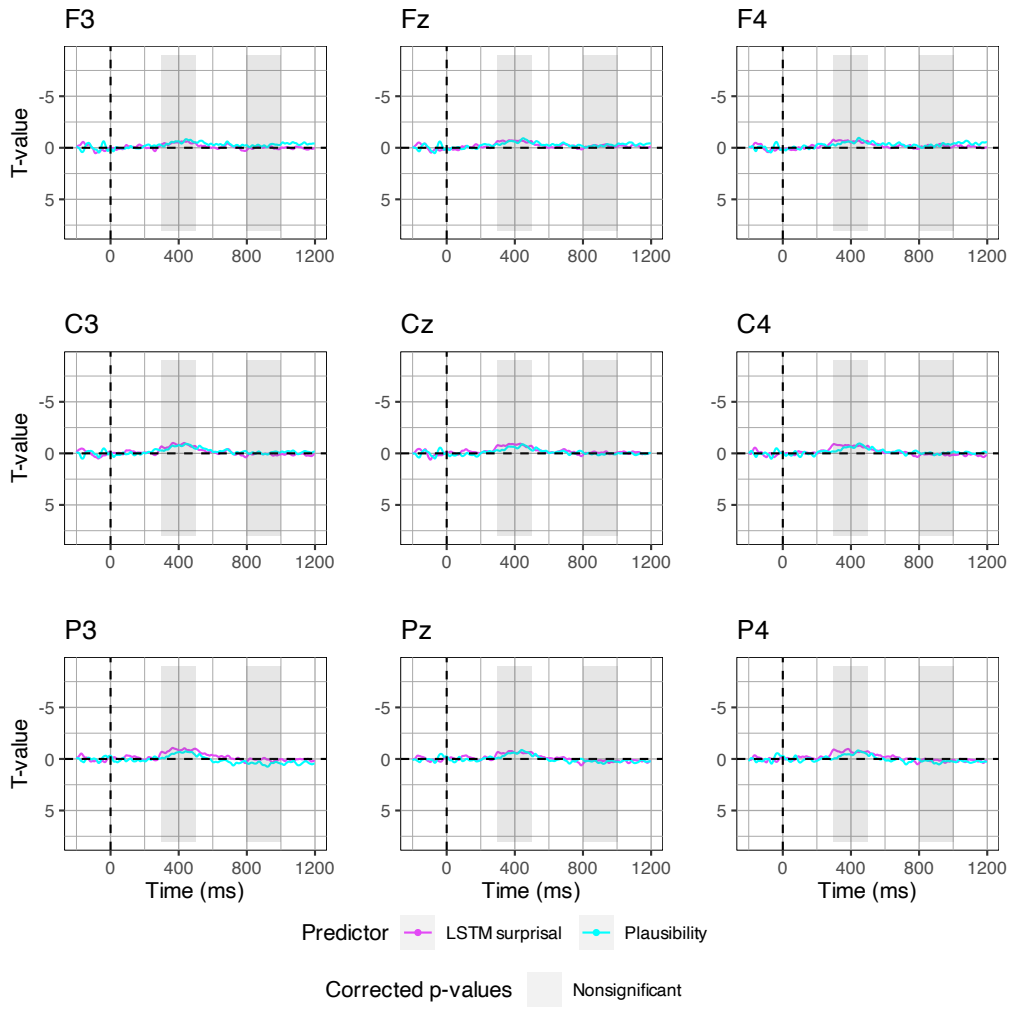

Figure B.12.: **DBC19**, **LSTM & Plausibility**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

B.2. DBC21

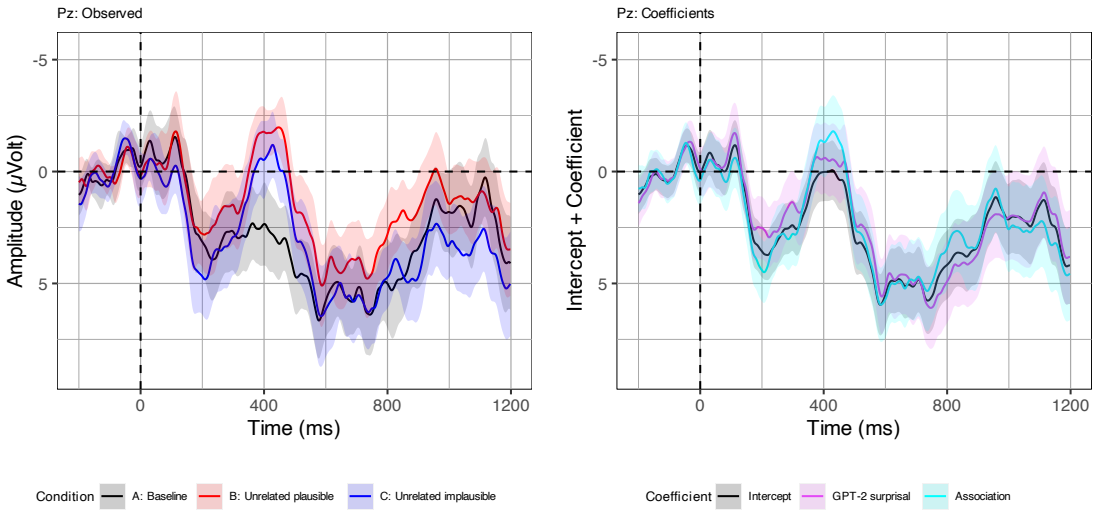

Figure B.13.: **DBC21**, **GPT-2 & Association**: observed voltages per condition (left) and the surprisal coefficient over time (right).

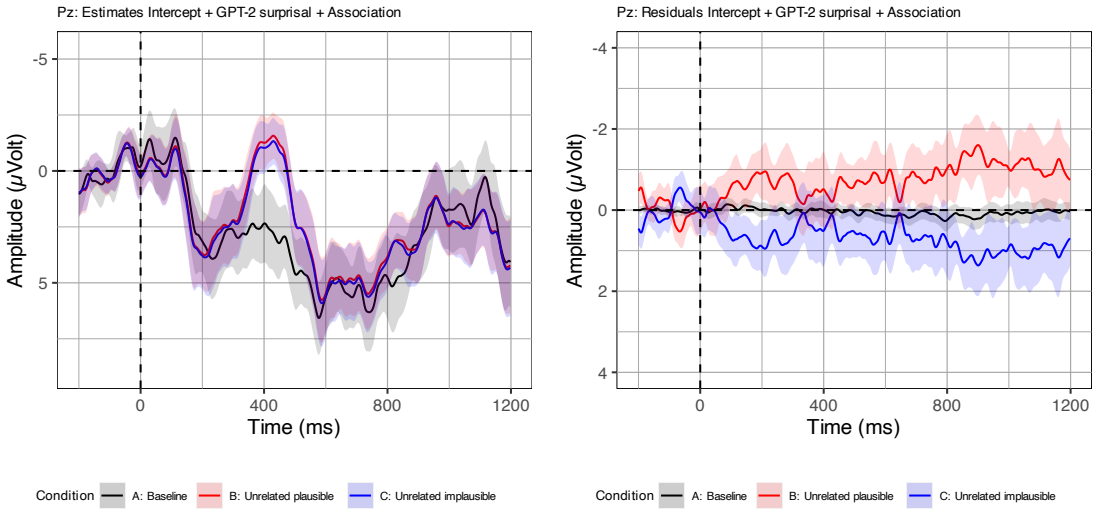

Figure B.14.: **DBC21**, **GPT-2 & Association**: estimated voltages (left) and residuals (right) per condition.

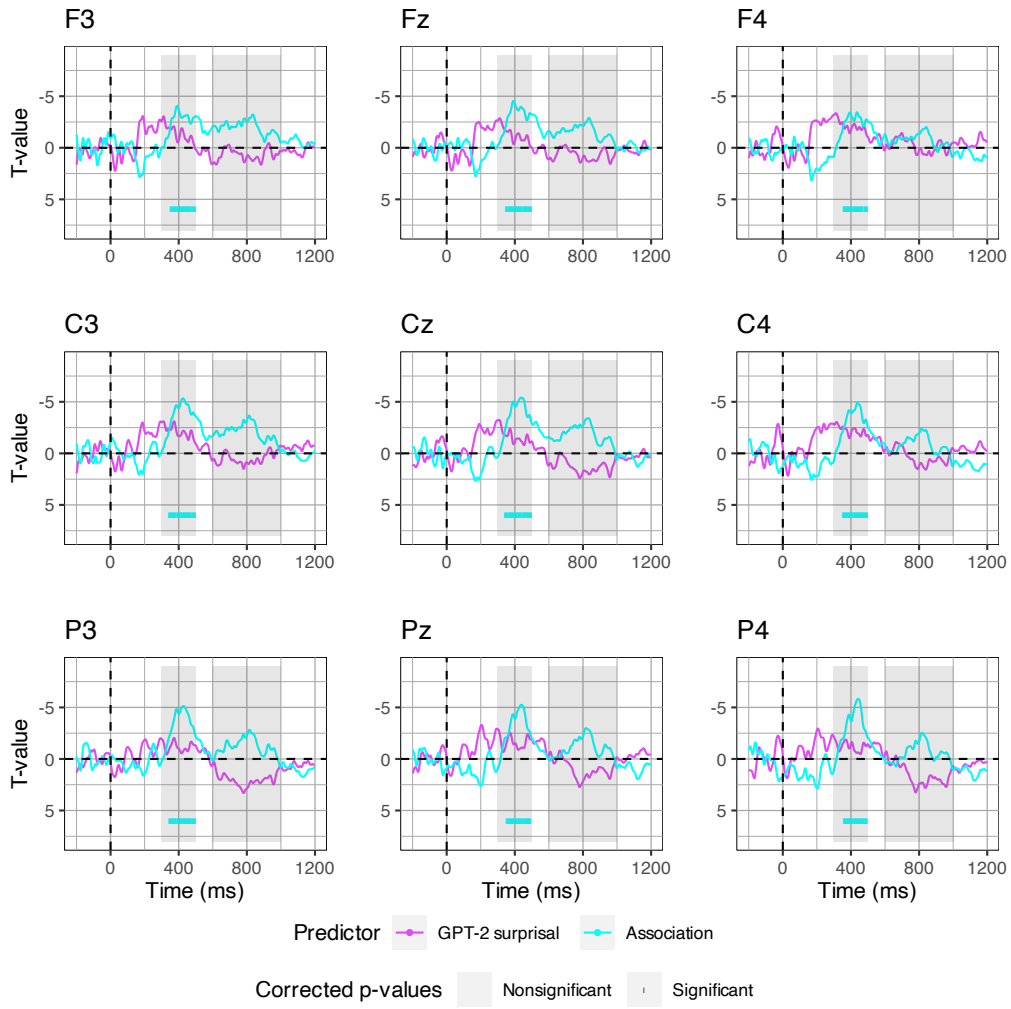

Figure B.15.: **DBC21, GPT-2 & Association**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

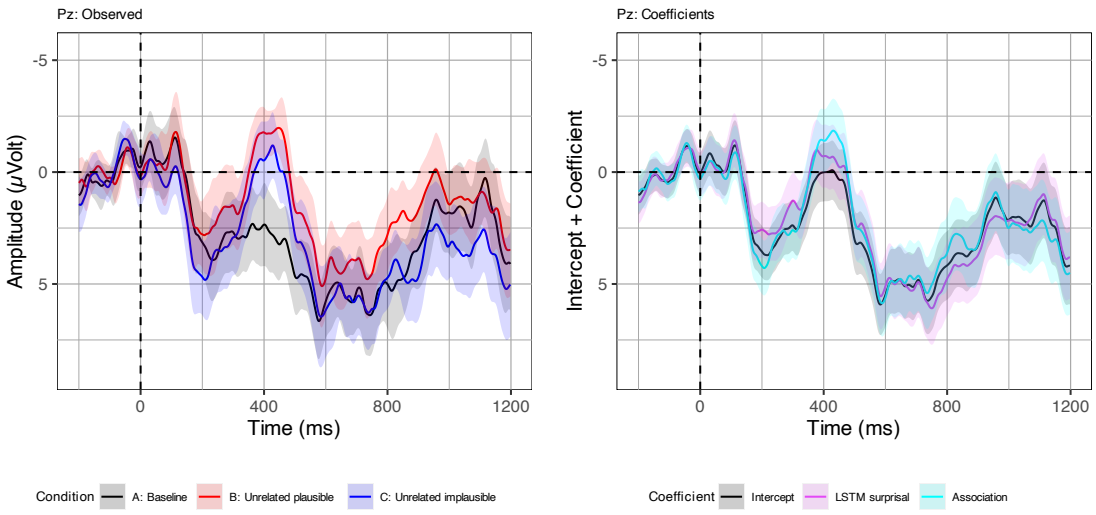

Figure B.16.: **DBC21**, **LSTM & Association**: observed voltages per condition (left) and the surprisal coefficient over time (right).

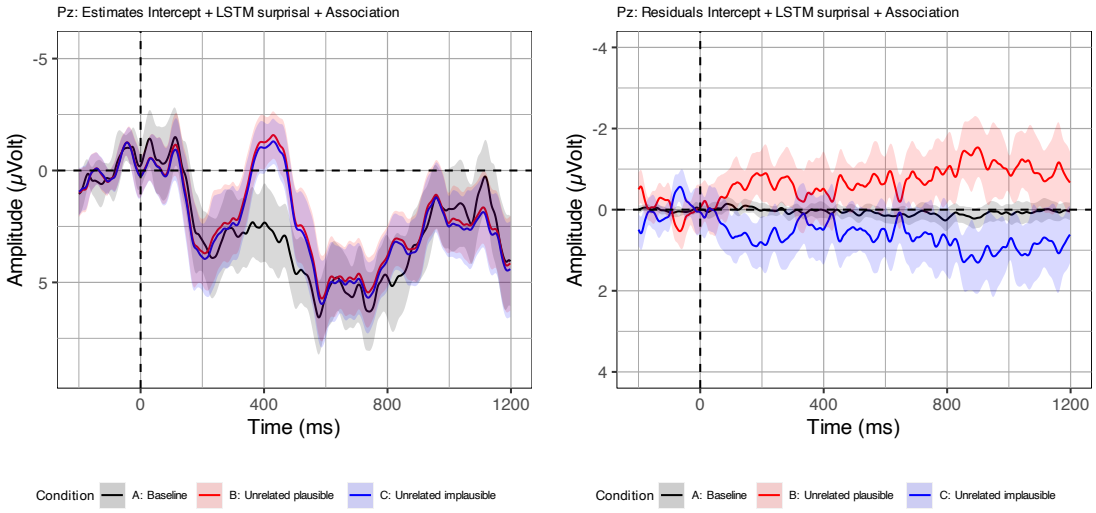

Figure B.17.: **DBC21**, **LSTM & Association**: estimated voltages (left) and residuals (right) per condition.

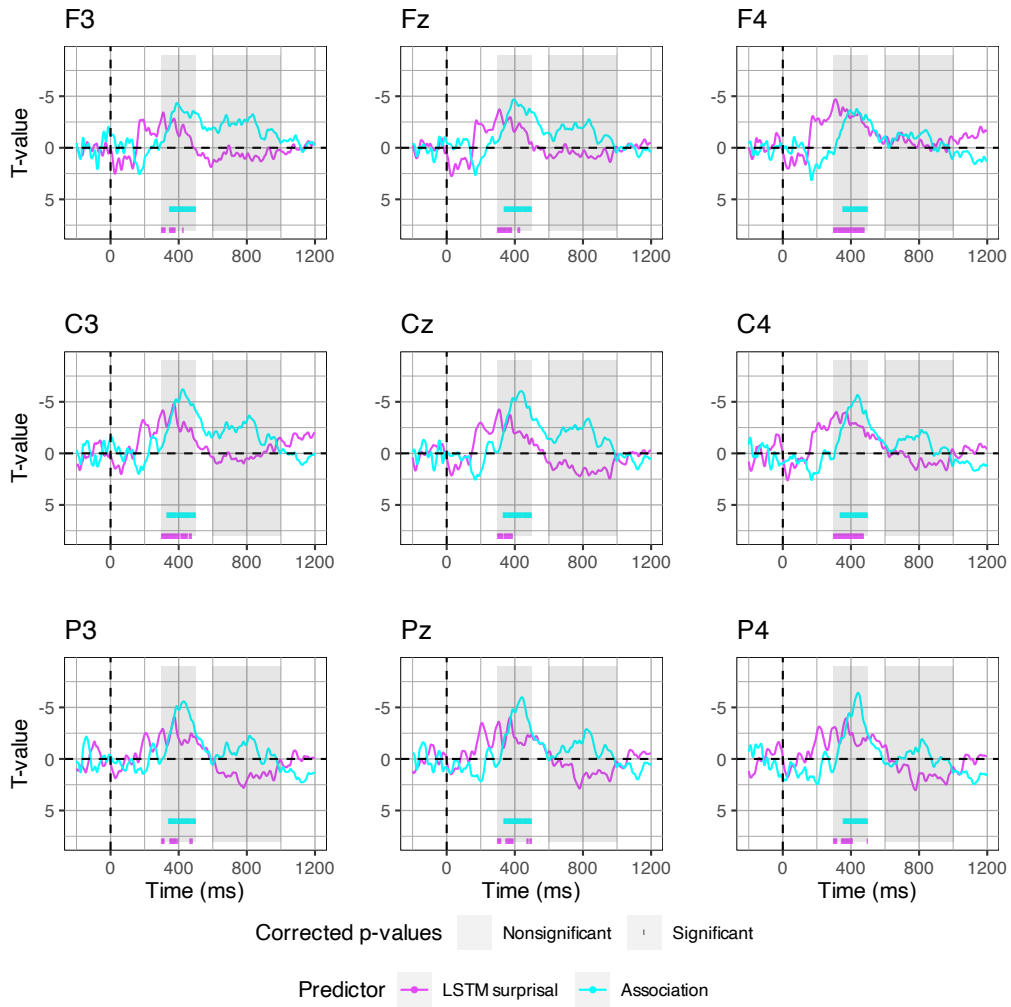

Figure B.18.: **DBC21**, **LSTM & Association**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

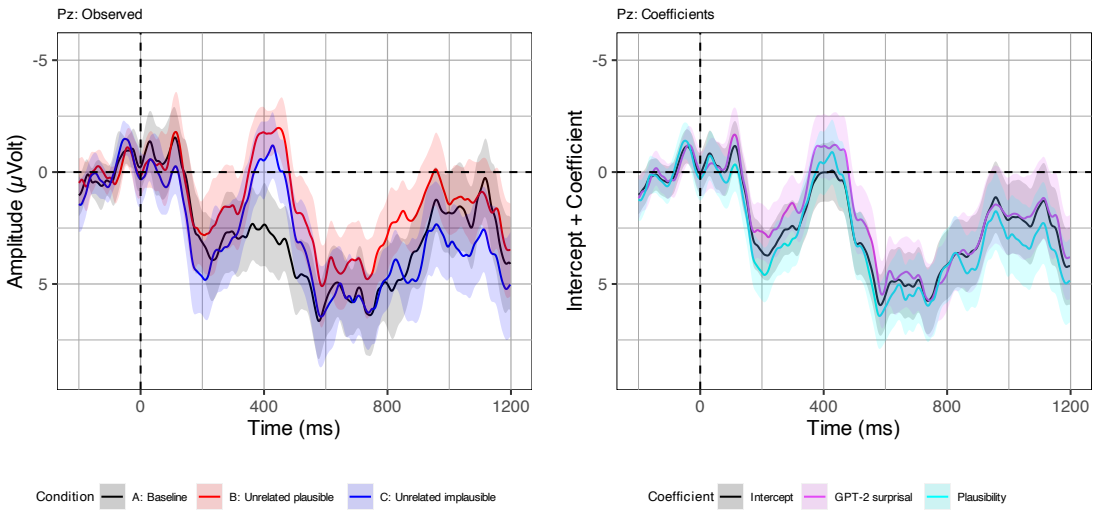

Figure B.19.: **DBC21**, **GPT-2 & Plausibility**: observed voltages per condition (left) and the surprisal coefficient over time (right).

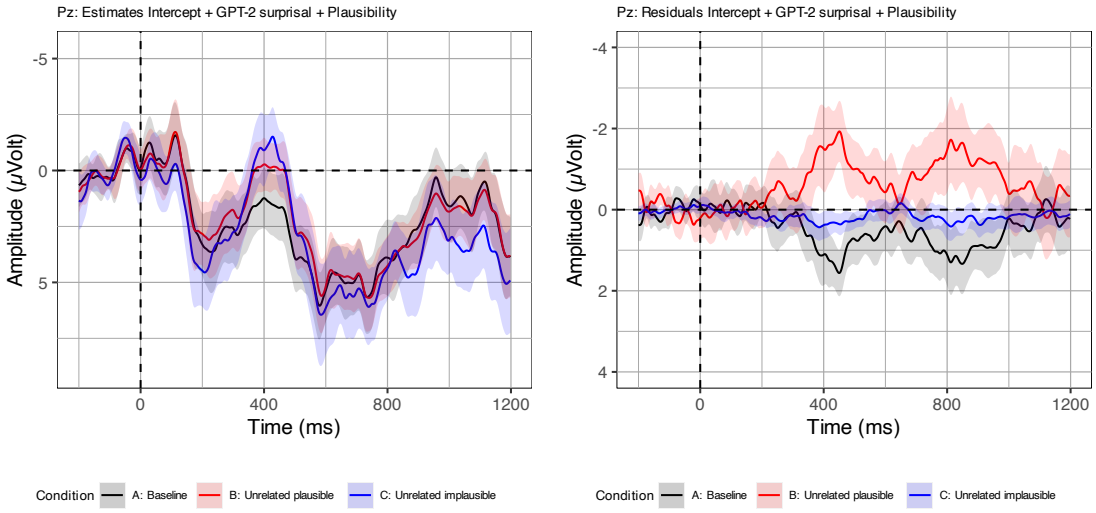

Figure B.20.: **DBC21**, **GPT-2 & Plausibility**: estimated voltages (left) and residuals (right) per condition.

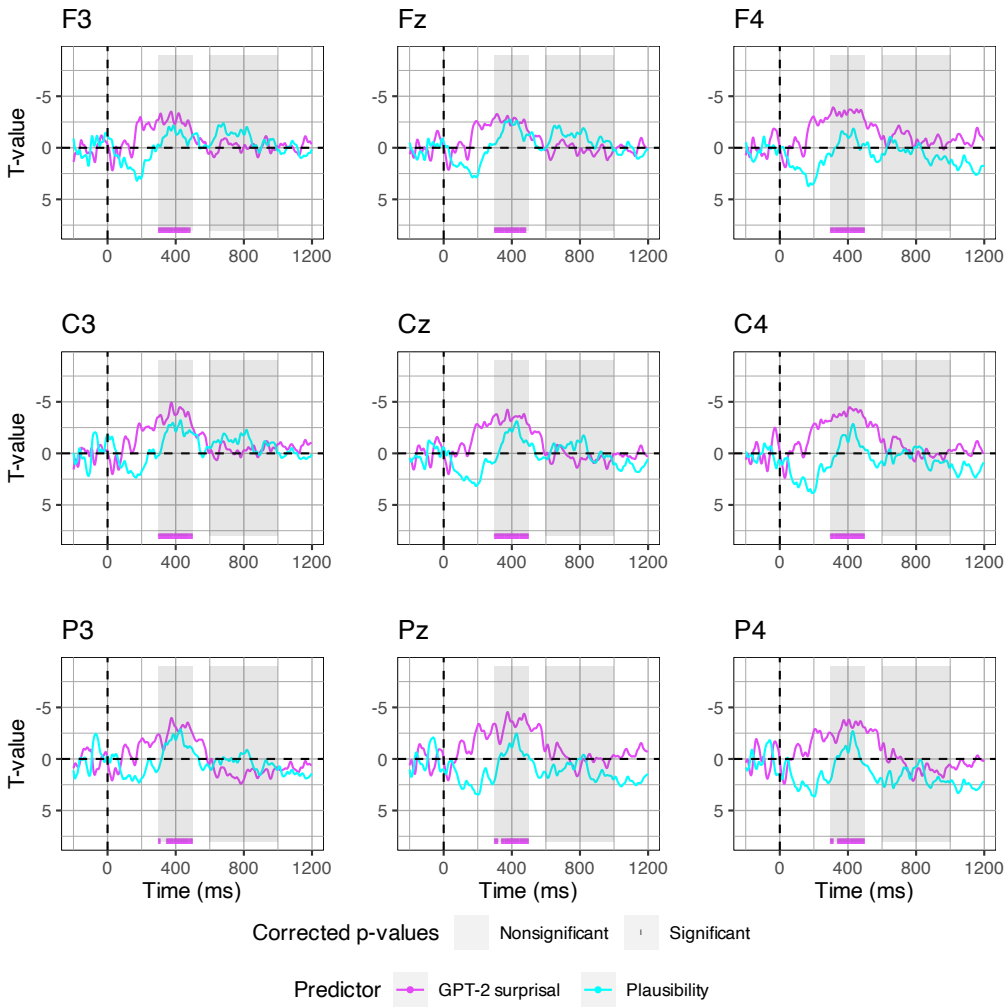

Figure B.21.: **DBC21, GPT-2 & Plausibility**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

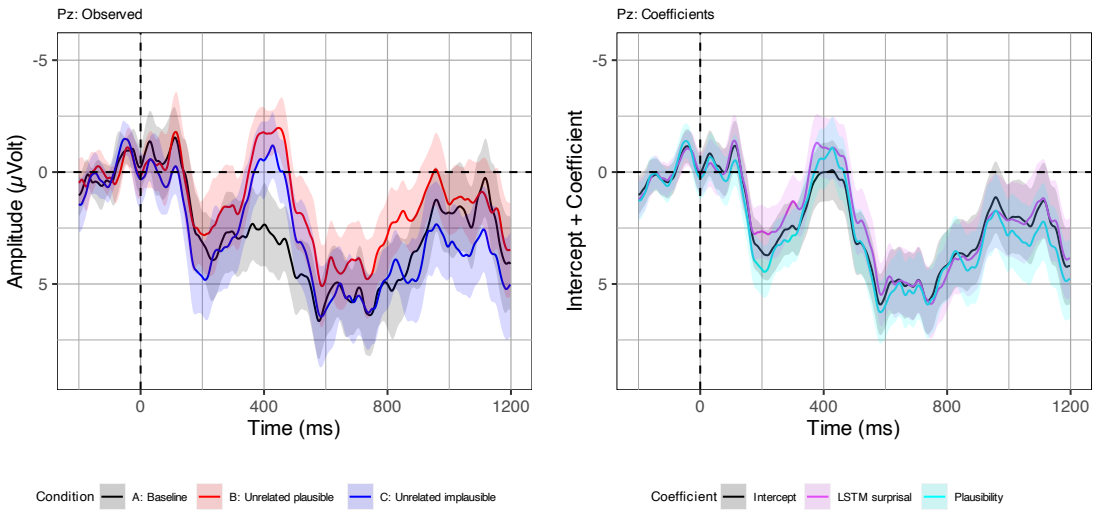

Figure B.22.: **DBC21**, **LSTM & Plausibility**: observed voltages per condition (left) and the surprisal coefficient over time (right).

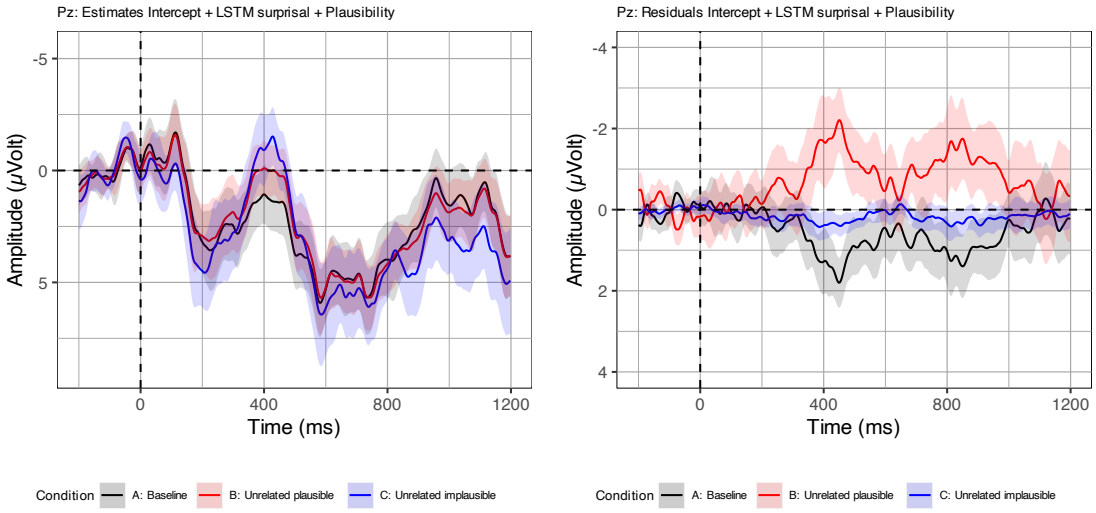

Figure B.23.: **DBC21**, **LSTM & Plausibility**: estimated voltages (left) and residuals (right) per condition.

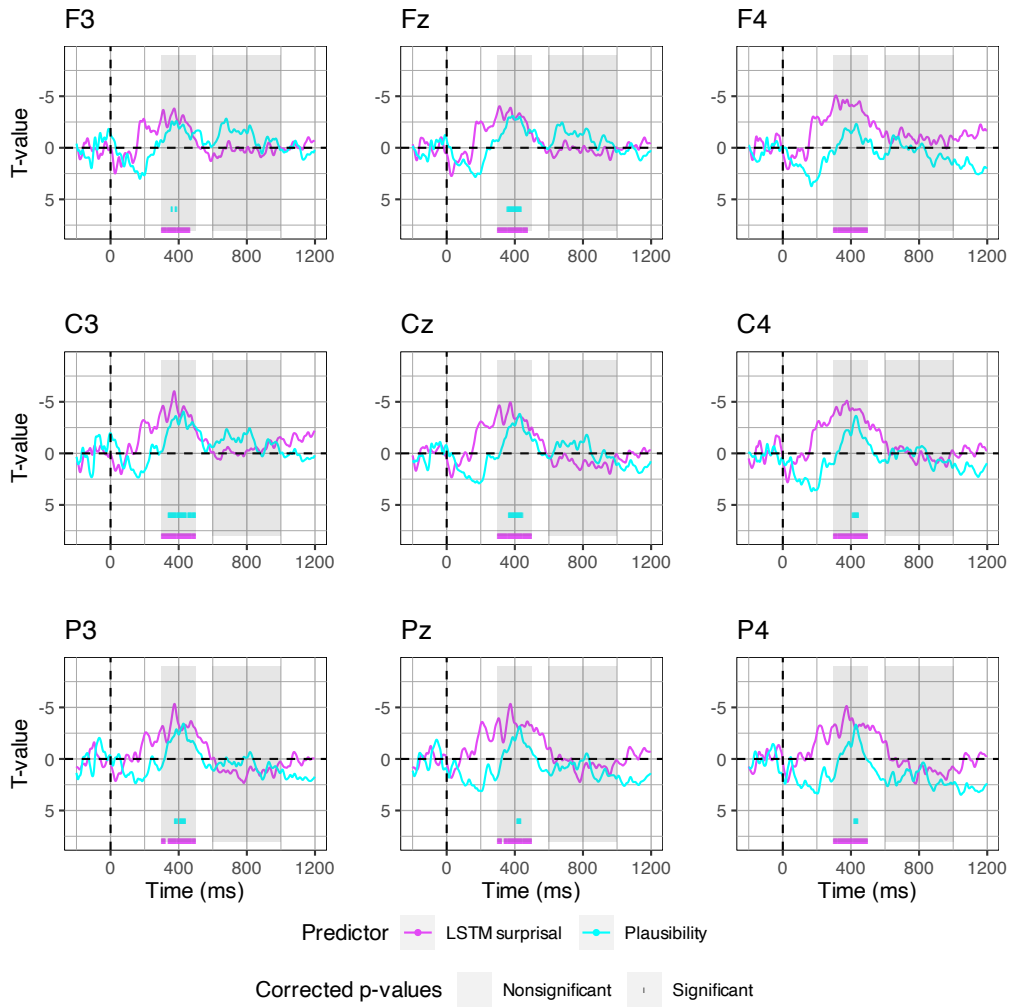

Figure B.24.: **DBC21**, **LSTM & Plausibility**: T-values that were obtained from across-subjects regression for the surprisal predictor, shown for nine central electrodes. The bars indicate time samples where p values were significant after correcting for multiple comparisons.

Appendix C.

LM Test Set

	Genre	Book	Author/Editor	Text	WordCount	Source	Link	License	License Link
	Fairy Tale	Japanische Märchen	Karl Alberti	Schlauheit schützt nicht vor Täuschung	214	Project Gutenberg	https://www.gutenberg.org/cache/epub/23393/pg23393-images.html	Public Domain	NA
	Fairy Tale	Rübezahl	Rosalie Koch	Das Rad	265	Project Gutenberg	https://www.gutenberg.org/cache/epub/37940/pg37940-images.html	Public Domain	NA
	Fairy Tale	Ludwig Bechsteins Märchenbuch	Ludwig Bechstein	Das Kätzchen und die Stricknadeln	307	Project Gutenberg	https://www.gutenberg.org/cache/epub/63465/pg63465-images.html#chap_165	Public Domain	NA
	Fairy Tale	Türkische Märchen	Friedrich Giese	Der Holzhauer, der zur Unzeit tanzte	311	Project Gutenberg	https://www.gutenberg.org/cache/epub/69949/pg69949-images.html	Public Domain	NA
	Fairy Tale	Rügen-Märchen	Ernst Moritz Arndt	Der Riese Balderich	382	Project Gutenberg	https://www.projekt-gutenberg.org/arndt/ruegen/chap011.html	Public Domain	NA
SUM					**1479**				
	Short story	NA	Dierk Seidel	Alles aus und vorbei	356	Kulturkater	https://www.kulturkater.de/alles-aus-und-vorbei/	CC BY-SA 3.0	https://creativecommons.org/licenses/by-sa/3.0/
	Short story	NA	Dierk Seidel	Der Baum	289	Kulturkater	https://www.kulturkater.de/geschichten-von-oben-prolog-der-baum/	CC BY-SA 3.0	https://creativecommons.org/licenses/by-sa/3.0/
	Short story	NA	Dierk Seidel	Prolog	299	Kulturkater	https://www.kulturkater.de/geschichten-von-oben-prolog-der-baum/	CC BY-SA 3.0	https://creativecommons.org/licenses/by-sa/3.0/
	Short story	NA	Dierk Seidel	Bloß nicht Zappeln	212	Kulturkater	https://www.kulturkater.de/bloss-nicht-zappeln/	CC BY-SA 3.0	https://creativecommons.org/licenses/by-sa/3.0/
	Short story	NA	Dierk Seidel	Ausgestiegen	310	Kulturkater	https://www.kulturkater.de/ausgestiegen/	CC BY-SA 3.0	https://creativecommons.org/licenses/by-sa/3.0/
SUM					**1466**				
	Newspaper	Der Standard - Kultur	Unknown	Hugo Race	313	10kGNAD	https://tblock.github.io/10kGNAD/	CC BY-NC-SA 4.0	https://creativecommons.org/licenses/by-nc-sa/4.0/
	Newspaper	Der Standard - Wissenschaft	Unknown	Wrack der Endeavour	281	10kGNAD	https://tblock.github.io/10kGNAD/	CC BY-NC-SA 4.0	https://creativecommons.org/licenses/by-nc-sa/4.0/
	Newspaper	Der Standard - Panorama	Unknown	Methan	274	10kGNAD	https://tblock.github.io/10kGNAD/	CC BY-NC-SA 4.0	https://creativecommons.org/licenses/by-nc-sa/4.0/
	Newspaper	Der Standard - Sport	Unknown	FBI	261	10kGNAD	https://tblock.github.io/10kGNAD/	CC BY-NC-SA 4.0	https://creativecommons.org/licenses/by-nc-sa/4.0/
	Newspaper	Der Standard - Web	Unknown	Barbie	265	10kGNAD	https://tblock.github.io/10kGNAD/	CC BY-NC-SA 4.0	https://creativecommons.org/licenses/by-nc-sa/4.0/
SUM					**1394**				
	Wikipedia	NA	NA	DDR-Pizza	279	Wikipedia 14.09.2023	https://de.wikipedia.org/wiki/Pizza	CC BY-SA 4.0	https://creativecommons.org/licenses/by-sa/4.0/deed.en
	Wikipedia	NA	NA	Riesenfaultier	288	Wikipedia 14.09.2023	https://de.wikipedia.org/wiki/Riesenfaultier	CC BY-SA 4.0	https://creativecommons.org/licenses/by-sa/4.0/deed.en
	Wikipedia	NA	NA	Warägergarde	145	Wikipedia 14.09.2023	https://de.wikipedia.org/wiki/War%C3%A4gergarde	CC BY-SA 4.0	https://creativecommons.org/licenses/by-sa/4.0/deed.en
	Wikipedia	NA	NA	Voyager Golden Record	286	Wikipedia 14.09.2023	https://de.wikipedia.org/wiki/Voyager_Golden_Record	CC BY-SA 4.0	https://creativecommons.org/licenses/by-sa/4.0/deed.en
	Wikipedia	NA	NA	Mount Erebus	223	Wikipedia 14.09.2023	https://de.wikipedia.org/wiki/Mount_Erebus	CC BY-SA 4.0	https://creativecommons.org/licenses/by-sa/4.0/deed.en
SUM					**1221**				
	Scientific	NA	Anne Herrmann-Werner, Rebecca Erschens, Stephan Zipfel, Teresa Loda	Medizinische Ausbildung	286	GMS Journal for Medical Eductation	https://www.egms.de/static/de/journals/zma/2021-38/zma001489.shtml	CC BY 4.0	https://creativecommons.org/licenses/by/4.0/
	Scientific	NA	Beate Littig	Eliten	140	Forum: Qualitative Social Research	https://www.qualitative-research.net/index.php/fqs/article/view/1000	CC BY 4.0	https://creativecommons.org/licenses/by/4.0/
	Scientific	NA	Gerrit Bayer-Hohenwarte	Zeitdruck	146	The Journal of Specialised Translation	https://jostrans.org/issue09/art_bayer.php	CC BY 2.0	https://creativecommons.org/licenses/by/2.0/
	Scientific	NA	Sandra Jaworeck, Stefan Stemler, Peter Kriwy	Glücksempfinden	210	Prävention und Gesundheitsförderung	https://link.springer.com/article/10.1007/s11553-022-00943-3	CC BY 4.0	https://creativecommons.org/licenses/by/4.0/deed.en
	Scientific	NA	Karsten Kluth, Pascal Jung, Dennis Wurm, Ingo Schmitz, Nicolas Sänger	Jaeger	200	Zeitschrift für Arbeitswissenschaft	https://link.springer.com/article/10.1007/s41449-023-00358-6	CC BY 4.0	https://creativecommons.org/licenses/by/4.0/deed.en
SUM					**982**				
TOTAL					**6542**				

Bibliography

Arehalli, S., Dillon, B., & Linzen, T. (2022). Syntactic Surprisal From Neural Models Predicts, But Underestimates, Human Processing Difficulty From Syntactic Ambiguities. *Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL)*, 301–313. https://aclanthology.org/2022.conll-1.20

Aurnhammer, C., Delogu, F., Brouwer, H., & Crocker, M. (2023). The P600 as a continuous index of integration effort. *Psychophysiology*. https://doi.org/10.1111/psyp.14302

Aurnhammer, C., Delogu, F., Schulz, M., Brouwer, H., & Crocker, M. (2021). Retrieval (N400) and integration (P600) in expectation-based comprehension. *PLOS ONE*, *16*(9), 1–31. https://doi.org/10.1371/journal.pone.0257430

Aurnhammer, C., & Frank, S. (2019a, September). Comparing gated and simple recurrent neural network architectures as models of human sentence processing. In A. Goel, C. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st annual conference of the cognitive science society* (pp. 112–118). Cognitive Science Society. https://scholar.google.de/citations?view_op=view_citation&hl=de&user=SPeiq88AAAAJ&citation_for_view=SPeiq88AAAAJ:9yKSN-GCB0IC

Aurnhammer, C., & Frank, S. (2019b). Evaluating information-theoretic measures of word prediction in naturalistic sentence reading. *Neuropsychologia*, *134*, 107198. https://doi.org/10.1016/j.neuropsychologia.2019.107198

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural Machine Translation by Jointly Learning to Align and Translate. *ArXiv*, *1409*.

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 238–247. https://doi.org/10.3115/v1/P14-1023

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*, 1–46.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent Is Difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166. https://doi.org/10.1109/72.279181

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information.

Botsch, R. (2011). Chapter 12: Significance and measures of association. *Scopes and Methods of Political Science*.

Brouwer, H., Crocker, M., Venhuizen, N., & Hoeks, J. (2017). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive Science*, *41*(S6), 1318–1352. https://doi.org/https://doi.org/10.1111/cogs.12461

Brouwer, H., & Crocker, M. W. (2017). On the Proper Treatment of the N400 and P600 in Language Comprehension. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01327

Brouwer, H., Delogu, F., & Crocker, M. (2021). Splitting Event-Related Potentials: Modeling Latent Components using Regression-based Waveform Estimation. *European Journal of Neuroscience*, *53*, 974–995. https://doi.org/10.1111/ejn.14961

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143. https://doi.org/https://doi.org/10.1016/j.brainres.2012.01.055

Cho, K., van Merrienboer, B., Bahdanau, D., & Bengio, Y. (2014, October). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches [arXiv:1409.1259 [cs, stat]]. Retrieved October 16, 2023, from http://arxiv.org/abs/1409.1259
Comment: Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8).

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014, September). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation [arXiv:1406.1078 [cs, stat]]. Retrieved October 16, 2023, from http://arxiv.org/abs/1406.1078
Comment: EMNLP 2014.

Chwilla, D., & Kolk, H. (2002). Three-step priming in lexical decision. *Memory & cognition*, *30*, 217–25. https://doi.org/10.3758/BF03195282

De Varda, A. G., Marelli, M., & Amenta, S. (2023). Cloze probability, predictability ratings, and computational estimates for 205 English sentences, aligned with existing EEG and reading time data. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02261-8

Degaetano-Ortlieb, S., & Teich, E. (2022). *Corpus Linguistics and Linguistic Theory*, *18*(1), 175–207. https://doi.org/doi:10.1515/cllt-2018-0088

Delogu, F., Brouwer, H., & Crocker, M. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition*, *135*, 103569. https://doi.org/https://doi.org/10.1016/j.bandc.2019.05.007

Delogu, F., Brouwer, H., & Crocker, M. (2021). When components collide: Spatiotemporal overlap of the N400 and P600 in language comprehension. *Brain Research*, *1766*, 147514. https://doi.org/10.1016/j.brainres.2021.147514

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/https://doi.org/10.1016/0364-0213(90)90002-E

Ethayarajh, K. (2019, September). How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings [arXiv:1909.00512 [cs]]. Retrieved October 16, 2023, from http://arxiv.org/abs/1909.00512
Comment: Accepted to EMNLP 2019.

Ettinger, A., Feldman, N. H., Resnik, P., & Phillips, C. (2016). Modeling N400 amplitude using vector space models of word representation. *Cognitive Science*.

Fernandez Monsalve, I., Frank, S. L., & Vigliocco, G. (2012). Lexical surprisal as a general predictor of reading time. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 398–408.

Fossum, V., & Levy, R. (2012). Sequential vs. Hierarchical Syntactic Models of Human Incremental Sentence Processing. *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2012)*, 61–69. https://aclanthology.org/W12-1706

Frank, S., & Hoeks, J. C. J. (2019). The interaction between structure and meaning in sentence comprehension: Recurrent neural networks and reading times. *Annual Meeting of the Cognitive Science Society*.

Frank, S. (2017). Word Embedding Distance Does not Predict Word Reading Time. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*.

Frank, S., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language*, *140*, 1–11. https://doi.org/https://doi.org/10.1016/j.bandl.2014.10.006

Frank, S., & Willems, R. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Language, Cognition and Neuroscience*, *32*(9), 1192–1203. https://doi.org/10.1080/23273798.2017.1323109

Goodkind, A., & Bicknell, K. (2018). Predictive power of word surprisal for reading times is a linear function of language model quality. *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*, 10–18. https://doi.org/10.18653/v1/W18-0102

Grice, H. P. (1967). Logic and Conversation. In P. Grice (Ed.), *Studies in the way of words* (pp. 41–58). Harvard University Press.

Haeuser, K. I., & Kray, J. (2022). How odd: Diverging effects of predictability and plausibility violations on sentence reading and word memory. *Applied Psycholinguistics*, *43*(5), 1193–1220. https://doi.org/10.1017/S0142716422000364

Hale, J. (2001). A Probabilistic Earley Parser as a Psycholinguistic Model. *Proceedings of NAACL 2001*, *2*. https://doi.org/10.3115/1073336.1073357

Han, J., Kamber, M., & Pei, J. (2012). Getting to Know Your Data. In *Data Mining* (pp. 39–82). Elsevier. https://doi.org/10.1016/B978-0-12-381479-1.00002-2

Hochreiter, S. (1991). Untersuchungen zu dynamischen neuronalen netzen.

Hochreiter, S. (1998). The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*, 107–116. https://doi.org/10.1142/S0218488598000094

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. *Neural computation*, *9*, 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python. https://doi.org/10.5281/zenodo.1212303

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813. https://doi.org/10.3758/BF03196544

Jo, J.-y., & Myaeng, S.-H. (2020, July). Roles and Utilization of Attention Heads in Transformer-based Neural Language Models. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 3404–3417). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.311

Jurafsky, D., & Martin, J. H. (n.d.). *Speech and Language Processing* (3rd ed.). Retrieved October 11, 2023, from https://web.stanford.edu/~jurafsky/slp3/

Keppel, G., & Postman, L. (1970). *Norms of Word Association, Edited by Leo Postman and Geoffrey Keppel*. https://books.google.de/books?id=1RUkcgAACAAJ

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, *8*(4), 533–572. https://doi.org/10.1080/01690969308407587

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*, 621–47.

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, *104*, 211–240.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/https://doi.org/10.1016/j.cognition.2007.05.006

Li, J., & Futrell, R. (2023). A decomposition of surprisal tracks the N400 and P600 brain potentials. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*. https://escholarship.org/uc/item/75c569dr

Luong, M.-T., Pham, H., & Manning, C. D. (2015, September). Effective Approaches to Attention-based Neural Machine Translation [arXiv:1508.04025 [cs]]. Retrieved October 16, 2023, from http://arxiv.org/abs/1508.04025
Comment: 11 pages, 7 figures, EMNLP 2015 camera-ready version, more training details.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/https://doi.org/10.1016/j.jml.2016.04.001

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.

Merkx, D., & Frank, S. (2021). Human Sentence Processing: Recurrence or Attention? *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 12–22. https://doi.org/10.18653/v1/2021.cmcl-1.2

Michaelov, J., Bardolph, M., Coulson, S., & Bergen, B. (2021). Different kinds of cognitive plausibility: Why are transformers better than RNNs at predicting N400 amplitude?

Michaelov, J., Bardolph, M., Van Petten, C., Bergen, B., & Coulson, S. (2023). Strong Prediction: Language Model Surprisal Explains Multiple N400 Effects. *Neurobiology of Language*, 1–71. https://doi.org/10.1162/nol_a_00105

Michaelov, J., & Bergen, B. (2020). How well does surprisal explain N400 amplitude under different experimental conditions? *Proceedings of the 24th Conference on Computational Natural Language Learning*, 652–663. https://doi.org/10.18653/v1/2020.conll-1.53

Michaelov, J., Bergen, B., & Coulson, S. (2022). So Cloze yet so far: N400 Amplitude is Better Predicted by Distributional Information than Human Predictability Judgements. *IEEE Transactions on Cognitive and Developmental Systems*, *PP*. https://doi.org/10.1109/TCDS.2022.3176783

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space.

Mitchell, J., Lapata, M., Demberg, V., & Keller, F. (2010). Syntactic and semantic factors in processing difficulty: An integrated measure. *Proceedings of the 48th annual meeting of the association for computational linguistics*, 196–206.

Mosbach, M., Steuer, J., & Brown, A. (2023). Language Model Toolkit [Unpublished].

Nair, S., & Resnik, P. (2023, October). Words, Subwords, and Morphemes: What Really Matters in the Surprisal-Reading Time Relationship? [arXiv:2310.17774 [cs]]. Retrieved November 18, 2023, from http://arxiv.org/abs/2310.17774
Comment: Accepted to Findings of EMNLP 2023; 10 pages, 5 figures.

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Matthew Husband, E., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., … Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791), 20180522. https://doi.org/10.1098/rstb.2018.0522

Nieuwland, M. S., & Van Berkum, J. J. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, *24*(3), 691–701. https://doi.org/https://doi.org/10.1016/j.cogbrainres.2005.04.003

Oh, B.-D., & Schuler, W. (2022). Why Does Surprisal From Larger Transformer-Based Language Models Provide a Poorer Fit to Human Reading Times?

Park, H. H., Zhang, K. J., Haley, C., Steimel, K., Liu, H., & Schwartz, L. (2021). Morphology Matters: A Multilingual Language Modeling Analysis [arXiv:2012.06262 [cs]]. *Transactions of the Association for Computational Linguistics*, *9*, 261–276. https://doi.org/10.1162/tacl_a_00365
Comment: To appear in TACL, a pre-MIT Press publication version; 15 pages, 3 figures; for the datasets, see https://github.com/hayleypark/MorphologyMatters.

Parviz, M., Johnson, M., Johnson, B., & Brock, J. (2012). Using Language Models and Latent Semantic Analysis to Characterise the N400m Neural Response.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Plüster, B. (2023, September). LeoLM: Igniting German-Language LLM Research. https://laion.ai/blog/leo-lm/

Pynte, J., New, B., & Kennedy, A. (2008). On-line contextual influences during reading normal text: A multiple-regression analysis. *Vision research*, *48*, 2172–83. https://doi.org/10.1016/j.visres.2008.02.004

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rabs, E., Delogu, F., Drenhaus, H., & Crocker, M. W. (2022). Situational expectancy or association? The influence of event knowledge on the N400. *Language, Cognition and Neuroscience*, *37*(6), 766–784. https://doi.org/10.1080/23273798.2021.2022171

Radford, A., & Narasimhan, K. (2018). Improving Language Understanding by Generative Pre-Training. https://api.semanticscholar.org/CorpusID:49313245

Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving Lexical and Syntactic Expectation-Based Measures for Psycholinguistic Modeling via Incremental Top-down Parsing. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, 324–333.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation.

Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language*, *162*, 42–45. https://doi.org/10.1016/j.bandl.2016.08.001

Schweter, S. (n.d.-a). Dbmdz German Bert [Accessed: 2023-10-30].

Schweter, S. (n.d.-b). German Secret GPT-2 model [Accessed: 2023-10-27].

Schweter, S. (2020, November). *German GPT-2 model* (Version 1.0.0). Zenodo. https://doi.org/10.5281/zenodo.4275046

Sennrich, R., Haddow, B., & Birch, A. (2016). Neural Machine Translation of Rare Words with Subword Units. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1715–1725. https://doi.org/10.18653/v1/P16-1162

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Slaats, S., & Martin, A. E. (2023, March). What's surprising about surprisal. https://doi.org/10.31234/osf.io/7pvau

Smit, P., Virpioja, S., Grönroos, S.-A., & Kurimo, M. (2014, April). Morfessor 2.0: Toolkit for statistical morphological segmentation. In S. Wintner, M. Tadić, & B. Babych (Eds.), *Proceedings of the demonstrations at the 14th conference of the European chapter of the association for computational linguistics* (pp. 21–24). Association for Computational Linguistics. https://doi.org/10.3115/v1/E14-2006

Smith, N., & Kutas, M. (2015). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*. https://doi.org/10.1111/psyp.12317

Smith, N., & Levy, R. (2008). Optimal processing times in reading: A formal model and empirical investigation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*(30).

Smolka, E., & Eulitz, C. (2018). Psycholinguistic measures for German verb pairs: Semantic transparency, semantic relatedness, verb family size, and age of reading acquisition. *Behavior Research Methods*, *50*, 1540–1562. https://doi.org/10.3758/s13428-018-1052-5

Sutskever, I., Vinyals, O., & Le, Q. V. (2014, December). Sequence to Sequence Learning with Neural Networks [arXiv:1409.3215 [cs]]. Retrieved October 16, 2023, from http://arxiv.org/abs/1409.3215
Comment: 9 pages.

Taylor, W. L. (1953). "Cloze Procedure": A New Tool for Measuring Readability. *Journalism & Mass Communication Quarterly*, *30*, 415–433.

The pandas development team. (2020, February). *Pandas-dev/pandas: Pandas* (Version 2.0.2). Zenodo. https://doi.org/10.5281/zenodo.3509134

Van Petten, C. (2014). Examining the N400 semantic context effect item-by-item: Relationship to corpus-based measures of word co-occurrence. *International Journal of Psychophysiology*, *94*, 407–419. https://doi.org/10.1016/j.ijpsycho.2014.10.012

Van Petten, C., & Luka, B. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, *83 2*, 176–90.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need.

Venhuizen, N., Crocker, M., & Brouwer, H. (2019). Expectation-based Comprehension: Modeling the Interaction of World Knowledge and Linguistic Experience. *Discourse Processes*, *56*, 229–255. https://doi.org/10.1080/0163853X.2018.1448677

Venhuizen, N., Hendriks, P., Crocker, M., & Brouwer, H. (2022). Distributional formal semantics [Special Issue: Selected Papers from WoLLIC 2019, the 26th Workshop on Logic, Language, Information and Computation]. *Information and Computation*, *287*, 104763. https://doi.org/https://doi.org/10.1016/j.ic.2021.104763

Vig, J. (2019). Visualizing Attention in Transformer-Based Language Representation Models [Publisher: arXiv Version Number: 2]. https://doi.org/10.48550/ARXIV.1904.02679

Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, *1*(4), 339–356. https://doi.org/https://doi.org/10.1016/0893-6080(88)90007-X