Is similarity-based interference caused by lossy compression or cue-based retrieval? A computational evaluation

Himanshu Yadav (hyadav@uni-potsdam.de)
Department of Linguistics, University of Potsdam
14476 Potsdam, Germany

Garrett Smith (gasmith@uni-potsdam.de)
Department of Linguistics, University of Potsdam
14476 Potsdam, Germany

Shravan Vasishth (vasishth@uni-potsdam.de)
Department of Linguistics, University of Potsdam
14476 Potsdam, Germany

Abstract

The similarity-based interference paradigm has been widely used to investigate the factors subserving subject-verb agreement processing. A consistent finding is facilitatory interference effects in ungrammatical sentences but inconclusive results in grammatical sentences. Existing models propose that interference is caused either by misrepresentation of the input (representation distortion-based models) or by mis-retrieval of the interfering noun phrase based on cues at the verb (retrieval-based models). These models fail to fully capture the observed interference patterns in the experimental data. We implement two new models under the assumption that a comprehender utilizes a lossy memory representation of the intended message when processing subject-verb agreement dependencies. Our models outperform the existing cue-based retrieval model in capturing the observed patterns in the data for both grammatical and ungrammatical sentences. Lossy compression models under different constraints can be useful in understanding the role of representation distortion in sentence comprehension.

Keywords: Similarity-based interference; lossy memory representation; cue-based retrieval

Introduction

Similarity-based interference in subject-verb agreement dependencies has played an important role in understanding the mechanisms underlying sentence comprehension (Wagers, Lau, & Phillips, 2009; Lago, Shalom, Sigman, Lau, & Phillips, 2015). In this paradigm, a noun phrase matching in agreement features with the verb—called a distractor—is presented along with the subject noun. For example, in the following sentences (a) and (c), the distractor noun phrase *the cabinet(s)* matches the number feature of the verb in contrast to conditions (b) and (d), where it does not.

(a) Grammatical, interference condition
The key to the cabinet unsurprisingly was rusty.
(b) Grammatical, no-interference condition
The key to the cabinets unsurprisingly was rusty.
(c) Ungrammatical, interference condition
* The key to the cabinets unsurprisingly were rusty.
(d) Ungrammatical, no-interference condition
* The key to the cabinet unsurprisingly were rusty.

A consistent finding is that of facilitation in ungrammatical conditions: reading times at the verb 'were' in condition (c) are, on average, faster than in condition (d) (Jäger, Engelmann, & Vasishth, 2017; Wagers et al., 2009; Lago et al., 2015; Dillon, Mishler, Sloggett, & Phillips, 2013; Jäger, Mertzen, Van Dyke, & Vasishth, 2020). By contrast, the results are inconclusive in grammatical conditions: reading times at the verb in condition (a) can be faster, slower, or comparable to condition (b). Figure 1 shows the observed interference effects in the grammatical and ungrammatical conditions from 11 published datasets.

Several models have been proposed to explain the facilitatory interference effect in the ungrammatical conditions, but these models cannot explain the range of effects in the grammatical conditions. Most of these models can be placed into one of two categories, cue-based retrieval accounts, and representation distortion-based accounts.

The cue-based retrieval account (Lewis & Vasishth, 2005) assumes that dependency completion between the subject and the verb is driven by a cue-based retrieval process: encountering a verb triggers a content-addressable search in memory using feature specifications such as [+subject] or [+plural], called retrieval cues. The cue-based retrieval model correctly predicts the facilitatory effect in ungrammatical conditions. But the model predicts an inhibitory effect in grammatical conditions: a slowdown in condition (a) compared to (b). This prediction is not supported by the interference effect data in the grammatical conditions shown in Fig. 1.

Representation distortion-based accounts assume that the representation of the pre-verbal sentence material—subject noun and/or distractor noun—undergoes distortion with time. One of the representation distortion-based accounts—the encoding-based model (Bock & Eberhard, 1993; Eberhard, 1997)—maintains that the plural feature of the distractor noun percolates up to the subject noun phrase causing a *misrepresentation* of the subject in a proportion of trials. The encoding-based model predicts facilitatory effect in both grammatical and ungrammatical conditions, which is not supported by the observed pattern of effects (see Fig. 1).

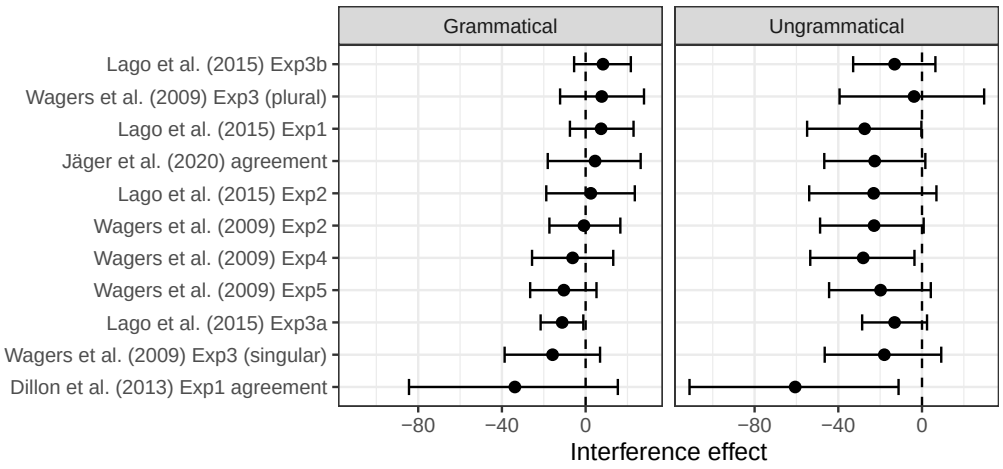

Figure 1: The pattern of interference effects in subject-verb agreement dependencies. Here, "interference effect" means the difference in reading times at the verb between the interference and no-interference conditions.

Another class of representation distortion-based models are based on lossy compression of the linguistic input (Futrell, Gibson, & Levy, 2020). These models assume that a comprehender obtains a distorted representation of the true intended message due to lossy memory encoding, and they reconstruct a set of possible true representations using their prior linguistic knowledge. A well-tested model of this type is the lossy-context suprisal model of Futrell et al. (2020). The model captures working memory effects within an expectation-based framework. It assumes that after reading or hearing a series of words, the words can be corrupted by deleting content words (e.g., nouns or verbs) at a constant rate. Processing difficulty at a new word is the expected surprisal of the word given this lossy memory representation of its preceding context. Futrell et al. (2020) show that the model explains structural forgetting effects (Vasishth, Suckow, Lewis, & Kern, 2010), and Futrell (2019) explain information locality across languages using the lossy compression model.

An important limitation of the literature on similarity-based interference effects is that researchers either invoke cue-based retrieval or some kind of lossy compression model to explain the data. Moreover, the two classes of model have never been pitted against each other in any systematic quantitative evaluation, even though a considerable amount of benchmark data are available on interference effects. A further intriguing possibility, which needs to be quantitatively evaluated, is that both lossy compression and cue-based retrieval could play a role in a hybrid model.

We address these open issues by implementing two lossy compression models of similarity-based interference to try to capture the observed effects in both grammatical and ungrammatical conditions in subject-verb agreement dependencies. We compare the performance of our models against the cue-based retrieval model of Lewis and Vasishth (2005); Vasishth, Nicenboim, Engelmann, and Burchert (2019) on interference effect data from the 11 publicly available datasets shown in figure 1.

A lossy compression model of interference effects

We implement a lossy-context surprisal model as described in Futrell et al. (2020) with some additional assumptions to model interference effects in subject-verb agreement dependencies.

Assumptions

Consider the sentence "The key to the cabinets unsurprisingly was rusty". The observed pre-verbal noun phrase in this sentence is *the key to the cabinets*. We call this input I. The lossy compression model assumes the following:

1. The linguistic input received by the comprehender has undergone lossy compression: there was some true representation r; due to lossy memory encoding, the true representation r distorts to the observed input I such that the plural marker on the nouns can either be deleted or inserted or left unchanged at constant rates
2. The comprehender reconstructs a set of possible true representations from input I conditioned on their prior linguistic knowledge and the rates of deletion/insertion in the system
3. The processing difficulty at the verb is the expected (average) surprisal of encountering the verb given all possible true representations of the input I

Next, we derive the processing difficulty and reading times at the verb in subject-verb agreement dependencies.

Calculating processing difficulty and reading times at the verb

In the sentence "The key to the cabinets unsurprisingly was rusty", the input is

$I = N \quad \text{P} \quad N.pl$

where N represents a noun, P represents a preposition, and $.pl$ represents a plural marker on a noun.

The possible true representations, r_i, that can lead to input I due to lossy compression are,

$r_1 = N.pl \quad \text{P} \quad N.pl \qquad r_2 = N.pl \quad \text{P} \quad N$

$r_3 = N \quad \text{P} \quad N.pl \qquad r_4 = N \quad \text{P} \quad N$

The processing difficulty for the upcoming verb will be proportional to the *expected suprisal* of the verb given all possible true representations $r_1, r_2, .., r_N$:

$$D(V|I) \propto \sum_{i=1}^{N} -\log P(V|r_i) \cdot P(r_i|I) \tag{1}$$

where $-\log P(V|r_i)$ is the surprisal — negative log conditional probability — of seeing a plural/singular verb after the context r_i; we compute conditional probabilities from the COW corpora (Schäfer, 2015; Schäfer & Bildhauer, 2012). And, $P(r_i|I)$ is the probability density of reconstructing a representation r_i from the given input, I. We can derive $P(r_i|I)$ using Bayes' rule,

$$P(r_i|I) \propto P(I|r_i)P(r_i) \tag{2}$$

where $P(r_i)$ is the prior probability density of a possible true representation r_i and can be estimated from corpus data. $P(I|r_i)$ represents the lossy memory encoding function: the likelihood that a representation r_i distorts to I given a constant deletion rate d and constant insertion rate a:

$$I|r_i \sim Memory(r_i, d, a) \tag{3}$$

where d is the rate of deleting a plural marker and a is the rate of inserting a plural marker. Table 1 shows the likelihood of obtaining I from each possible representation r_i. Finally, we transform processing difficulty into reading times using a linear linking function. Reading times in j^{th} trial, RT_j, will be:

$$RT_j = S \cdot D(V|I) + \varepsilon_j \tag{4}$$

where S is a scaling parameter and ε_j is the random noise in the j^{th} trial such that $\varepsilon_j \sim Normal(0, 20)$. The model has thus 3 free parameters: deletion rate d, insertion rate a and scaling parameter S.

Possible true representation r_i	**Likelihood of generating** I **from** r_i $P(I\|r_i)$
$N.pl \quad \text{P} \quad N.pl$	$d(1-d)$
$N.pl \quad \text{P} \quad N$	da
$N \quad \text{P} \quad N.pl$	$(1-a)(1-d)$
$N \quad \text{P} \quad N$	$(1-a)a$

Table 1: The lossy memory encoding function: the likelihood of obtaining the observed input I (N P $N.pl$) from lossy compression of a possible true representation r_i

Prior predictions

We use the model equations stated in the previous section and generate prior predictions from the model. This allows us to determine the range of effects the model can generate and compare them against the observed interference effect data. The joint distribution of interference effects in grammatical and ungrammatical conditions — $\{E_{gram}, E_{ungram}\}$ — is assumed to come from the lossy compression model conditional on its free parameters, the deletion rate d, the insertion rate a, and the scaling parameter S

$$\{E_{gram}, E_{ungram}\} \sim Model(d, a, S) \tag{5}$$

We specified the priors as follows. For deletion rate d and insertion rate a, we choose a weakly informative prior because we do not want to make any strong assumptions about these parameters:

$$d \sim Normal_{lb=0,ub=1}(0, 0.25) \tag{6}$$

$$a \sim Normal_{lb=0,ub=1}(0, 0.25) \tag{7}$$

where $lb = 0$ and $ub = 1$ indicate a lower bound of 0 and upper bound of 1 respectively. For the scaling parameter S, we choose a Gaussian prior centered at 25 and with standard deviation of 5; this range was chosen so that the model does not generate unreasonably large or small reading times (see Jäger et al., 2017, for meta-analysis estimates of reading times):

$$S \sim Normal_{lb=0}(25, 5)$$

Figure 2 shows the prediction space of the lossy compression model against the observed interference effect data. The model is able to predict a facilitatory effect in ungrammatical conditions and positive, zero or negative effects in grammatical conditions. Thus, the prior predictions of the model are consistent with qualitative pattern of the observed interference effects, but the magnitudes of predicted effects do not often align with the human data.

The lossy compression model, presented here, assumes that the link between the lossy memory representations and reading time effects is the average surprisal of the upcoming word. However, one is free to choose a different linking function. In the next section, we introduce a hybrid model that integrates the lossy compression and cue-based retrieval mechanisms in order to predict reading times at the verb.

Lossy-compression-plus-retrieval model

Recent work has shown that a model unifying representation distortion- and retrieval-based mechanisms shows a better fit to interference effect data from subject-verb agreement dependencies (Yadav, Smith, & Vasishth, 2021). Given this modeling evidence, it is interesting to explore a model combining lossy compression and cue-based retrieval in a single set of processes. We implement the lossy compression-plus-retrieval model with the idea that the cue-based retrieval at the verb is preceded by lossy memory representation of the intended message. Here, reading times are determined by the cue-based retrieval mechanisms, but the retrieval process operates on a noisy version of the intended input.

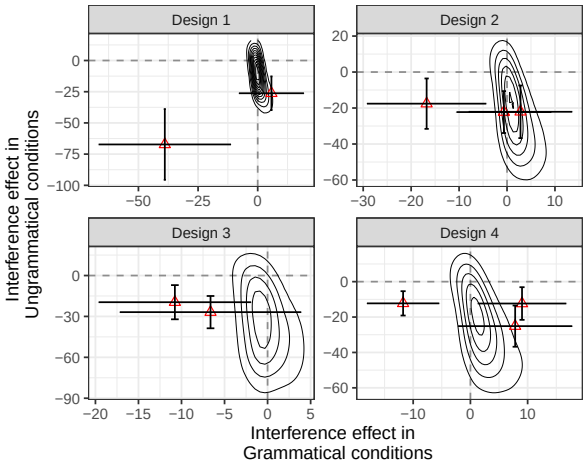

Figure 2: The prior predictive interference effect (in milliseconds) generated by the lossy compression model is shown as a contour of the joint distribution of effects in the grammatical and ungrammatical conditions. The red triangular points and errors bars around them represent the observed interference effects and their 95% credible intervals obtained from published datasets. The predictions differ across experimental designs because prior density of possible true representations $p(r_i)$ is estimated to be different for each design. Design 1: English subject relative clause constructions; Dillon et al. (2013), Exp 1 and Jäger et al. (2020). Design 2: English object relative clause constructions; Wagers et al. (2009) Exp 2, Exp 3, and Lago et al. (2015) Exp 2. Design 3: English prepositional phrase constructions; Wagers et al. (2009) Exp 4, Exp 5. Design 4: Spanish relative clause constructions; Lago et al. (2015) Exp1, Exp 3a, and Exp 3b.

Assumptions

The lossy compression-plus-retrieval model assumes that

1. Dependency completion between the subject and the verb is driven by a cue-based retrieval process
2. Cue-based retrieval is affected by changes in representaton of the subject and the distractor nouns due to lossy compression of the intended message (as described in the previous section)

Next, we derive the updated retrieval time equation to account for representation change due to lossy compression.

Calculating retrieval times

The retrieval time at the verb in the j^{th} trial, RT_j, is an exponential function of the activation of the retrieved chunk,

$$RT_j = Fe^{-A_{j,retrieved}} \tag{8}$$

where F is a scaling parameter called the latency factor which reflects overall processing speed.

Under cue-based retrieval, the chunk with the highest activation gets retrieved in each trial. The activation of the chunk retrieved in j^{th} trial would be the maximum of the activation of the subject and the distractor noun:

$$A_{j,retrieved} = \max\{A_{j,subject}, A_{j,distractor}\} \tag{9}$$

The activation of the subject and the distractor in a trial is determined by the amount of activation they receive via cue-feature match. The noun phrase that matches more cues receives a higher activation. Thus, activation of the subject and the distractor in j^{th} trial is a function of their representation,

$$\{A_{j,subject}, A_{j,distractor}\} \sim Activation(r_j) \tag{10}$$

where r_j is the representation of subject and distractor noun in the j^{th} trial. The lossy compression-plus-retrieval model assumes that the representation in the j^{th} trial is sampled from probability density of reconstructing r from input I,

$$r_j \sim P(r|I,a,d) \tag{11}$$

where a and d are the insertion and deletion rates, respectively. The probability density function $P(r|I,a,d)$ can be derived in the same way as in equation 2. Using these equations, the lossy compression-plus-retrieval model allows us to make reading time predictions at the verb, which we now compare to reading time data from 11 experiments.

Prior predictions

We generate prior predictions from the lossy compression-plus-retrieval model conditional on its three free parameters, the deletion rate d, the insertion rate a, and the latency factor F. For the deletion rate and the insertion rate, we specify the same priors as in equation 6 and 7. For the latency factor, we used a truncated normal distribution:

$$F \sim Normal_{lb=0.1}(0.15, 0.03)$$

where lb = 0.1 indicates a lower bound of 0.1 on latency factor values. We choose this lower bound because a latency factor of less than 0.1 generates unreasonably fast reading times.

Figure 3 shows the prediction space of the lossy compression-plus-retrieval model against observed interference effect data. The model predictions are consistent with the facilitatory effect in ungrammatical conditions, but inconsistent with the range of effects in grammatical conditions.

Model comparison

We compare the performance of the lossy compression models (which assume that representation undergoes distortion due to information loss) against the cue-based retrieval model (which assumes that representation is intact and a retrieval process drives processing) on 11 published datasets. We use stratified k-fold cross-validation for model comparison: (1) We split each dataset into 6 folds (subsets) such that each fold contained observations from all participants for all conditions, (2) we prepared 6 sets of training and test data by leaving out one fold as test data and taking other 5 as training

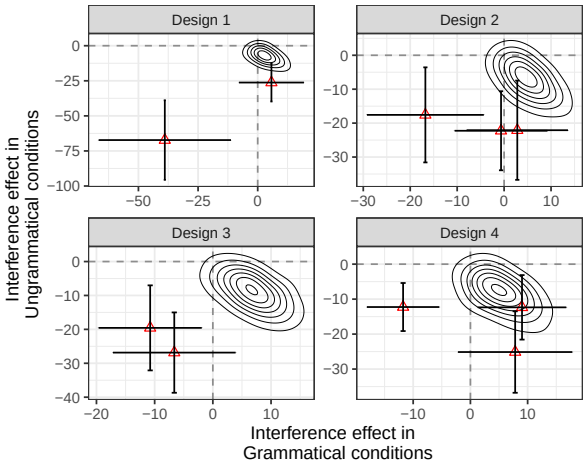

Figure 3: The prior predictive interference effect (in milliseconds) generated by the lossy compression-plus-retrieval model is shown as a contour of joint distribution of effects in the grammatical and ungrammatical conditions. The red triangular points and errors bars around them represent observed interference effects.

data, (3) in each iteration, we fit the models on training data using Approximate Bayesian Computation[1] (Sisson, Fan, & Beaumont, 2018) and computed the predictive accuracy of the fitted model on the test data in terms of log pointwise predictive density. Figure 4 shows the comparison of estimated log pointwise predictive density ($\widehat{elpd}$) of the models on 11 datasets. We find that:

1. The $\widehat{elpd}$ values for the lossy compression model are larger than the cue-based retrieval model for 6 out of 11 datasets suggesting stronger evidence in the favor of lossy-compression model. The models are indistinguishable for the remaining five datasets.

2. The lossy-compression-plus-retrieval model shows higher predictive accuracy than the cue-based retrieval model for six out of 11 datasets.

3. The lossy-compression-plus-retrieval model and the lossy compression model show comparable performance.

Overall, the results suggest that a lossy compression model or a lossy compression-plus-retrieval model can explain the data better than the standard cue-based retrieval model.

Discussion

We have implemented two models—a lossy compression model and a lossy compression-plus-retrieval model—and investigated whether they can outperform the cue-based retrieval model. More specifically, we investigated whether, compared to the cue-based retrieval model, these two models can furnish a better account for the pattern of interference effects in grammatical and ungrammatical subject-verb agreement dependencies. Both models are based on the idea of lossy memory representations of the intended message. The lossy compression model assumes that the linguistic input received by a comprehender is subject to information loss, and that the comprehender infers a set of possible true representations from the given input using their prior linguistic knowledge. Reading times are then predicted to be proportional to the expected surprisal of the next word given the set of possible true representations. By contrast, the hybrid lossy compression-plus-retrieval model assumes that dependency completion is driven by a cue-based retrieval process which is affected by a change in the representation of memory chunks due to lossy compression. Reading time predictions here are derived from the assumptions of cue-based retrieval (Lewis & Vasishth, 2005).

The evaluation of the three models' predictive performance shows that both lossy compression and lossy compression-plus-retrieval models are better at explaining the interference effect data than the cue-based retrieval model of Lewis and Vasishth (2005). An important implication of the modeling results is that the cognitive processes underlying dependency completion in sentence comprehension might involve representation distortion due to lossy compression of the intended message.

An interesting open question is whether the deletion and insertion rates assumed in the lossy memory encoding function are sensitive to factors like the syntactic position of the nouns and the distance between the nouns and the verb. For example, a noun that appears earlier in the sentence may enjoy a primacy advantage (Häussler & Bader, 2015), and therefore be less likely to be distorted by deletion/insertion noise. Similarly, there could be a subject advantage in memory such that the representation of subject nouns is distorted at slower rates than other noun phrases (Futrell et al., 2020). Another reasonable assumption can be that the memory representation of nouns is susceptible to only deletion noise and not insertion noise. Our model is currently agnostic to these factors. But they can be explored by developing constraints on deletion and insertion rate for different noun phrases in our model. We plan to take this up in future work.

In sum, the modeling presented here demonstrates, for the first time, that a prominent and well-accepted explanation for interference effects—cue-based retrieval—is outperformed by models that assume lossy compression. The fact that the two lossy compression models (the one with and without cue-based retrieval) show comparable fits raises an interesting question: is the cue-based retrieval assumption needed at all to explain interference effects? This is an important open question that should be addressed in future work.

[1]Approximate Bayesian Computation (ABC) allows us to fit complex models when the likelihood of a model cannot be expressed mathematically. We use a particle filtering-based ABC algorithm to estimate posterior distributions of free parameters in the models.

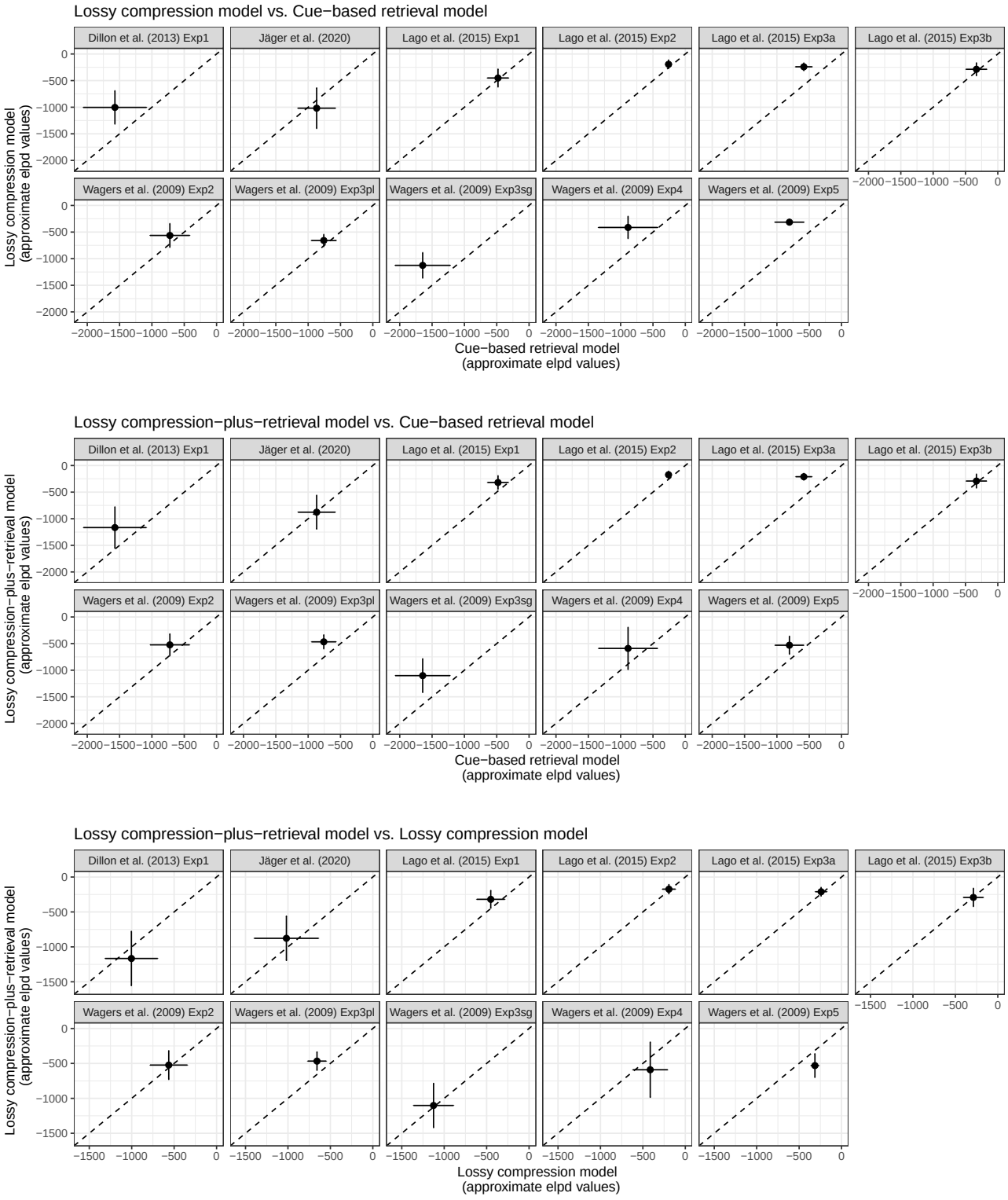

Figure 4: Estimated log pointwise predictive density for each model for each dataset based on stratified k-fold cross validation.

Acknowledgments

We thank the anonymous reviewers for helpful suggestions.

References

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, *8*(1), 57–99.

Dillon, B. W., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*, 85–103.

Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language*, *36*, 147–164.

Futrell, R. (2019). Information-theoretic locality properties of natural language. In *Proceedings of the first workshop on quantitative syntax (quasy, syntaxfest 2019)* (pp. 2–15).

Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-context surprisal: An information-theoretic model of memory effects in sentence processing. *Cognitive science*, *44*(3), e12814.

Häussler, J., & Bader, M. (2015). An interference account of the missing-vp effect. *Frontiers in Psychology*, *6*, 766.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2017). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316-339.

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language*, *111*.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement processes in Spanish comprehension. *Journal of Memory and Language*, *82*, 133–149.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

Schäfer, R. (2015). Processing and querying large web corpora with the cow14 architecture.

Schäfer, R., & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. In *Lrec* (pp. 486–493).

Sisson, S. A., Fan, Y., & Beaumont, M. (2018). *Handbook of approximate bayesian computation*. CRC Press.

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences*, *23*, 968–982. doi: https://doi.org/10.1016/j.tics.2019.09.003

Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, *25*(4), 533–567.

Wagers, M., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*, 206–237.

Yadav, H., Smith, G., & Vasishth, S. (2021). *Feature encoding modulates cue-based retrieval: Modeling interference effects in both grammatical and ungrammatical sentences.* PsyArXiv. Retrieved from `https://psyarxiv.com/76aex`