GLOSSA psycholinguistics

Chromý, J., Brand, J., Laurinavichyute, A., & Lacina, R. (2023). Number agreement attraction in Czech and English comprehension: A direct experimental comparison. *Glossa Psycholinguistics*, 2(1): 14, pp. 1–20. DOI: https://doi.org/10.5070/G6011235

eScholarship University of California

Number agreement attraction in Czech and English comprehension: A direct experimental comparison

Jan Chromý, Charles University, Faculty of Arts, Prague, Czechia, jan.chromy@ff.cuni.cz

James Brand, Charles University, Faculty of Arts, Prague, Czechia, james.brand.ac@gmail.com

Anna Laurinavichyute, University of Potsdam, Department of Linguistics, Potsdam, Germany, anna.laurinavichyute@uni-potsdam.de

Radim Lacina, Osnabrück University, Institute of Cognitive Science, Osnabrück, Germany & University of Potsdam, Department of Linguistics, Potsdam, Germany, radim.lacina@uni-osnabrueck.de

Number agreement attraction in comprehension is a phenomenon that has been documented in various typologically diverse languages. This evidence has led to claims about the cross-linguistic uniformity of agreement attraction effects and its independence from the grammatical features of a particular language. However, recent research has shown that in Czech, number agreement attraction effects are either absent or negligible in size. This directly contradicts the cross-linguistic uniformity hypothesis. The current paper aims to further corroborate this finding and presents a direct experimental comparison of Czech and English. Two comparable self-paced reading experiments were conducted using stimuli that were translation equivalents in Czech and English. Our analyses demonstrate a preference for the null model in Czech (no agreement attraction), unlike in English, where an interaction between verb number and attractor number was preferred. Moreover, when we compare the data from the two experiments directly, we find that the interaction between language and attraction was also preferred. In sum, we provide evidence against the cross-linguistic uniformity hypothesis for agreement attraction effects and suggest that processing patterns may differ between languages even for almost identical structures, such as agreement relations.

1. Introduction

Speakers and comprehenders sometimes make errors in their parsing or dependency formation decisions. One of the most studied phenomena is *agreement attraction*. In the production of English, the following type of speech error has been well documented (Bock & Miller, 1991):

(1) *The time for fun and games are over.

The sentence in (1) is ungrammatical due to the mismatch in number between the subject head (*time*) and the verbal auxiliary that requires number agreement with the subject head. While the subject is singular, the verb (*are*) carries plural marking. Crucially, this type of agreement error is more common when the noun modifying the subject head (inside the PP) is also plural. In other words, when it is *games,* rather than singular *fun*, speakers are much more likely to produce an ungrammatical sentence such as (1) (Bock & Miller, 1991). This tendency for speakers to produce such errors has been termed *agreement attraction*.

Interestingly, it has been documented that agreement attraction is present not only in production, but also in comprehension. For example, Wagers et al. (2009) have shown differences in reaction times (RTs) between sentences such as (2) and (3), using the self-paced reading paradigm. Both sentences are ungrammatical, but they differ in the number of the noun phrase standing between the subject (*the key*) and the verb (*were*). When this noun phrase, known as the attractor, was plural, Wagers et al. (2009) reported a statistically significant speed-up in the processing of (3), in comparison to (2).

(2) The key to the cell unsurprisingly were rusty from many years of disuse.

(3) The key to the cells unsurprisingly were rusty from many years of disuse.

Such findings have been replicated in English (e.g. Tanner et al., 2014; Cunnings & Sturt, 2018; Parker & An, 2018; Laurinavichyute & von der Malsburg, 2022), and similar effects have also been found for other languages, such as French (Franck et al., 2015; Franck & Wagers, 2020), Spanish (Acuña-Fariña et al., 2014; Lago et al., 2015), German (Lago & Felser, 2018), Arabic (Tucker et al., 2015, 2021), Russian (Slioussar, 2018), Armenian (Avetisyan et al., 2020), and Greek (Paspali & Marinis, 2020). The evidence for agreement attraction effects has also been supported by a meta-analysis of 16 previous experiments (Jäger et al., 2017). From all of these studies, we observe that agreement attraction effects typically arise on the verb + 1 region (i.e. *rusty* in example (3)) and can spill over to the following region(s). There are two general types of explanation for these effects: representational (e.g. Hammerly et al., 2019) and memory access-oriented approaches (Engelmann et al., 2019). Crucially, for the purposes of this study, both of these approaches predict agreement attraction effects to arise in sentences such as (3).

Agreement attraction effects in comprehension have been attested not only in reaction time studies, but also in experiments employing acceptability judgments (e.g. Wagers et al., 2009), which have reported that ungrammatical sentences – with an attractor matching the verb – have a tendency to be judged as grammatical more often than ungrammatical sentences with a mismatching attractor. This effect is referred to as the *grammaticality illusion* (Phillips et al., 2011).

The overwhelming evidence for agreement attraction effects in various languages has led to some claims that agreement attraction is a cross-linguistically uniform phenomenon, which is independent of the structural features of languages, e.g. case marking, word order etc. (Lago et al., 2015; Avetisyan et al., 2020; Paspali & Marinis, 2020). An influential theory attempting to explain these effects, namely cue-based parsing models (e.g. Engelmann et al., 2019), sees their root in general memory mechanisms that also ought to be independent of language-specific features.

However, recent research on Czech has provided evidence against the idea of cross-linguistic uniformity of agreement attraction effects. Lacina and Chromý (2022) conducted two experiments using Czech sentences, which failed to replicate the effects reported for Russian (Slioussar, 2018), a language structurally similar to Czech (i.e. a Slavic language with a relatively free word order, and complex case marking). Moreover, Chromý et al. (submitted) conducted four experiments on various sentence types in Czech and found either negligible or no attraction effects. The study by Chromý et al. (submitted) also presents a corpus analysis of written and spoken Czech, showing that agreement attraction errors are practically non-existent in Czech production. Finally, Lacina (2023) ran an untimed acceptability judgment pilot study, where attraction sentences were presented to Czech speakers as regular text. He found a substantial ungrammaticality effect, but no evidence of agreement attraction. The study does, however, show that Czech comprehenders are sensitive to agreement violations in sentences similar in structure to those used in the above-mentioned reading studies both on Czech and other languages. Taken together, there is an emerging picture that number agreement attraction effects are not uniform and that in Czech, agreement attraction effects are at least substantially weaker than the effects attested in studies in other languages.

Despite this, it is not possible to definitely conclude whether the cross-linguistic comparison is entirely valid, as a number of outstanding issues remain. Firstly, studies on agreement attraction employ different items (considering their length, semantics, predictability, plausibility, etc.). We cannot exclude the possibility that the results for Czech were due to the specific characteristics of the chosen experimental sentences (for example, all Czech reading time experiments used the future tense auxiliary, whereas the experiments in other languages typically employ the past tense). Secondly, differences exist in the number of stimuli presented – not only experimental items (cf. Laurinavichyute & von der Malsburg, 2022), but also the number of fillers. Possibly, task adaptation effects (Fine et al., 2013; Arehalli & Wittenberg, 2021; Laurinavichyute

& von der Malsburg, 2023) may have influenced results of different studies to a different extent. Moreover, it is unlikely that more subtle differences in the methodologies of different experiments would have been replicated directly, for example, the wording of instructions, font sizes or the experimental software/platform. Such aspects may have somehow confounded the results. In other words, we cannot exclude the possibility that the findings on Czech differed from findings on other languages due to certain methodological and technical differences in running the experiments and analysing their results.

2. Current study

In the present study, we compare Czech – a language where number agreement attraction effects were previously not clearly documented, with English – a language where number agreement attraction has been documented reliably. This allows us to directly test the hypothesis of cross-linguistic uniformity of agreement attraction effects. Crucially, this approach allows us to control for various factors in the experimental design, so that a reliable comparison between the two languages can be made.

To this end, we conducted two parallel web-based self-paced reading experiments in Czech and English, which incorporated the following key considerations:

- The stimuli (both experimental items and fillers) for the Czech and English experiments were carefully designed to be translation equivalents.
- The experiments were both programmed in PC Ibex Farm (Zehr & Schwarz, 2018) using the same script (differing only in the set of items used). All settings were thus equal.
- The instructions given to the participants were as similar as possible – they were first created in Czech and then translated into English.
- Raw data were processed identically and the analysis was conducted following an identical workflow.

The two experiments targeted sentences in which the attractor noun stood between a subject and a verb, such as in (2). There were several reasons for this. First, such sentences were previously examined in various languages and were typically found to yield agreement attraction effects in comprehension. Second, it is relatively easy to create such sentences in both Czech and English, so the sentences can align relatively well (i.e. the critical part of the sentence has the same number of words in the identical order), thus making them the ideal candidates for a direct comparative study. The only problem was the presence of articles in English (due to Czech not having this word category). This ultimately led to longer sentences in English (counted by the number of words used), but crucially, the regions of interest (meaning the verb, verb + 1 and possibly also the attractor and attractor + 1 regions) were aligned exactly in the two languages.

3. Experiment 1: Czech

3.1 Method

3.1.1 Participants

We collected data from 194 native Czech speakers. The participants were recruited from a university-wide participant pool of undergraduate students from Charles University, who received course credit for their participation. Three participants were excluded due to their accuracy on filler item probes being lower than 75%. The resulting sample consisted of 191 participants (159 female, 31 male, 1 did not disclose; median age of 21 years). After the exclusion of the three participants, the average response accuracy for filler items was 94.14%, with the median being 95.01%.

3.1.2 Materials

We used 24 experimental sentence sets. Each sentence consisted of four conditions, based on the number of the attractor noun (singular or plural) and the number of the verb (singular or plural). Thus, the experiment employed a 2×2 within-subject design.

Each experimental sentence consisted of eight words:

1. the subject in nominative singular
2. a preposition *pro* ('for')
3. an attractor noun in the accusative case (singular or plural)
4. an adverb
5. the past-tense auxiliary (singular or plural)
6. another adverb
7. a past participle form of a verb
8. an agent noun

An example of an experimental sentence is shown in **Table 1**. Importantly, the plural forms of the attractor noun were syncretic, i.e. the accusative plural form was homonymous with the nominative plural form (e.g. the ending *-y* in the attractor *archivářky*, 'female archivers' has the same form for both the nominative plural and the accusative plural). The singular form was non-syncretic (i.e. uniquely accusative singular). The main reason for using sentences with syncretic plural attractors were the findings on Russian (Slioussar, 2018), a language structurally close to Czech. Slioussar showed that case syncretism was a key factor influencing the presence of agreement attraction effects in Russian. Thus, we assumed that the constructions with the highest probability of yielding agreement attraction sentences in Czech need to contain syncretic attractors.

Table 1: Example of experimental item from Experiment 1, together with glosses according to the Leipzig glossing rules (Comrie et al., 2008).

	1	2	3	4	5	6	7	8
a	Složk-a	pro	archivářk-u	nejspíš	byl-a	pevně	zapečetěn-a	úřad-y.
	file-NOM.SG	for	archiver-ACC.SG	probably	was	tightly	sealed-SG	authority-INS.PL
b	Složk-a	pro	archivářk-y	nejspíš	byl-a	pevně	zapečetěn-a	úřad-y.
	file-NOM.SG	for	archiver-ACC.PL = NOM.PL	probably	was	tightly	sealed-SG	authority-INS.PL
c	Složk-a	pro	archivářk-u	nejspíš	byl-y	pevně	zapečetěn-y	úřad-y.
	file-NOM.SG	for	archiver-ACC.SG	probably	were	tightly	sealed-PL	authority-INS.PL
d	Složk-a	pro	archivářk-y	nejspíš	byl-y	pevně	zapečetěn-y	úřad-y.
	file-NOM.SG	for	archiver-ACC.PL = NOM.PL	probably	were	tightly	sealed-PL	authority-INS.PL

'A file for an archiver/archivers probably was/were tightly sealed by the authorities.'

Each experimental sentence was followed by a yes–no comprehension question. These questions always targeted the relation between the sentence agent (region 8), verb (region 7) and object (region 1). For example, the sentence shown in **Table 1** had the question: *Zapečetili složku úřady?* ('Did the authorities seal the file?'). Half of the items had "yes" as the correct answer, the other half had "no" (in such cases, a different subject was used in the question than in the sentence, but the verb and object were identical).

Three sentences were used as practice items at the beginning of the experiment. As fillers, we used another 96 sentences. Thus, the ratio of experimental to filler items was 1:4. Altogether, participants read 123 sentences. Each practice item and filler sentence was followed by a yes–no comprehension question, but these were not created according to a specific pattern and targeted various aspects of the sentences.

3.1.3 Procedure

The web-based experiment employed the moving window word-by-word self-paced reading paradigm, and was conducted on the PC Ibex Farm platform (Zehr & Schwarz, 2018).

Participants were first presented with the information about the experiment, including a consent form, and were asked to fill out a brief demographic questionnaire. Next, they were presented with three practice sentences to acquaint them with the procedure.

In the experiment, participants were asked to use the space bar to move from word to word. After each sentence, they had to answer a yes–no comprehension question targeting the sentence content by mouse clicking. No time-out was set for answering the question, and participants did not receive any feedback about the accuracy of their answer.

Each participant saw only one of the conditions of each item, and the conditions were distributed based on the Latin-square design. Altogether, each participant was exposed to six exemplars of each condition. The order of sentences was fully randomized for each participant. On average, the experiment took about 20 minutes to complete.

3.1.4 Data analysis

Before the analysis, RT outliers were removed. The minimum cut-off point was 100 ms, the maximum cutoff point was decided based on the visual inspection of the spread of RTs (only clearly discontinuous data points were excluded, cf. Baayen & Milin, 2010). All RTs higher than 7,000 ms were removed (around 0.143% of the whole data set).

Analyses were conducted using the R programming language (R Core Team, 2022). Data were analyzed using generalized linear mixed models, fit in the Bayesian framework (Vasishth et al., 2018) using the R package *brms* (Bürkner, 2017). The results are reported in terms of the posterior mean, the 95% Credible intervals (95%-CrI), the $P(\beta > 0)$, and Bayes factors (BF) for the critical effects. The details of the Bayesian model fitting, including the specification of regularizing priors, can be found in Appendix A.

Every model assumed log-normal distribution of the dependent variable and included the main effects of attractor and verb number and their interaction: singular (grammatical) verb form was coded as 1, plural as –1; singular attractor was coded as 1, plural attractor as –1. The models had a full random effect structure: by-subject and by-item random intercepts and slopes for the main effects and their interaction.

3.2 Results

3.2.1 Descriptive statistics

Figure 1 shows the log-transformed RTs for each condition and sentence region. Raw RTs for each condition are presented in **Table 4** in Appendix B.

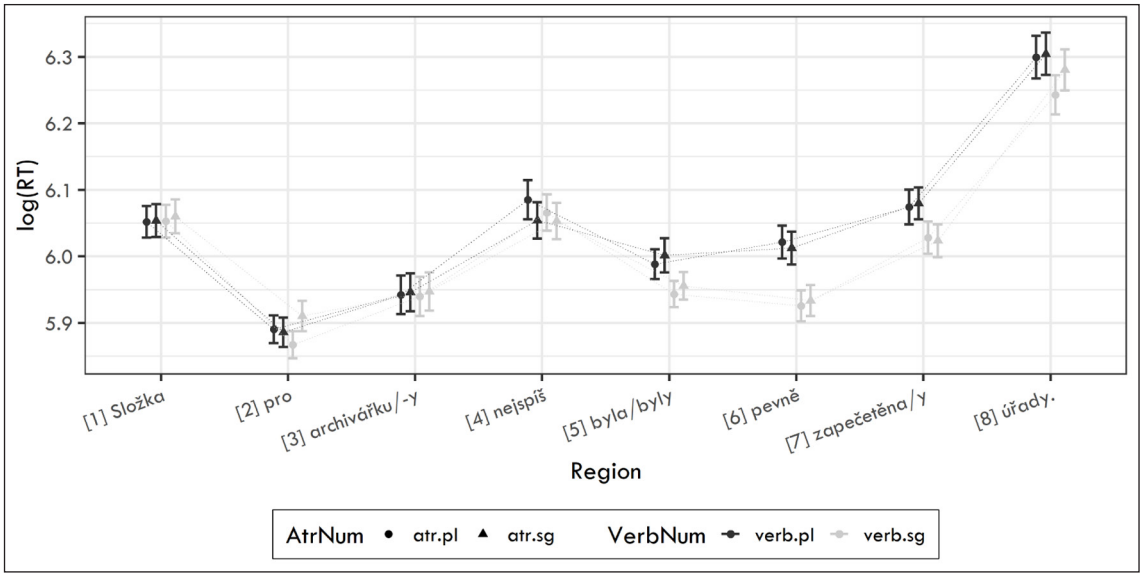

Figure 1: Mean log RTs together with 95% confidence intervals for the four conditions used in Experiment 1.

3.2.2 Bayesian modeling

In the precritical region, there was no indication of any main effect. In the critical and postcritical region, there was a main effect of grammaticality: a speedup in grammatical conditions as compared to ungrammatical conditions with plural verb (critical: $\hat{\beta} = -18$ ms, 95%-CrI: [–26, –9] ms, $P(\beta < 0) > 0.99$; postcritical: $\hat{\beta} = -34$ ms, 95%-CrI: [–42, –26] ms, $P(\beta < 0) > 0.99$). There was no indication for an attraction effect (i.e., an interaction between attractor number and verb number) in the critical or postcritical region (critical: $\hat{\beta} = -0.15$ ms, 95%-CrI: [–7.1, 6.7] ms, $P(\beta < 0) = 0.52$; postcritical: $\hat{\beta} = 3.2$ ms, 95%-CrI: [–4.6, 11] ms, $P(\beta > 0) = 0.8$). In **Figure 2**, the Bayes factor analysis is summarized. Bayes factors show how much more likely the model with the attraction effect is than a model that does not include the attraction effect, given the data and the specific priors. **Figure 2** shows that for the whole range of priors, the model without the attraction effect is strongly preferred.[1]

3.3 Discussion

In sum, we did not find any evidence for agreement attraction in this experiment. This is fully in line with previous experiments on Czech (Lacina & Chromý, 2022; Chromý et al., submitted).

[1] But note that regularizing priors generally favor the null model.

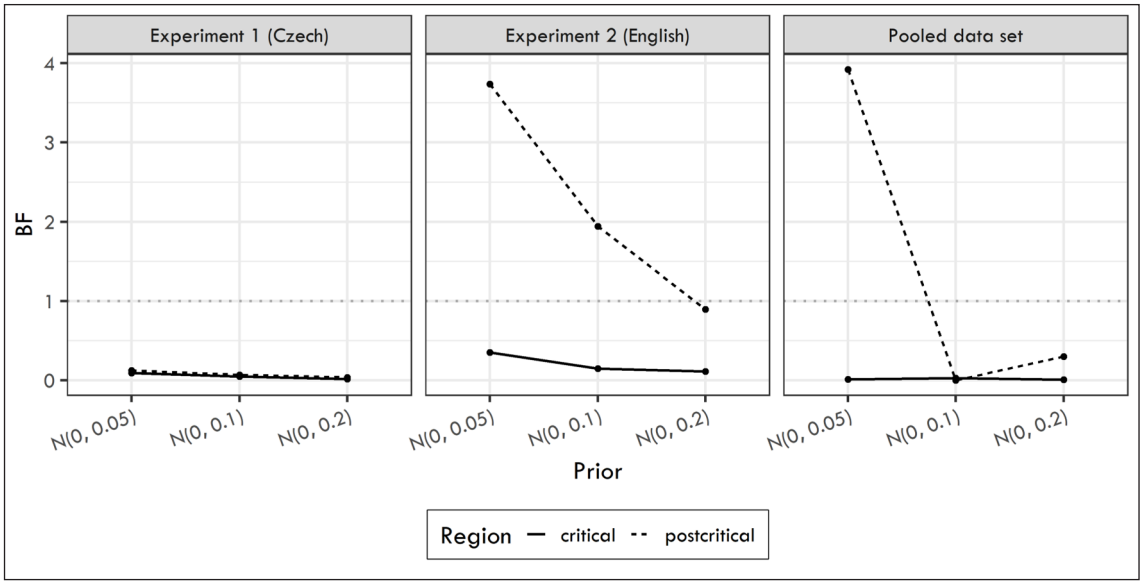

Figure 2: Summary of the Bayes factor analyses for Experiment 1, Experiment 2, and the pooled data set. For Experiments 1 and 2, Bayes factors show how much more likely a model with the attractor number by verb number interaction (i.e., attraction effect) is than the model without the interaction. For the pooled data set, Bayes factors show how much more likely a model with the three-way interaction (i.e. a difference in attraction effects between languages) is than the model without the three-way interaction. Bayes factor values greater than 1 favor the interaction, and the values less than 1 favor the null model.

4. Experiment 2: English

4.1 Method

4.1.1 Participants

We collected data from 200 native speakers of English, who participated in the experiment for payment of £8/hour through Prolific (www.prolific.co). Two participants were excluded due to their accuracy on filler item probes being lower than 75%. The resulting sample consisted of 191 participants (117 female, 76 male, 5 did not disclose; median age of 24 years). After the exclusion of the two participants, the average response accuracy for filler items was 94.1%, with the median being 95.34%. The participants were university students and thus had a similar educational background to the Czech participants in Experiment 1.

4.1.2 Materials

We used the same number of experimental sentences, fillers and practice sentences as in Experiment 1. All stimuli were translation equivalents of the Experiment 1 stimuli. An example experimental sentence is shown in **Table 2**.

Table 2: Example of experimental sentence from Experiment 2.

	1	2	3	4	5	6	7	8	9	10	11	12
a	The	file	for	the	archiver	likely	was	tightly	sealed	by	the	authorities.
b	The	file	for	the	archivers	likely	was	tightly	sealed	by	the	authorities.
c	The	file	for	the	archiver	likely	were	tightly	sealed	by	the	authorities.
d	The	file	for	the	archivers	likely	were	tightly	sealed	by	the	authorities.

4.1.3 Procedure

The experimental procedure was identical to that of Experiment 1.

4.1.4 Data analysis

The analysis workflow was identical to that of Experiment 1. The exclusion of RTs below 100 ms and above 7000 ms represented 0.3% of the data set. The upper cut-off point was decided based on a visual inspection of the data using Q-Q plots, following the recommendations by Baayen and Milin (2010).

4.2 Results

4.2.1 Descriptive statistics

Figure 3 shows the log-transformed RTs for each condition and sentence region. Raw RTs for each condition are presented in **Table 5** in Appendix C.

4.2.2 Bayesian modeling and discussion

In the precritical region, there was no indication of any main effect. At the critical and postcritical region, there was an indication of an effect of grammaticality: a speedup in grammatical sentences as compared to ungrammatical conditions with a plural verb (critical: $\hat{\beta} = -4.7$ ms, 95%-CrI: [–11, 2] ms, $P(\beta < 0) = 0.92$; postcritical: $\hat{\beta} = -25$ ms, 95%-CrI: [–33, –16] ms, $P(\beta < 0) > 0.99$). In both critical and postcritical regions, there was an indication of an attraction effect (critical: $\hat{\beta} = -5.6$ ms, 95%-CrI: [–13, 1.5] ms, $P(\beta < 0) = 0.94$; postcritical: $\hat{\beta} = -9.7$ ms, 95%-CrI: [–17, –2.4] ms, $P(\beta < 0) > 0.99$). In **Figure 2**, the evidence in favor of the attraction effect in English is summarized: For a range of informative priors, the model with the attraction effect is preferred.

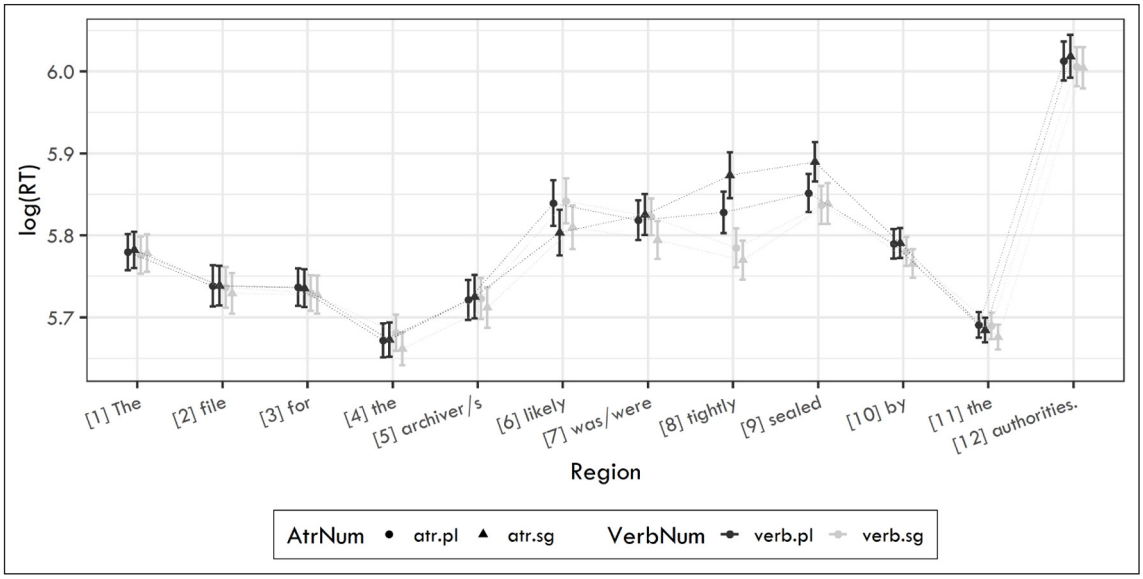

Figure 3: Mean log RTs together with 95% confidence intervals for the four conditions used in Experiment 2.

5. Aggregate analysis: The effect of language

The two experiments yielded different results: Experiment 1 showed a clear preference for the null model (i.e. no evidence for attraction effects in Czech), whereas Experiment 2 yielded a preference for the model containing an interaction between attractor number and verb number for English (i.e. evidence for typical agreement attraction effects). To directly compare attraction effects in Czech and English, we fitted a model on the pooled data set with language as a predictor: Czech coded as –1 and English as 1.[2] The model included the fixed effects of attractor number, verb number, and language, as well as all possible interactions between those effects. Differences in attraction effects between languages should manifest as a three-way interaction. The random effects structure was the same as in the models for Experiments 1 and 2 and as this was a between-subjects manipulation, it did not include random slopes for language.

In all regions, a main effect of language was found: English stimuli were read faster compared to Czech ones (precritical: $\hat{\beta} = -93$ ms, 95%-CrI: [–120, –63] ms, $P(\beta < 0) > 0.99$; critical: $\hat{\beta} = -58$ ms, 95%-CrI: [–80, –35] ms, $P(\beta < 0) > 0.99$; postcritical: $\hat{\beta} = -58$ ms, 95%-CrI: [–82, –35] ms, $P(\beta < 0) > 0.99$). In the precritical region only, there was a main effect

[2] The presence of an effect in one experiment and the absence of an effect in the other experiment does not constitute evidence for the difference between the effects across experiments.

of attractor: singular attractors were read faster than plural ones ($\hat{\beta} = -11$ ms, 95%-CrI: [–17, –4.4] ms, $P(\beta < 0) > 0.99$). In the critical and postcritical regions, there was a main effect of grammaticality (critical: $\hat{\beta} = -11$ ms, 95%-CrI: [–16, –5.6] ms, $P(\beta < 0) > 0.99$; postcritical: $\hat{\beta} = -29$ ms, 95%-CrI: [–34, –23] ms, $P(\beta < 0) > 0.99$). In the critical region, there was also an interaction between grammaticality and language ($\hat{\beta} = 5.7$ ms, 95%-CrI: [0.68, 11] ms, $P(\beta > 0) = 0.98$), suggesting that the grammaticality effect was less pronounced in English. Crucially, in the postcritical region, the expected three-way interaction was observed: $\hat{\beta} = -6.7$ ms, 95%-CrI: [–12, –1.8] ms, $P(\beta < 0) > 0.99$.

In **Figure 2**, the evidence supporting the difference in attraction effects between Czech and English is summarized: When informative priors are on the same scale as the effect sizes, the model that allows for differences between languages in attraction effects is preferred.

6. General discussion

The two experiments presented in this paper aimed to directly compare the presence and magnitude of number agreement attraction effects in two typologically different languages, namely, Czech and English.

Experiment 1, which tested the processing of Czech sentences by native Czech speakers, revealed no evidence in favor of the presence of agreement attraction effects. This is in line with the previous results of Lacina and Chromý (2022) and Chromý et al. (submitted). At the same time, robust ungrammaticality effects were observed – Czech comprehenders were reliably slower upon encountering a verb whose number mismatched that of the subject head. This shows that Czech participants were paying enough attention to the stimuli and were sensitive to morpho-syntactic violations. Experiment 2, which tested processing of English sentences by native English speakers, showed a different pattern: Here, we documented a clear effect of agreement attraction (plural attractors speeded up processing of ungrammatical, i.e. plural, verbs). Likewise, we saw a main effect of ungrammaticality. Crucially, the analysis of the aggregated data from the Czech and English experiments supported the difference in the attraction profile between the languages.

Whilst our results highlight the fact that differences exist across the two languages, we can only speculate about what is driving the differences. A possible explanation may be the presence of overt case marking in Czech. It is possible that the overt marking of the attractor noun as accusative actually blocks agreement attraction effects. However, this explanation is somewhat doubtful, because there are studies which show agreement attraction effects in languages with overt case marking, such as Russian (Slioussar, 2018) or Armenian (Avetisyan et al., 2020) and Slovak in production (Badecker & Kuminiak, 2007). In their paper, Chromý et al. (submitted) offer a different explanation of the differences between Czech and other languages that exhibit agreement attraction effects in comprehension, namely, that it is the strong reliance of Czech

on the formal agreement (and very limited use of semantic agreement) which effectively overcomes possible attraction in comprehension. According to this view, Czech speakers would be especially attentive to agreement and its violations (which is supported by the slowdown due to ungrammaticality being smaller in English than in Czech) and they would thus be less influenced by interfering noun phrases.

In general, there are many structural differences between Czech and English which could influence processing of sentences containing agreement attraction structures, and which cannot be easily distinguished based on the experiments presented here (and in previous studies). For example, English has fixed word order, whereas Czech word order is relatively free and the position of the NP in the sentence is not directly related to its syntactic status. Czech also has overt case marking that can be ambiguous (syncretic), but also highly informative in the given context. For example, overt marking can unambiguously specify the syntactic status of a certain NP, such as the subject NPs in the Czech experiment, which contained an unambiguously nominative ending. In contrast to English, Czech strongly relies on formal agreement, especially in the case of grammatical number. For example, Czech has many nouns which are formally plural, but semantically singular, such as *dveře* ('door'), or *plavky* ('swimsuit'). There are also nouns that are formally singular, but semantically plural, such as *námořnictvo* ('navy'), or *dobytek* ('cattle'). Such nouns would be ideal candidates for semantic agreement. However, even for these nouns, subject-verb agreement and relative pronoun agreement obligatorily match their formal number specification, not their semantic number (Ziková & Caha, 2004). It is extremely unlikely that a native Czech speaker would refer to a door or a swimsuit with a singular pronoun, or would use a plural pronoun to refer to the navy or cattle. Czech speakers are forced to be attentive to the morphological information of the NPs in order to overcome possible comprehension problems. Therefore, they might encode subject NPs more strongly than speakers of English, which may result in their resistance to agreement attraction effects. Moreover, there might be differences in processing information conveyed by the attractor NP. Chromý and Vojvodić (2023) report that optional/additional information is recalled to a much lesser extent than core information in Czech. In contrast, English speakers may be more attentive to such information, which would lead to the attractor number being more prominent.

Although the two experiments we present here were carefully designed to be matched, the participant samples were different. In the Czech experiment, participants were undergraduate students who received course credit, whereas in the English experiment, participants were native English speakers (with a similar education background and age compared to the Czech participants) who were paid for participating. This may have presented a confound – however, several factors suggest that this is not the main cause of the difference. First, both groups had a very high rate of response accuracy for comprehension questions: for fillers, the mean was 94.14% (Czech) and 94.1% (English) and the median was 95.01% (Czech) and 95.34% (English),

for experimental items, the mean was 93.74% (Czech) and 93.96% (English), with 95.83% being the median in both experiments. This indicates that the participants comprehended the stimuli equally well in both experiments. Second, both groups showed strong ungrammaticality effects, which can be taken as another sign of attentive reading. Third, the samples were matched to be relatively similar, based on their age, education and the student status. Fourth, other studies on English employed students too (Wagers et al., 2009; Cunnings & Sturt, 2018, etc.) and found typical attraction effects.

Another interesting observation is that the difference in general reading speed between languages is relatively large: 68 to 90 ms per word. This poses a highly interesting question on the general cross-linguistic differences of the reading process. Such differences are not unexpected (see Siegelman et al., 2022 and Liversedge et al., 2016, for interesting analyses of differences in eye-movements between native speakers of various languages in reading comparable stimuli in their respective languages), but are rather beyond the scope of the current paper. Crucially, we do not believe this difference between English and Czech participants in RTs would be a sign of a general difference in the degree of attention to the presented materials. As mentioned above, both groups completed the comprehension questions with about 95% median response accuracy, which directly contradicts a possible speed–accuracy trade-off. There may, of course, be some unseen differences between the groups that might have influenced the differences in RTs, but it does not seem to be the case that the Czech participants were more concentrated, focused or cautious than their English counterparts.

We have to stress that the potential explanations for cross-linguistic differences are rather speculative at this point, as the present data set does not allow us to distinguish between them. At the same time, they present interesting challenges and opportunities for future research. Our study highlights the importance of exploring cross-linguistic processing differences. New data may show whether the observed patterns reported in the current paper are due to the nature of the two languages, possible sample-related factors or even deep differences in processing mechanisms between the speakers of different languages.

7. Conclusion

The two experiments presented in this paper offer evidence against the cross-linguistic uniformity of agreement attraction effects. The agreement attraction effect was observed in English, but not in the translation-equivalent Czech experiment. A direct comparison of attraction effects on the pooled data set provided evidence supporting cross-linguistic differences in the attraction patterns. Our results suggest that language processing patterns may differ between languages even for structures that are almost identical.

Appendix A. Statistical modeling

The priors used for the models and their interpretation can be seen in **Table 3**. For the parameters not specified in **Table 3**, default brms priors were used. The models were run using four chains, with the default 2000 iterations per chain for effect size estimation, and the first half of iterations were discarded as warm-up iterations. For Bayes factor analyses, each chain ran for 30000 iterations with 2000 warmup iterations. For all models, convergence diagnostics indicated no convergence issues, i.e. $\hat{R}$ values were close to 1.

Table 3: Prior specifications for the models estimating effect sizes and their explanation. Reading times are modeled on the log-normal scale.

Parameter	Prior	Interpretation
Reading times		
Intercept	$\mathcal{N}(\mu = 5, \sigma = 1)$	RTs likely lie between 20 ms and 1100 ms.
β	$\mathcal{N}(\mu = 0, \sigma = 0.5)$	The estimated effects likely lie between –1880 ms and 1880 ms.
sd(β)	$\mathcal{N}(\mu = 0, \sigma = 0.5)$	By-item or by-participant deviation of any effect from the overall estimate likely lies between –1884 ms and 5122 ms.
σ	$\mathcal{N}(\mu = 0, \sigma = 1)$	Residual error likely lies between 0 ms and 7000 ms. Regularizing prior for the correlation matrix.
cor	$lkj(2)$	Strong correlations are discouraged.

Appendix B. Experiment 1 (Czech): Raw mean RTs in ms

Table 4: Raw mean RTs in ms (together with 95% confidence intervals) for the regions 3–8 for the four conditions used in Experiment 1.

	attractor	verb	3 attractor	4 attractor + 1	5 verb	6 verb + 1	7 verb + 2	8 verb + 3
a	sg	sg	445.04 [425.58, 464.49]	482.69 [464.92, 500.47]	415.68 [403.51, 427.85]	418.07 [400.82, 435.32]	460.43 [442.39, 478.47]	635.96 [605.57, 666.34]

(Condt.)

	attractor	verb	3 attractor	4 attractor + 1	5 verb	6 verb + 1	7 verb + 2	8 verb + 3
b	pl	sg	451.71 [426.78, 476.65]	486.91 [469.7, 504.12]	406.16 [396.21, 416.1]	414.15 [397.63, 430.66]	457.13 [442.76, 471.5]	607.66 [576.77, 638.55]
c	sg	pl	446.06 [424.82, 467.3]	483.07 [465.39, 500.74]	456.23 [437.17, 475.29]	451.82 [437.34, 466.3]	482.93 [465.46, 500.4]	660.78 [626.86, 694.7]
d	pl	pl	444.47 [423.58, 465.37]	509.93 [488.28, 531.58]	432.83 [420.88, 444.79]	455.94 [442.09, 469.79]	492.46 [471.23, 513.69]	660.33 [625.46, 695.19]

Appendix C. Experiment 2 (English): Raw mean RTs in ms

Table 5: Raw mean RTs in ms (together with 95% confidence intervals) for regions 5–10 for the four conditions used in Experiment 2.

	attractor	verb	5 attractor	6 attractor + 1	7 verb	8 verb + 1	9 verb + 2	10 verb + 3
a	sg	sg	344.38 [325.82, 362.93]	378.49 [364.87, 392.12]	358.87 [348.99, 368.75]	355.4 [343.32, 367.47]	385.96 [368.94, 402.97]	334.78 [328.51, 341.06]
b	pl	sg	345.92 [329.47, 362.37]	392 [378.57, 405.43]	367.27 [357.36, 377.18]	362.61 [347.22, 378]	376.07 [362.87, 389.27]	341.96 [333.62, 350.29]
c	sg	pl	357.42 [336.17, 378.68]	383.26 [366.68, 399.83]	377.36 [364.84, 389.89]	408.33 [392.41, 424.26]	397.34 [385.97, 408.72]	345.79 [338.13, 353.46]
d	pl	pl	340.35 [328.46, 352.24]	392.73 [378.71, 406.74]	371.87 [360.56, 383.19]	380.06 [367.72, 392.4]	379.49 [369.51, 389.47]	344.42 [337.4, 351.44]

Data availability statement

The experimental items, fillers, data, and analysis code from both experiments are available on the OSF: http://doi.org/10.17605/OSF.IO/2KJY5. Both experiments were also preregistered: https://doi.org/10.17605/OSF.IO/4DZ7J.

Ethics and consent

The study was approved by the Research Ethics Committee of Faculty of Arts, Charles University (Ref. No.: UKFF/272908/2022).

Acknowledgements

JC was supported by Alexander von Humboldt Foundation and by the Charles University institutional program Cooperatio. AL received funding from Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Project-ID 317633480, SFB 1287. JB received funding from Charles University's PRIMUS grant scheme, 21/HUM/015. RL was supported by the German Research Foundation (DFG) as part of the Emmy Noether project awarded to Nicole Gotzner (Grant Nr. GO 3378/1-1).

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Jan Chromý: conceptualization; data curation; methodology; resources; software; writing – visualization; original draft, writing – review & editing; project administration

James Brand: conceptualization; methodology; resources; writing – review & editing

Anna Laurinavichyute: formal analysis; visualization; validation; writing – review & editing

Radim Lacina: conceptualization; methodology; resources; writing – original draft, writing – review & editing; project administration

References

Acuña-Fariña, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua. International review of general linguistics, 143,* 108–128. DOI: https://doi.org/10.1016/j.lingua.2014.01.013

Arehalli, S., & Wittenberg, E. (2021). Experimental filler design influences error correction rates in a word restoration paradigm. *Linguistics Vanguard, 7*(1), 20200052. DOI: https://doi.org/10.1515/lingvan-2020-0052

Avetisyan, S., Lago, S., & Vasishth, S. (2020). Does case marking affect agreement attraction in comprehension? *Journal of Memory and Language, 112,* 104087. DOI: https://doi.org/10.1016/j.jml.2020.104087

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12–28. DOI: https://doi.org/10.21500/20112084.807

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language, 56*(1), 65–85. DOI: https://doi.org/10.1016/j.jml.2006.08.004

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive psychology, 23*(1), 45–93. DOI: https://doi.org/10.1016/0010-0285(91)90003-7

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software, 80*, 1–28. DOI: https://doi.org/10.18637/jss.v080.i01

Chromý, J., Lacina, R., & Dotlačil, J. (submitted). Number Agreement Attraction in Czech Comprehension: Negligible Facilitation Effects. *Open Mind.*

Chromý, J., & Vojvodić, S. (2023). When and where did it happen? Systematic differences in recall of core and optional sentence information. *Quarterly Journal of Experimental Psychology*. DOI: https://doi.org/10.1177/17470218231159190

Comrie, B., Haspelmath, M., & Bickel, B. (2008). The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses. https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf

Cunnings, I., & Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language, 102*, 16–27. DOI: https://doi.org/10.1016/j.jml.2018.05.001

Engelmann, F., Jäger, L. A., & Vasishth, S. (2019). The effect of prominence and cue association on retrieval processes: A computational account. *Cognitive Science, 43*(12), e12800. DOI: https://doi.org/10.1111/cogs.12800

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS one, 8*(10), e77661. DOI: https://doi.org/10.1371/journal.pone.0077661

Franck, J., Colonna, S., & Rizzi, L. (2015). Task-dependency and structure-dependency in number interference effects in sentence comprehension. *Frontiers in psychology, 6*, 349. DOI: https://doi.org/10.3389/fpsyg.2015.00349

Franck, J., & Wagers, M. (2020). Hierarchical structure and memory mechanisms in agreement attraction. *Plos one, 15*(5), e0232163. DOI: https://doi.org/10.1371/journal.pone.0232163

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive psychology, 110,* 70–104. DOI: https://doi.org/10.1016/j.cogpsych.2019.01.001

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. DOI: https://doi.org/10.1016/j.jml.2017.01.004

Lacina, R. (2023). Under no illusion: An acceptability study on Czech agreement attraction. *PsyArXiv*. DOI: https://doi.org/10.31234/osf.io/3tjw6

Lacina, R., & Chromý, J. (2022). No agreement attraction facilitation observed in Czech: Not even syncretism helps. *Proceedings of the Annual Meeting of the Cognitive Science Society, 44*, 2423–2430. https://escholarship.org/uc/item/2vq4g622. DOI: https://doi.org/10.31234/osf.io/ktyv8

Lago, S., & Felser, C. (2018). Agreement attraction in native and nonnative speakers of German. *Applied Psycholinguistics, 39*(3), 619–647. DOI: https://doi.org/10.1017/S0142716417000601

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in Spanish comprehension. *Journal of Memory and Language, 82*, 133–149. DOI: https://doi.org/10.1016/j.jml.2015.02.002

Laurinavichyute, A., & von der Malsburg, T. (2022). Semantic attraction in sentence comprehension. *Cognitive Science, 46*(2), e13086. DOI: https://doi.org/10.1111/cogs.13086

Laurinavichyute, A., & von der Malsburg, T. (2023). Agreement attraction in grammatical sentences and the role of the task. DOI: https://doi.org/10.31234/osf.io/n75vc

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition, 147*, 1–20. DOI: https://doi.org/10.1016/j.cognition.2015.10.013

Parker, D., & An, A. (2018). Not all phrases are equally attractive: Experimental evidence for selective agreement attraction effects. *Frontiers in psychology, 9,* 1566. DOI: https://doi.org/10.3389/fpsyg.2018.01566

Paspali, A., & Marinis, T. (2020). Gender agreement attraction in Greek comprehension. *Frontiers in psychology, 11*, 717. DOI: https://doi.org/10.3389/fpsyg.2020.00717

Phillips, C., Wagers, M., & Lau, E. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. *Experiments at the Interfaces* (pp. 147–180). Brill. DOI: https://doi.org/10.1163/9781780523750_006

R Core Team. (2022). *R: A language and environment for statistical computing* (manual). Vienna, Austria. https://www.R-project.org/

Siegelman, N., Schroeder, S., Acartürk, C., Ahn, H.-D., Alexeeva, S., Amenta, S., Bertram, R., Bonandrini, R., Brysbaert, M., Chernova, D., et al. (2022). Expanding horizons of crosslinguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior research methods,* 1–21. DOI: https://doi.org/10.3758/s13428-021-01772-6

Slioussar, N. (2018). Forms and features: The role of syncretism in number agreement attraction. *Journal of Memory and Language, 101,* 51–63. DOI: https://doi.org/10.1016/j.jml.2018.03.006

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of memory and language, 76,* 195–215. DOI: https://doi.org/10.1016/j.jml.2014.07.003

Tucker, M. A., Idrissi, A., & Almeida, D. (2015). Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in psychology, 6*, 347. DOI: https://doi.org/10.3389/fpsyg.2015.00347

Tucker, M. A., Idrissi, A., & Almeida, D. (2021). Attraction effects for verbal gender and number are similar but not identical: Self-paced reading evidence from Modern Standard Arabic. *Frontiers in psychology,* 3774. DOI: https://doi.org/10.3389/fpsyg.2020.586464

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of phonetics, 71*, 147–161. DOI: https://doi.org/10.1016/j.wocn.2018.07.008

Wagers, M., Lau, E., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language, 61*(2), 206–237. DOI: https://doi.org/10.1016/j.jml.2009.04.002

Zehr, J., & Schwarz, F. (2018). PennController for internet based experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832

Ziková, M., & Caha, P. (2004). Neshody s děvčetem (k hierarchii jmenných rysů). In P. Karlík (Ed.), *Korpus jako zdroj dat o češtině* (pp. 71–78). Masarykova univerzita.