

Acceptability Rating of Ungrammatical Colloquial Latvian: How Native Speakers Judge Different Error Types

Jurģis ŠKILTERS¹, Līga ZARIŅA¹,
Eglė ŽILINSKAITĒ-ŠINKŪNIENĒ², Velta SKOLMEISTERE¹

¹University of Latvia, Laboratory for Perceptual and Cognitive Systems at the Faculty of Computing, Faculty of Computing, Raina blvd. 19, Rīga, Latvia

²Vilnius University, Institute for the Languages and Cultures of the Baltic, Faculty of Philology, Universiteto st. 5, Vilnius, Lithuania

`jurgis.skilters@lu.lv`, `liga.zarina@lu.lv`,
`egle.zilinskaite@gmail.com`, `velta.sk@gmail.com`

Abstract. The aim of the current study is to test what types of sentences according to their grammatical structure are (a) considered to be more plausible from a native speaker's view and (b) perceived as better understandable even if not entirely grammatical.

A quasi-experimental task in a repeated measures design was conducted. 83 native speakers of Latvian have rated a randomized and balanced set of most typical grammatical errors (together with correct sentences and the filler sentences). Importantly our sample contained all regions of Latvia with different dialects (and even a different writing system).

According to our results, errors in usage of verbal prefixes, definite / indefinite endings, and word order errors are the closest to the ratings of correct sentences. The most crucial errors recognized by native speakers are the errors of prepositional usage and word coordination.

To our knowledge, this is the first acceptability rating experiment for the Baltic languages.

Keywords: grammaticality, acceptability rating, colloquial Latvian, types of grammatical errors

1. The problem

Latvian is a fusional language that belongs to morphologically rich but small (in terms of speaker population) languages that are neglected in terms of analysis of their grammatical structure and its correspondence to cognitive processing. Even more under-explored are the features of acceptability of certain types of sentences. In this study we are focusing on a subpart of the problem: we are aiming at exploring native language acceptability of colloquial Latvian which is considered ungrammatical according to traditional view of Latvian grammar.

We assume the framework of graded grammaticality and explore which features of grammaticality are more important for the understanding of the sentence and which features of ungrammaticality are tolerated by native speakers (i.e., which are the sentence types that native speakers of Latvian accept as good enough although they are

not entirely grammatical)? The relations between acceptability rating and the structure of graded grammaticality are documented in several studies. E.g., according to some previous research there are some constructions that are considered as ungrammatical according to theoretical criteria but still are judged as acceptable by speakers (Bresnan, 2007).

A wide range of evidence supporting the idea that grammaticality varies comes from empirical work on acceptability (e.g., Cowart, 1997, Sprouse, 2008, Schütze and Sprouse, 2013).

The idea that grammaticality is a binary feature (i.e., that there are either grammatical or non-grammatical sentences) is also challenged in classical approaches in computational linguistics; according to Manning and Schütze (1999, 9), “there are many reasons to be interested in the frequency with which different sentences and sentence types are used, and simply dividing sentences into grammatical and ungrammatical sentences gives no information about this. For instance, very often non-native speakers say or write things that are not in any way syntactically ungrammatical, but just somehow odd.”

Assuming that there are graded structure of grammaticality, what are the links of this structure to native speaker interpretation? One way of exploring this is to run some of the canonical sentence judgment or sentence acceptability tests. This was what we were doing in our study.

2. Theoretical framework

The current work uses the acceptability judgement test (Cowart, 1997, Sprouse, 2008, Schütze, 1996/2016, Schütze and Sprouse, 2013) in the framework of graded structure of grammatical knowledge: we are assuming that grammaticality is a graded phenomenon having different degrees of grammatical and cognitive prominence.

Gradience seems to be pervasive in grammaticality categories; according to a variety of empirical studies, grammaticality is a continuous feature with certain consistency regarding the acceptability ratings among responses (Lau et al., 2016, Schütze and Sprouse, 2013). Further, gradience in grammaticality seems to be more compatible with connectionist and neural network computational models than with the rule-based and top-down (Allen and Seidenberg, 1999). An important question from the perspective of computability of natural language: what are the relations between *acceptability* conducted by human raters and *probability* computed by a deterministic device (where the likelihood of occurrence is determined by factors such as sentence length and lexical frequency). Lau et al. (2016) argues that although it is not simply possible to reduce acceptability to probability, acceptability can be predicted based on probability. A crucial generalization from the work by Lau et al. (2016) indicates that linguistic knowledge is to a significant extent probabilistic. (Cp. also Bresnan (2007), who argues that grammaticality rather reflects probabilities instead of categorical constraints.)

Graded grammaticality framework has several alternative explanations and possible criticisms:

Gradience is a result of performance and is due to processing features and, therefore, does not entail gradience in grammar (Hofmeister and Sag, 2010 after Lau et al., 2016). However, processing and grammatical properties seem to systematically overlap and can be predictably explored in the same time keeping in mind that there is interaction between both. However, acceptability judgements are not equally sensible to all processing effects: syntactic knowledge seems to be directly related to acceptability ratings, whereas semantic – not (Sprouse, 2008).

Then it might be also argued that *probabilistic and binary views on grammaticality are not exclusive* since non-grammatical sentences might have non-zero probability of being rated as unacceptable whereas grammatical sentences are zero probability sentences.

Still another possible criticism is that *grammaticality as such is not metalinguistic task, whereas acceptability grading – is*. However, different areas where acceptability grading tasks have been applied (i.e., not only language but, e.g., perception) indicate systematic pattern of data and valid possibilities to generalize the results into theories (Schütze and Sprouse, 2013, 28f.). Metalinguistic features of acceptability grading task are also related to the fact that acceptability grading is an off-line measure of language processing (i.e., measured phenomena are after their initial processing). But still there seem to be a convincing set of evidences supporting the methodological validity of acceptability rating (cp. Schütze, 1996/2016, Sprouse et al., 2013) which seems to be not the case with some other measurements based on psychophysical principles. (E.g., magnitude estimation is less meaningful and valid than other acceptability rating techniques, cp. Sprouse, 2011).

To sum up, we assume that acceptability is a complex cognitive phenomenon (1) arising spontaneously in reaction to certain linguistic stimuli – word strings that are more or less similar to sentences; (2) integrating both grammatical and other cognitive processes that, in turn, yields consistent and robust results. Acceptability measuring is not substantially different from psychophysical approaches where subjects (e.g., seeing or hearing stimuli) have to report the intensity of a dimension of a corresponding perceptual domain along a scale (cp. also Schütze and Sprouse, 2013, 28). Further, acceptability grading is a unique measure since it cannot be replaced by another type of measure (e.g., neural or corpus) – neural analyses would not provide the researcher with a sufficiently sensitive measure to ungrammatical sentences only and a sentence that appears or does not appear in a corpus might or might not be grammatical – corpus does not provide an evidence for or against it (cp. Schütze and Sprouse, 2013, 29; for an argument against overgeneralization of web-search results and the necessary and complimentary role of acceptability judgement test see also Schütze, 2009). Moreover, low frequency lexical and syntactic items seem to be more plausible for acceptability rating than corpus-based frequency analysis (Divjak, 2017).

A limitation of acceptability measuring is the lack of a direct link between the results of acceptability reflecting more general cognitive processes (consistently including

grammatical ones)¹ and syntactic system which means that the rating of linguistic and grammatical features is co-determined by general cognitive processing factors that are gradient and are shaped by the features specific to task and subject (Schütze and Sprouse, 2013, Schütze, 1996/2016, Cowart, 1997, see Figure 1).

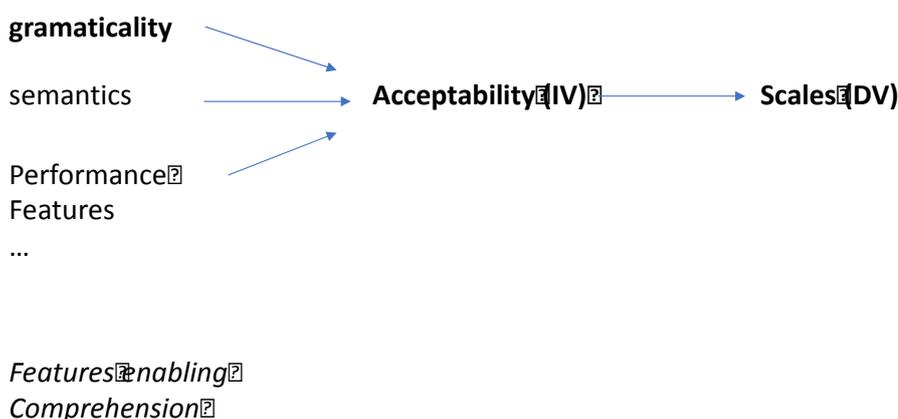


Fig. 1. Overall structure of acceptability rating task indicating the relations between independent and dependent variables (IV and DV)

Sometimes acceptability and grammaticality are distinguished (e.g., Sprouse et al., 2013). Agreeing with their results, we are assuming that sentence acceptability is a larger phenomenon integrating grammaticality, semantics, performance effects and other effects of cognitive processing. We also agree that “Perceptions of acceptability [...] arise as an automatic consequence of sentence comprehension, as they cannot be consciously suppressed by native speakers. Acceptability is generally considered a composite property, as several factors appear to affect acceptability judgments.” (Sprouse et al., 2013, 220f.). Although there is a systematic pattern of grammaticality rating, Sprouse et al. suggest to call these measures acceptability judgements instead of grammaticality judgements.²

An interesting issue for future studies is to explore how other cognitive processes impact grammaticality and to what extent. Memory limitations can constrain the understanding of entirely grammatical sentences if they are large and contain a complex structure and specific lexical constituents (cp. Cowart, 1997, 8).

¹ We agree with Cowart (1997, 7) that “theories of grammar are partial models of the human cognitive system that should, among other things, help to explain why judgements pattern as they do.”

² Again to agree with Cowart, “[T]here is no such thing as an absolutely acceptable or unacceptable sentence, although there might be such a thing as an absolutely grammatical or ungrammatical sentence.” (Cowart, 1997, 9).

Another option that is also plausible in the light of the current work: syntax is a consequence of other cognitive phenomena and grammatical representations are the result of implementation dependent system (cp. Phillips, 2013, Phillips and Lewis, 2013). In the same time evidence supports the view that grammatical system and language processing system are different aspects of the same cognitive system³ (Lewis and Phillips, 2015).

The assumption that grammaticality grading is a part of a larger configuration of cognitive processes that are actually graded raises the questions regarding the overall structure of language processing and the ways it is determined by overall cognitive and perceptual processes. A framework that provides some explanatory background is Townsend and Bever's (2001, 164–167) model according to which sentence processing consists of three stages: (1) assigning a likely meaning and a tentative syntax (*pseudosyntax*). In this stage the likely lexical items and heads are computed such that syntactic structure relates arguments to predicates. More in details it means (a) lexical recognition, (b) phrase segregation, and (c) assignment of a tentative configurational syntactic structure. (2) A stage of candidate real syntax (*forming the syntax*): a tentative meaning / form structure is mapped from pseudosyntax onto structures leading to syntactic derivation. (3) Complete syntax is assigned; detailed syntactic description (word order, agreement markings etc.) is matched to the original input. In this step, the output of the stage 2 is checked against the initial input and if needed reanalysis is started.

The complete syntax and pseudo-syntax are computed independently which might explain the relative insensitivity to certain grammatical errors. The well-formedness constraints also function on the third stage.

Underlying intuition of Townsend and Bever is that the sentence is understood before “a complete and correct parse is assigned” (Townsend and Bever, 2001, 164). This also supports the idea that humans are sensitive to semantic information before complete syntactic structure is established. Furthermore, this provides with an explanation of pragmatic effects in discourse: meaning containing pragmatic information comes before verb morphology (Townsend and Bever, 2001, 203–206).

The theoretical framework of the present study is also consistent with some recent prominent work on language processing (Christiansen and Chater, 2016) according to which language processing has some general perceptual and cognitive constraints: first, perceptual input in general (and linguistic input in particular as a part of it) must be compressed and recoded as fast and economical as possible; second, compressed encodings are transformed into several representational levels; third, cognitive systems use anticipation principle according to which prior information constrains “the recoding

³ “By *system* we mean a collection of cognitive mechanisms with a distinct purpose, operating over representations of distinct kind.” (Lewis and Phillips, 2015, 28)

of current perceptual input” (Christiansen and Chater, 2016, 4).⁴ Also this might provide explanations why certain strings are tolerated even if not entirely grammatical.

3. Overall Design

A set of random ungrammatical sentences (that are automatically (as in Lau et al., 2016) or semi-automatically generated or extracted from native language speakers’ corpora) are graded regarding acceptability by the set of native-speakers’ subjects.

The acceptability is graded according to Likert 5-point scale, where 1 is the most acceptable and 5 – the least acceptable.

Sentence stimuli for grading were prepared according to the most frequent errors in a corpus of ungrammatical sentences. The error-annotated corpus of Latvian that was used as input material for stimuli sentences was elaborated by software company Tilde Ltd; it was based on student papers (5157 sentences) and balanced texts (10563 sentences) (Deksne and Skadina, 2014). Additionally, a test corpus (5157 sentences) and a corpus of Latvian as a second language (679 sentences) were used. It has to be noted that the *error* here is used as a cover term which encompasses the stimuli ranging from ambiguous to irregular or clearly erroneous cases.

The categories of grammatical errors are not exactly corresponding to the morphological features of Latvian but reflect the actual use of language and its errors.

4. Method

We used a sentence acceptability judgement rating method (Cowart, 1997, Sprouse, 2008, Schütze, 1996/2016, Schütze and Sprouse, 2013) in a repeated measures (in-group) quasi-experimental research design. Experiment was conducted online (for a similar procedure cp. Sprouse et al. 2013).

By using descriptive statistics and statistical tests (Friedman test, Kruskal-Wallis test, t-test, Wilcoxon signed-rank test, Kendall and Spearman correlation) we determined and characterized differences between acceptability evaluations regarding certain groups of sentences distinguished after grammatical error types and within each such group of sentences. We compared these evaluations with evaluations of anchor sentences, grammatically correct sentences and filler sentences. We also evaluated if there can be

⁴ “Therefore, using prior information to *predict* future input is likely to be essential to *successfully encoding that future input* (as well as helping us to react faster to such input). Anticipation allows faster, and hence more effective, recoding [...]” (Christiansen and Chater, 2016, 4).

detected regularities depending on sentences' length and word commonness. The tool for statistical processing was IBM SPSS Statistics 22.

5. Participants

Although usually in acceptability rating tasks a bit more than 20 participants are used (but see Divjak, 2017, who used 285 subjects), in our study we had 83 subjects; all native speakers. Our sample was constructed to include all regions of Latvia (Riga, Vidzeme, Kurzeme, Latgale, Zemgale), all age groups, and an equal gender distribution.

The rationale behind these sampling principles was to check whether there are some effects on the grammaticality among different regions of Latvia (we assumed that there might be some regional differences and some impacts on grammaticality ratings because there are different dialects spoken in different parts of Latvia)⁵ and different age groups (since the language spoken by the older generation might be influenced by Russian and the language of the younger generation has a larger impact of English, we might assume that there are some minor syntactic impacts). The resulting sample was not entirely corresponding to the initial plan but still consisted of all sampling categories, although to different degree.

Sample consisted of 63 female and 20 male participants. The distribution regarding age was: 27% 16–30, 25% 31–45, 33% 46–60 and 16% over 61 years old. The sample is relatively balanced among all Latvian regions – Vidzeme 27%, Latgale 13%, Kurzeme 18%, Zemgale 13% and the capital of Latvia – Rīga 29%.

Most of the participants have the higher education (78%). The participants' education corresponds to different fields (main fields are humanities or social sciences (43%) and natural sciences or engineering sciences (35%)).

The native language of all respondents is Latvian. Most of the respondents are with other language knowledge (63% indicated Russian language as the second most well known language and 31% reported English), but 25% had learned Latvian language additionally at the university besides family and school.

6. Stimuli

The stimuli set consists of anchor sentences, non-grammatical sentences, grammatically correct sentences and filler sentences. Non-grammatical sentences correspond to 8

⁵ Our data come from the regions that are represented by all three Latvian dialects: Livonian dialect, Middle Latvian and High Latvian. The former two are called Low Latvian and are spoken in the Western and Central part of Latvia (Kurzeme and Riga regions and part of Zemgale and Vidzeme regions), and the latter one is spoken in Eastern Latvia. Part of this dialect, spoken mainly in Latgale region, has its standard language called Latgalian. Written Latgalian stems from the year 1753 when the first Latgalian book was published.

criteria (Table 1; Sprouse et al. (2013) used 150 acceptable and 150 unacceptable sentences). Different types of sentences (non-grammatical, correct and filler sentences) were used to avoid negative or affirmative bias in subjects (more concerning use of fillers in stimuli generation cp. Schütze, 1996/2016, 181).

Table 1: Stimuli types used in the study

| Sentence type | Number of sentences |
|---|---------------------|
| Anchoring sentence: acceptable | 2 |
| Anchoring sentence: moderate acceptable | 2 |
| Anchoring sentence: unacceptable | 2 |
| 1. Errors in definite / indefinite ending | 8 |
| 2. Errors in number and case of nouns | 8 |
| 3. Errors in verb forms | 8 |
| 4. Errors in usage of pronouns | 8 |
| 5. Errors in usage of prepositions | 8 |
| 6. Errors in word coordination | 8 |
| 7. Errors in word order | 8 |
| 8. Errors in usage of verbal prefixes | 8 |
| Correct sentence | 64 |
| Filler sentence: question | 32 |
| Filler sentence: exclamation | 32 |
| Total | 198 |

Six of sentence types of the input stimuli are coherent but two of them have rather heterogeneous structure. Errors in number and case of nouns include sentences with *plurale tantum* used in singular and vice versa while the errors of case represent the usage of the accusative instead of the nominative object marking in debitive construction, the nominative instead of the genitive marking of a negated object in a negative possessive construction and adverb *cik* ‘how many’ governing the nominative instead of the genitive case. Errors in word coordination encompass instances of nominal agreement (6 sentences) and erroneous prepositional governing (2 sentences).

The relative frequency of the words and word and sentence length were controlled taking into account the characteristics of corpus provided by Tilde (cp. also Schütze, 1996/2016, 181) (Figure 2).

We avoided extremes in imaginable and concrete sentences, and sentences with highly unrepresentative or unnatural semantic content (cp. Schütze, 1996/2016, 183).

All stimuli were generated according to the principles that rare, untypical, abstract, long, specific words, extreme differences in emotional valence (both negative or positive) are avoided; the content of the sentences is different but not too abstract and not complex. The average length of words in each group of sentences was from 5–7 letters while the

minimal amount of letters in each group was 2 and maximal from 9–18 letters (average maximal length – 14 letters).

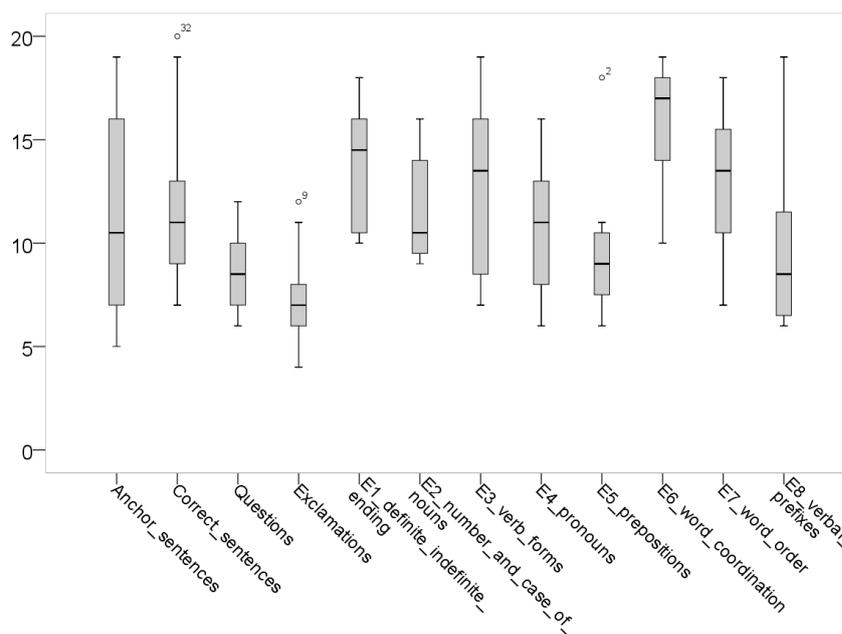


Fig. 2: Lengths of sentences used in the study, amount of words

We avoided extremes in imaginable and concrete sentences, and sentences with highly unrepresentative or unnatural semantic content (cp. Schütze, 1996/2016, 183).

All stimuli were generated according to the principles that rare, untypical, abstract, long, specific words, extreme differences in emotional valence (both negative or positive) are avoided; the content of the sentences is different but not too abstract and not complex. The average length of words in each group of sentences was from 5–7 letters while the minimal amount of letters in each group was 2 and maximal from 9–18 letters (average maximal length – 14 letters).

7. Design and procedure

For online questionnaire preparation we used QuestionPro™ tool.

Instructions:

Before showing stimuli subjects were instructed (Schütze and Sprouse, 2013, 36):

Imagine that the following sentences are being spoken by a friend. Would these sentences make you sound like a native speaker of Latvian?

The emphasis on spoken modality was given to enable the evaluation of acceptability according to native-speaker sense of interpreting language rather than just plausibility or frequency. Subjects were also instructed that our study is not concerned with prescriptive rules of grammar, truth and plausibility of its content.

The rationale of the instruction is to induce subjects' native speaker intuition. "The instructions essentially ask the subjects to consider whether they feel the stimuli sound like possible English sentences for them, and to concentrate on structure." (Schütze, 1996/2016, 184).

Subjects had to rate sentences (according to 5 pt Likert-scale) how acceptable each sentence is.

Sequence of stimuli:

We are aware of order of presentation and context effects which is the reason why we randomized every set of stimuli for each participant therefore distributing the order effects across the experimental conditions (Coward, 1997, 51).

First 6 anchoring items in randomized order were presented and after followed the experimental phase with different type of sentences (randomized non-grammatical, grammatically correct and filler stimuli). Finally, the demographic information was asked: age, gender, geographic / regional affiliation, native language, second best known language, Latvian language courses in the past, education, education field, occupation and handedness (Divjak, 2017, Cowart, 1997).

Anchoring items (two acceptable, two unacceptable and two moderate) are provided showing the lowest and highest points in the scale; these 6 items were used for practicing purposes to familiarize the subject with the task (these are the unannounced practice items; the subjects were not informed that the first six are practicing items. They included almost all variety and range of acceptability, cp. Schütze and Sprouse, 2013, 37, Schütze, 1996/2016, 185).

Filler sentences (questions and exclamations) did not include particular types of ungrammaticality or particular types of error. The rationale for using filler sentences was

(a) to decrease the probability that the subjects might get aware of our ungrammaticality criteria, (b) to reduce scale bias and to vary the possibilities of rating.⁶

8. Results

The filling of the questionnaire was relatively time consuming (average time spent – 53 min while 70% of respondents filled the questionnaire until 50 min and minimal time spent was 17 min).

The anchoring items set interval of acceptability from average 1.72 respective to grammatically correct sentences to average 4.88 respective to non-grammatical sentences. The average evaluations of correct sentences and filler sentences is from average 1.89 to average 1.97 (Figure 3). The anchor sentences within each group are evaluated homogenously. However, regarding the correct and filler sentences, the data reflect that even grammatically correct sentences are evaluated with a rather large amplitude respective to acceptability.

In average, the most tolerated error type according to our results is misuse of a verbal prefix (Figure 3, error type 8, average evaluation 2.3), whereas the error type that was rated as the most significant (most unacceptable in respect to native language intuitions) is improper usage of prepositions (Figure 3, error type 5, average evaluation 3.4).

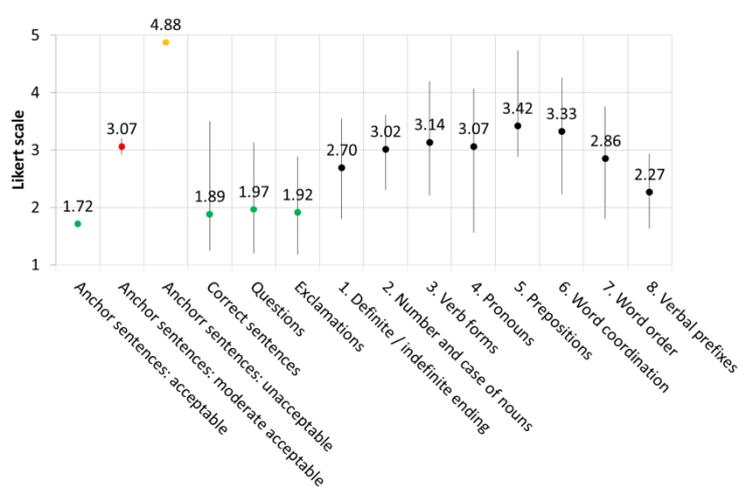


Fig. 3. Average, maximal and minimal evaluation values of average evaluations of sentences in each group of sentence types included in the study (Table 1).

⁶ In our approach we did not use fillers as a direct background against which experimental sentences are judged (Cowart 1997, 52).

However, to have a more comprehensive picture of our results we also have to take into account the differences that correspond to sentences' evaluations within each group of non-grammatical sentences (Figure 4). The Friedman test pointed to significant differences ($\alpha=0.05$) comparing mean values regarding each group of sentences, except anchor sentences.

This might be due to semantic reasons, e.g., less typical or eventually ambiguous words, sentence length. As for the sentence length: there does not seem to be an unambiguous link between acceptability rating and sentence length. However, in some types of errors there are separate sentences which indicate stronger positive or negative (the longer – the more unacceptable and the other way around) associations (Spearman rank correlation). This, however, shows no systematic overall pattern. We assume that the sentence meaning might contribute in those cases instead of the length of sentences. This is supported by a closer analysis of filler sentences: the length of filler sentences and the acceptability have a weak association ($\rho<0.3$).

Although spatial and non-spatial uses of prepositions were included, no significant difference can be observed. However according to the ratings, the most unacceptable is the sentence with explicitly spatial use (*Lielais un jaunais dzīvoklis bija uz piektā stāva* 'The large and new apartment was on the fifth floor' (in Latvian the Locative case *piektajā stāvā* instead of [*uz + Gen.*] 'on' should be used)).

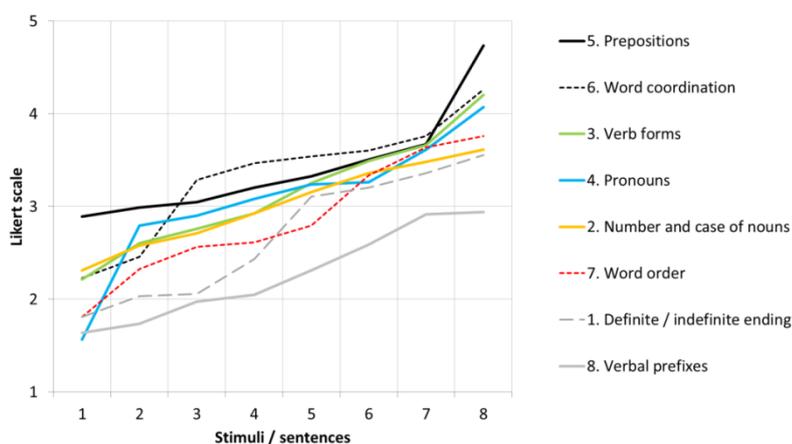


Fig. 4. Average ratings of all incorrect sentences included in the study (distinguished according grammatical error types).

Data indicate that the evaluation of one sentence could significantly influence general average value (Figure 5a); in the same time the distribution of evaluations for most sentences is relatively wide (Figure 5b).

Such tendencies occur in every error type; however, it is possible to observe general trends of error evaluations (Figure 6) according to which the most unacceptable ones refer to the error types 5 and 6, but the most acceptable ones are the types 1 and 8. The next more acceptable error types are 2 and 7, which are followed by the error types 3 and 4.

These tendencies are also reflected in the results of t-tests indicating significant differences ($\alpha=0.05$) between average evaluations of different types of errors. We were able to distinguish groups of error types that are similar regarding average evaluations to each other, but statistically significantly differ from other types of errors (Figure 7). These results show that incorrect verbal prefix statistically is not evaluated differently from filler sentences regarding acceptability.

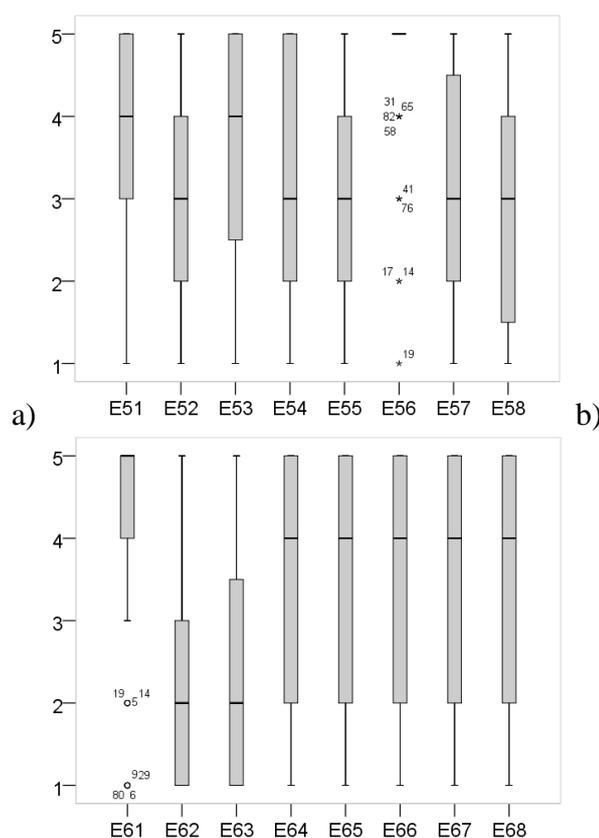


Fig. 5. Differences of sentences' ratings within a) errors in prepositions (error type 5) and b) errors of word coordination (error type 6).

Kruskal-Wallis test was conducted to compare the average values depending from the demographic factors. According to our results, there are no systematic overall pattern corresponding to significant impacts of demographic factors (e.g., place of residence or the time spent in a particular part of Latvia that might be linked to a dialect does not have an impact on rating acceptability; also age and other demographic variables seem to have no impact on the results). Also the correlation analysis (Spearman rank correlation) indicates that the demographic factors have no significant associations ($\alpha < 0.05$) or that there is a weak correlations only ($\rho < 0.35$).

Finally, our results show a pattern of overall validity because the filler items (questions or exclamations) were rated as grammatically relatively acceptable (they did not include any grammatical errors); also the starting anchoring items were rated according to different degrees of acceptability depending on their degree of deviation.

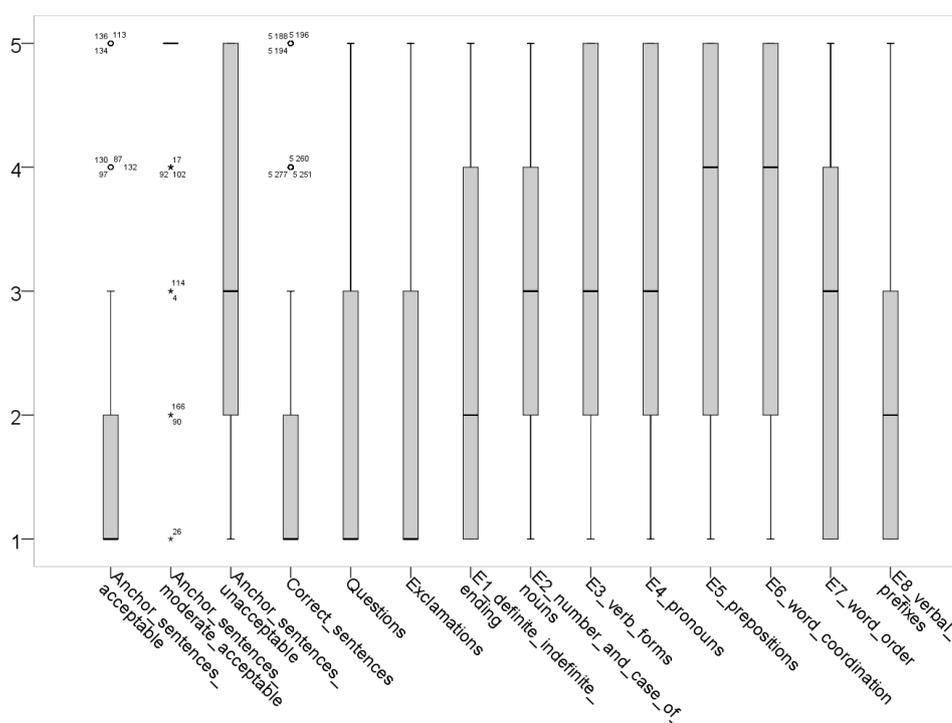


Fig. 6. General trends in evaluation of different sentence types (Table 1) that were tested in our study.

9. Discussion

Our current study is to our knowledge the first exploring Latvian from the point of view of acceptability. We have not observed any impacts of demographic factors which might indicate that the standard Latvian with its inherent conception of grammaticality is

dominating even in very different areas of colloquial dialects. The same concerns age of subjects: no significant impacts can be observed. (Initially we were hypothesizing that in virtue of the impact of English on the younger generation Latvian speakers or Russian on the older generation Latvian speakers, there might be some age differences in rating different types of grammaticality; according to our results, no significant effects are attested.)

The Figure Nr. 7 presents the overall picture of the results. They indicate that several errors are close to being fully acceptable (e.g., errors in usage of verbal prefix, definite / indefinite endings and errors in word order).

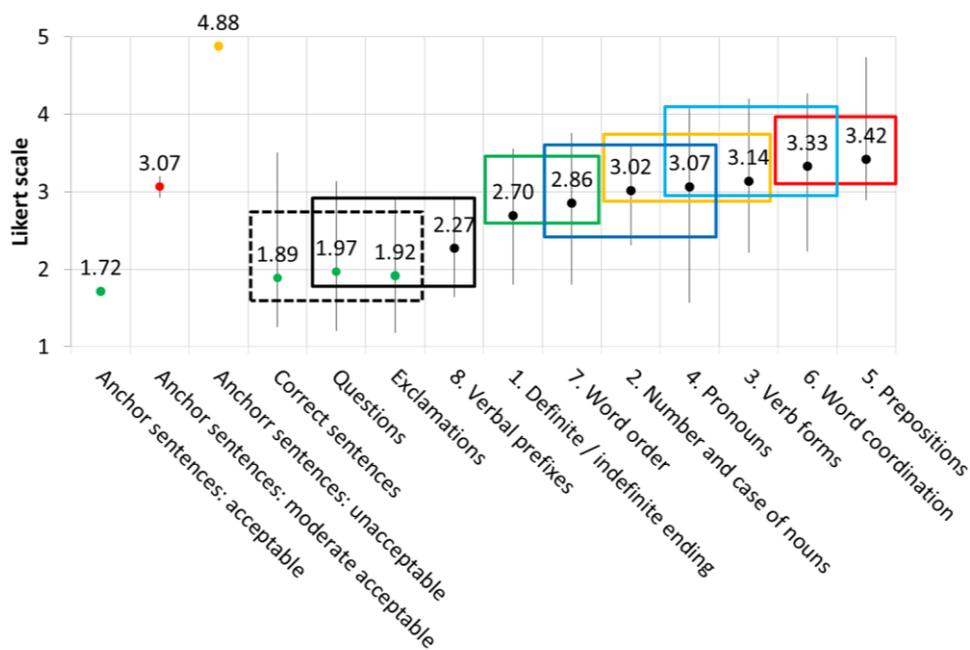


Fig. 7. Groups of sentence types that reflect significant differences ($\alpha=0.05$) between average evaluations of different types of errors.

Verbal prefixes in Latvian are the main means for expressing aspectual meaning, namely, perfectiveness. But only semantically bleached verbal prefixes may have the sole aspectual function (e.g., as prefix *pa-* in *darīt* : *padarīt* ‘do’, *uz-* in *rakstīt* : *uzrakstīt* ‘write’ or *no-* in *griezt* : *nogriezt* ‘cut’) as in the vast majority of cases verbal prefixes are derivational morphemes which functions range from slight modification of the verbal meaning (e.g., *lūkoties* ‘look at’ : *palūkoties* ‘take a look’) to ascribing a new meaning (cp.: *stāvēt* ‘stand’ : *aizstāvēt* ‘protect’, *piestāvēt* ‘fit’, *pastāvēt* ‘exist’, Nītiņa, 2001, 91). Thus, the relation between the verbal forms in derivational pair is primarily lexical the prefixed verbs being neutral in respect to aspect or biaspectual (Holvoet 2001, 132–136,

idem). The derivative function of verbal prefixes and their diverse semantic scope offer an explanation of high acceptability of erroneous or doubtful usage of prefixed verbs. In this group of errors, the sentences that were rated as most unacceptable (mean Likert scale score 2,9) are the cases of semantic incompatibility: either verbal prefix is needless as it adds the meaning of perfective aspect to verb with atelic interpretation (1) or the meaning of the prefix is not suitable in given context (2):

- (1) **Iz-analizēj-ot zinātnisk-o literatūr-u, darb-a*
 PVB-analyse-CVB scientific-ACC.SG.F.DEF literature-ACC.SG work-GEN.SG
autor-e no-šķir kopīg-u Lieldien-u
 author-NOM.SG PVB-separate.PRS.3 common-ACC.SG.F Easter-GEN.PL
motīv-u virkn-i vis-ai kriev-u literatūr-ai.
 motive-ACC.SG row-ACC.SG whole-DAT.SG.F Russian-GEN.PL literature-DAT.SG
 ‘[After analysing > analysing] scientific literature the author of the work distinguishes the commonseries of Easter motifs for all Russian literature.’
- (2) *Viņ-š vēl-as literatūr-u no-dalī-t jaun-aj-ā*
 3-NOM.SG.M wish.PRS-3.RFL literature-ACC.SG PRV-separate-INF new-DEF-LOC.SG.F
un jau pazīst-am-aj-ā.
 and already know-PRS.PP-DEF-LOC.SG.F
 ‘He wants to [separate > divide] the literature into the new and already known.’

The verb *nodalīt* ‘separate’ (: *dalīt* ‘divide, distribute’) has the meaning of separation of one part from the whole and usually requires different syntactic structure (*nodalīt* [*no* + Gen.] ‘separate from’), thus in the sentence (2) the verb with another prefix, viz. *iedalīt* ‘divide’, would be preferable for two reasons: a) it means the division into two groups and b) it is compatible with the Locative argument of the verb.

Some of the errors of prefixed verbs were the matter of prescriptive character. For example, even though *pielietot* ‘apply’ is suggested to substitute with *lietot* ‘use’ or *izmantot* ‘use, exploit’ (*pielietot metodes* > *lietot metodes* ‘use methods’) as the prefix *pa-* does not add any meaning in this case, native speakers considered *pielietot* rather acceptable.

High tolerance for errors of definite / indefinite ending of modifiers in a noun phrase may be due to their frequency in colloquial Latvian. The utterance that was rated as most unacceptable (mean Likert scale score 3,6) provides the clearest context for the usage of definite ending as the adjective is preceded by demonstrative (3, **šie atsevišķi uzdevumi* > *šie atsevišķie uzdevumi* ‘these separate tasks’). The second most unacceptable case requires the definite participle due to semantics of the verb (4, **izmantotas lietratūras* > *izmantotās literatūras*).

- (3) **Š-ie atsevišķ-i apakšnoz-ar-es uzdevum-i var*
 DEM-NOM.PL.M separate-NOM.PL.M subsector-GEN.SG task-NOM.PL can.PRS.3
būt sauk-t-i par tautsaimniecīb-as pakalpoj-um-iem.
 be-INF call-PST.PP-NOM.PL.M about economy-GEN.SG service-DAT.PL
 ‘These individual tasks of the sub-sector can be called economic services.’
- (4) **Pielikum-i tiek ievieto-t-i pēc*
 appendix-NOM.PL arrive.PRS.3 insert-PST.PP-NOM.PL.M after

izmanto-t-as *literatūr-as* *un* *avot-u* *sarakst-a.*
 use-PST.PP-GEN.SG.F literature-GEN.SG and source-GEN.PL list-GEN.SG
 ‘Appendices are inserted after the list of used literature and sources.’

Even though the predominant and unmarked word order in Latvian is SVO for transitive sentence, AdjN, GenN, AdvAdj for binary units and Latvian has both prepositions and postpositions, the sequence of words is relatively free and may vary according to information structure. This explains the high acceptability of irregular word order.

Errors in number and case show that the former are judged more critically. In the input examples the *plurale tantum* (La. *vēlēšanas* [NOM.PL] ‘election’) is used in singular and vice versa (La. *vienotība* [NOM.SG] ‘unity’) and is hardly acceptable by native speakers. On the contrary, the case errors are more tolerated as such error types occur frequently in every day language. Case errors include accusative-marked noun object in debitive construction instead of nominative-marked object (5, **jāfokusē uzmanību* > *jāfokusē uzmanība* ‘need to focus attention’), adverb *cik* ‘how many’ governing the nominative instead of the genitive (6, **cik projekti* > *cik projektu* ‘how many projects’) and *nav* [BE.3.NEG] usage with the nominative instead of the genitive marking of negated object in the negative possessive construction (7, **nav atbilstība* > *nav atbilstības* ‘there is no compliance’):

(5) **Ir* *jā-fokusē* *uzmanīb-u* *uz*
 be.PRS.3 DEB-focus attention-ACC.SG to
trūkum-u *un problēm-u* *konstruktīv-iem* *risinājum-iem.*
 defect-GEN.PL and problem-GEN.PL constructive-DAT.PL.M solution-DAT.PL
 ‘We must focus attention on constructive solution of the defects and problems.’

(6) **Jā-informē,* *cik* *projekt-i* *varē-tu* *būt* *ar*
 DEB-inform how_many project-NOM.PL can-IRR be.INF with
termiņ-u *līdz* *2015. gad-am.*
 deadline-ACC.SG till 2015 year-DAT.SG
 ‘It should be informed how many projects there could be with the deadline until the year 2015.’

(7) **Ir* *student-i,* *kur-iem* *nav* *atbilstīb-a*
 be.PRS.3 student-NOM.PL which-DAT.PL.M NEG.be.PRS.3 compliance-NOM.SG
vien-ai *no* *prasīb-ām.*
 one-DAT.SG.F from requirement-DAT.PL
 ‘There are students who do not meet one of the requirements.’

The group of pronoun errors consists of instances with erroneous usage of demonstratives, reflexive, possessive and relative pronouns as well as the anaphoric usage of personal pronouns for inanimate referents. In the latter case demonstratives *tas* [NOM.SG.M], *tā* [NOM.SG.F] should be used, although this rule often remains unfollowed in colloquial Latvian. This explains why such sentences are quite acceptable within this group of errors. The sentence that was rated as the most unacceptable (mean Likert scale score 4,1) exhibits misuse of reflexive pronoun (8) and is followed by the instance with demonstrative *tāds* ‘such (that)’ which is used instead of *šāds* ‘such (this)’ (9). (8) is ungrammatical as it lacks verbal argument indicating a place as a goal of the direction.

- (8) **Direktor-s* *ir* *ie-gāj-is* *pie* *sev-is*.
 director-NOM.SG be.PRS.3 PRV-go.PST.PA-NOM.SG.M at RFL-GEN.SG
 ‘The director has entered himself.’
- (9) **Ir* *atklā-t-i* *daž-i* *mehānism-i*
 be.PRS.3 discover-PST.PP-NOM.PL.M some-NOM.PL.M mechanism-NO
un *tād-u* *reklām-as* *kategorij-as* *tekst-u*
 and dem-GEN.PL advertising-GEN.SG category-GEN.SG text-GEN.PL
iedarbīb-as modeļ-i.
 effect-GEN.SG model-NOM.PL
 ‘Some mechanisms and models of exposure of such texts of advertising category have been discovered.’

Verbal errors encompass confusion of indicative and imperative plural 2nd form, inappropriate usage of indeclinable participle and the usage of infinitive in subordinate clauses denoting condition or purpose. Confusion of imperative and indicative is very common in colloquial Latvian – this is also reflected in our results as this type of verbal errors exhibits the highest tolerance of native speakers. The sentence (10) is rated as the most unacceptable in the group of verbal errors (mean Likert scale score 4,2): infinitive in impersonal subordinate conditional or purposive clauses in Latvian originates from Russian.

- (10) **Tagad* *mēģin-u* *sapras-t* *liet-as* *attiecīb-ā* *uz*
 now try-PRS.1 understand-INF thing-ACC.PL relation-LOC.SG to
iepirkum-u *noris-es* *pārbaud-ēm, lai* *ne-veik-t*
 purchase-GEN.PL procedure-GEN.SG checking-DAT.PL COMPL NEG.carry_out.INF
t-o *paš-iem*.
 DEM-ACC.SG.F self-DAT.PL.M
 ‘Now I am trying to understand the things related to checking of purchase procedure in order not to do that ourselves.’

Errors in word coordination (the cases of agreement and government) are very close to the most unacceptable instances. Agreement errors usually involve incorrect gender (and sometimes number) of the modifier in the noun phrase or incorrect gender of the passive participle in periphrastic passive construction in which the participle should agree with the subject in gender, number and case (as in (11): **politika* [policy.NOM.SG.F] *ir* [be.PRS.3] *vērsti* [focus:PST.PP.NOM.PL.M] > *politika ir vērsta* [focus:PST.PP.NOM.SG.F] [...]) ‘the policy is focused on [...]’). Agreement errors are considered less grammatical than instances of incorrect prepositional governing (12, **par nodrošināšanai* [about provision:DAT.SG] > *par nodrošināšanu* [about provision:ACC.SG]).

- (11) **Stratēģisk-ā* *industriāl-ā* *politik-a* *ir*
 strategic-NOM.SG.F.DEF industrial-NOM.SG.F.DEF policy-NOM.SG be.PRS.3
vērs-t-i *uz* *stratēģisk-iem* *sektor-iem*
 focus-PST.PP-NOM.PL.M to strategic-DAT.PL.M sector-DAT.PL
 ‘Strategic industrial policy focuses on strategic sectors.’

- (12) **Ne vienmēr vecāk-iem un skolotāj-iem ir skaidr-s, kā tiks risinā-t-s jautāj-um-s par mācīb-u līdzek-ļ-u nodrošināšan-ai.*
 not always parents-DAT.PL and teachers-DAT.PL be.PRS.3
 clear-NOM.SG.M how arrive-FUT.3 solve-PP.PST-NOM.SG.M question-NOM.SG
 about education-GEN.PL tool-GEN.PL provision-DAT.SG
 ‘It is not always clear to parents and teachers how the question of teaching tools will be solved.’

Finally, we can observe a clear case of unacceptable construction in native speaker ratings – prepositional errors. This group deals with several errors of adpositional use: the word order of adpositional phrase (prepositional use of postposition), the wrong choice of preposition (13), prepositional phrase instead of certain case form (e.g., 14).

- (13) **Prec-es vienmēr savlaicīg-i piegādāj-a uz pasūtīj-um-u.*
 good-ACC.PL always on_time-ADV deliver-PST.3 to order-ACC.SG
 ‘The goods were always delivered on time according to the order.’

- (14) **Par seš-iem gad-iem realizē-t-o māj-u projekt-u skait-s palielināj-ās septiņ-as reiz-es.*
 about six-DAT.PL.M year-DAT.PL carry_out-PST.PP-GEN.PL.M house-GEN.PL
 project-GEN.PL amount-NOM.SG increase-PST.3.RFL seven-ACC.PL.F
 time-ACC.PL
 ‘[About six years > in six years] the number of realised projects of houses increased seven times.’

In (13) another preposition, namely *pēc pasūtījuma* ‘according to the order’, should be used and in (14) the prepositional phrase should be replaced by the Locative of time *sešos gados* [six:LOC.PL.M year:LOC.PL].

Hence, adpositional structures are the most sensitive ones in interpreting a natural language sentence in Latvian. Worth mentioning that the stimuli sentences were not only spatial but also abstract or temporal. Interestingly, the sentence that was rated as most unacceptable (15, mean Likert scale score 4,7) has erroneous prepositional use of spatial meaning as instead of the prepositional phrase **uz piektā stāva* the Locative *piektajā stāvā* ‘on the fifth floor’ should be used.

- (15) **Liel-ais un jaun-ais dzīvokl-is bija uz piekt-ā stāv-a.*
 big-NOM.SG.M.DEF and new-NOM.SG.M.DEF flat-NOM.SG
 be.PST.3 on fifth-GEN.SG.M.DEF floor-GEN.SG
 ‘The large and new apartment was on the fifth floor.’

As already observed, several types of errors are close to being acceptable. We could tentatively assume that this kind of errors is frequently reflected in colloquial Latvian use and will in future disappear as an error (except perhaps in prescriptive grammars).

Although different non-grammatical sentences are rated differently, a more general (and independent on Latvian in particular) reason why certain grammatical insufficiencies are tolerated might be in virtue of a two-stage processing principle: the semantics is understandable before or independently of understanding the complete syntax of the sentence. This might be a consistent explanation with the model by Townsend and Bever (2001). Moreover, this might also explain why prepositional usage in Latvian is measured most critically. Errors of the usage of prepositions (wrong prepositional choice or usage of preposition instead of the locative case) alter the meaning of the phrase or sentence more than other error types resulting in more effort for first-stage processing, viz., understanding the semantics of the input sentence.

The tolerance to ungrammatical information might also be a result of a multilevel processing where in case of rating only some of the levels are entirely involved (Christiansen and Chater, 2016): e.g., it might be the case that incremental interpretation is completed, but multiple level structure is not completely generated.

Our study has also certain limitations: we were not controlling the frequency of constructions in the language, semantic and pragmatic properties of the sentences, emotional connotations (although we were avoiding emotionally extremely valences words or words denoting abstract objects). However, according to previous canonical work, we assume that the general response patterns to sentence types are reliable (Cowart, 1997, Schütze, 1996/2016).

Further, a separate study with a factorial design could be conducted to explore some of the sentence types in a more detail (e.g., different degrees of case coordination between adjective(s) and noun(s)) (Cowart, 1997).

Abbreviations

ACC – accusative
ADV – adverb
COMPL – complementizer
CVB – converb
DAT – dative
DEB – debitive
DEF – definite
DEM – demonstrative
F – feminine
GEN – genitive
INF – infinitive
IRR – irrealis
LOC – locative
M – masculine
NEG – negation
NOM – nominative

PA – active participle
 PL – plural
 PP – passive participle
 PRS – present
 PST – past
 PVB – preverb
 RFL – reflexive
 SG – singular

Acknowledgements

The research has been supported by the European Regional Development Fund within the project “Neural Network Modelling for Inflected Natural Languages” No. 1.1.1.1/16/A/215.

References

- Allen, J., Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.), *The emergence of language* (pp. 115–151). Mahwah: Lawrence Erlbaum.
- Bresnan, J. (2007). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In S. Featherston, W. Sternefeld (Eds.), *Roots: Linguistics in search of its evidential base* (pp. 75–96). Berlin: Walter de Gruyter.
- Christiansen, M. H., Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39.
- Cowart, W. (1997). *Experimental syntax: applying objective methods to sentence judgements*. Thousand Oaks: SAGE.
- Deksne, D., Skadina, I. (2014). Error-Annotated Corpus of Latvian. In Baltic HLT, 163–166 (doi:10.3233/978-1-61499-442-8-163).
- Divjak, D. (2017). The Role of Lexical Frequency in the Acceptability of Syntactic Variants: Evidence from that-Clauses in Polish. *Cognitive science*, 41(2), 354–382.
- Hofmeister, P., Sag, I. (2010). Cognitive constraints and island effects. *Brain and Language*, 86(2), 366–415.
- Holvoet, A. (2001). *Studies in the Latvian Verb*. Kraków: Wydawnictwo uniwersytetu Jagiellońskiego.
- Lau, J. H., Clark, A., Lappin, S. (2016). Grammaticality, acceptability, and probability: a probabilistic view of linguistic knowledge. *Cognitive Science*, 1–40.
- Lewis, S., Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research*, 44(1), 27–46.
- Manning, C. D., Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: The MIT press.
- Nītiņa, D. (2001). *Latviešu valodas morfoloģija. (Morphology of Latvian language)*. Rīga: Rīgas Tehniskā universitāte.
- Phillips, C. (2013). Some arguments and nonarguments for reductionist accounts of syntactic phenomena. *Language and Cognitive Processes*, 28(1–2), 156–187.

- Phillips, C., Lewis, S. (2013). Derivational order in syntax: Evidence and architectural consequences. *Studies in Linguistics*, 6, 11–47.
- Schütze, C. T. (1996/2016). *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*. 2nd edition. Berlin: Language Science Press.
- Schütze, C. T. (2009). Web searches should supplement judgements, not supplant them. *Zeitschrift für Sprachwissenschaft*, 28(1), 151–156.
- Schütze, C. T., Sprouse, J. (2013). Judgement data. In R. J. Podesva, D. Sharma (Eds.), *Research methods in linguistics* (pp. 27–50). Cambridge: Cambridge University Press.
- Sprouse, J. (2008). The differential sensitivity of acceptability judgments to processing effects. *Linguistic Inquiry*, 39(4), 686–694.
- Sprouse, J. (2011). A test of the cognitive assumptions of magnitude estimation: Commutativity does not hold for acceptability judgments. *Language*, 87(2), 274–288.

Received April 24, 2018, accepted May 29, 2018