

Processing of ambiguity in Russian: Form, function and frequency

Word forms, i.e. the surface forms of words that occur in natural language, do not map uniquely to the functions they fulfill or the properties they represent. Word forms may exhibit ambiguity of form where a single form may represent more than one set of properties, e.g. Russian *zakon* 'law.NOM=ACC.SG' can be nominative or accusative depending on context. Word forms may also exhibit ambiguity of function, e.g. *zakona* 'law.GEN.SG' may either express the function of possession or partitivity (among other things) depending on context. Form and function ambiguities are interconnected; a form that may represent more than one case can represent the set of functions associated with those cases. Previous research suggests that form and function ambiguities have an effect on processing (e.g. Clahsen et al. 2001; Verissimo & Clahsen 2009; Katz, Rexer & Mira 1995; Kostić 1991). However, different research programs have come to opposite conclusions about whether such ambiguities facilitate or inhibit processing. In this paper I investigate the effects of form and function ambiguities based on two processing experiments with Russian speakers. By having participants respond to the same set of stimuli with and without context, I control for each ambiguity to the extent possible. My results suggest that ambiguity of function inhibits processing when controlling for form ambiguities, but that form ambiguities do not have an effect after accounting for form frequency. This suggests that at least some form ambiguity effects found previously may be due to their frequencies and their relationship with ambiguities of function.

Ambiguity of form is when a language has a single form where two might be expected. In Russian, some nouns have distinct forms for the nominative and accusative, e.g. *kniga* ~ *knigu* 'book.NOM ~ book.ACC', whereas others do not *zakon* 'law.NOM.SG=ACC.SG'. In such instances of syncretism, the same form represents more than one case/number combination. Clahsen and colleagues (e.g. 2001; 2009) suggests that ambiguity of such features leads to faster reaction times in experimental tasks. The core idea is that forms that represent more than one case/number combination must be underspecified to be compatible with the properties they represent. Even though the accusative in Russian is [+directional] (based on Jakobson (1984)), *zakon* would be underspecified for directionality so that the form is compatible with the (lack of) features of both the nominative and accusative. Fewer features encoded in a lexical entry leads to faster processing. For Clahsen and colleagues, ambiguity of form is facilitatory in processing.

Another type of ambiguity is based on the semantic role/syntactic function of a form. For example, the genitive case can function to express possession and partitivity, among other things. Kostić and colleagues (e.g. 1991; 1995) propose that functions of word forms and their frequencies determine how quickly they are processed. They show that forms with higher 'information content', calculated based on the average frequency per semantic function of forms within the paradigm, take longer to process. For equally frequent forms, the form with more semantic functions will be processed slower. The core idea is that ambiguity requires additional processing because there are multiple functions that compete when a word form is being processed. For Kostić and colleagues, ambiguity of function is inhibitory in processing.

Ambiguity of function and form are necessarily interrelated. Syncretic forms, i.e. forms which represent more than one case, can fulfill more semantic roles than non-syncretic forms because each case has a variety of semantic roles it fulfills. Thus, it is not clear how ambiguity of function could inhibit processing whereas ambiguity of form which entails ambiguity of function could have the opposite effect. Furthermore, the contradictory findings are based on different tasks and in different languages making it all but impossible to make direct comparisons. This

raises questions about how the relationship between ambiguity of form and function interact and affect language processing.

To explore the relationship between form and function ambiguities in processing, two processing experiments were run with 154 native Russian participants. 20 masculine/Class I and 20 feminine/Class II lexemes were matched for length, neighborhood density, frequency, number of semantic meanings and morphological regularity. Stimuli were presented in accusative and dative case forms so that ambiguity of form existed in one class in the accusative (*zakon* ‘law.NOM=ACC’ vs. *knigu* ‘book.ACC’), and in the other class in the dative (*zakonu* ‘law.DAT’ vs. *knige* ‘book.DAT=LOC’). Forms were presented to speakers in two tasks: a lexical decision task (LDT) where forms were processed in isolation and a self-paced reading task (SPR) where forms were in a sentence context. In the self-paced reading task, stimuli were presented after the verb in one condition making them unambiguous for function, and before the verb in a second condition making their semantic function ambiguous. See example stimuli in Table 1.

Task	Stimulus (in bold)	Form ambiguity?	Function ambiguity?
LDT	вокзалу	no	yes
SPR	Утром к вокзалу мать везет дочь встречать друзей	no	yes
SPR	Утром мать везет дочь к вокзалу встречать друзей	no	no
LDT	победе	yes	yes
SPR	На самом деле это привело меня к победе в полуфинале	no	no
SPR	На самом деле к победе это привело меня в полуфинале	no	yes

Table 1. Example stimuli

Results were analyzed using mixed-effects linear regression models (using lme4 (Bates et al. 2015)). In models for both experiments, the dependent variable was reaction times with form frequency, class, case and their two-way interactions as fixed effects and random intercepts for item. Condition was also included as a fixed effect for the self-paced reading results. In the lexical decision task, only form frequency was a significant factor ($t = -6.14$, Std. Error = 0.025). Surprisingly, this was true even though forms in Class I had different reaction times when form frequency is not considered ($t(665) = -4.56$, $p < 0.001$). In the self-paced reading task, form frequency ($t = -3.72$, Std. Error = 0.016) and condition ($t=4.3$, Std. Error = 0.041) were significant. Stimuli that were presented before the verb were processed slower than stimuli presented after the verb.

The results of the self-paced reading task suggest that ambiguity of function is inhibitory during language processing, consistent with findings by Kostić and colleagues. The lexical decision task results based on the same word forms, however, show that when frequency is controlled for, ambiguity of form does not have a significant effect. This is somewhat surprising given that ambiguity of form entails ambiguity of function. In the lexical decision task, speakers did not necessarily have to determine the function of the form because it was processed in isolation which may explain the different results in each task. Ambiguity of form is closely connected to form frequency in lexical decision whereas ambiguity of function plays a greater role when the context requires speakers to disambiguate between multiple possible functions.