

Adversative conjunction choice in Russian (*no*, *da*, *odnako*): Semantic and syntactic influences on lexical selection

VSEVOLOD KAPATSINSKI

University of Oregon

ABSTRACT

This article presents a multivariate analysis of adversative conjunction choice (among *no*, *da*, and *odnako*) in Russian, drawing implications for sentence production and semantic theory. The two main factors shown to influence conjunction choice are the types of the conjoined constituents and the semantic subtype of the adversative relation. One of the conjunctions, *da*, is favored when the conjoined elements are of different syntactic types and disfavored when they are of the same type, which is argued to suggest that the conjunction is chosen at a point in sentence production when the types of both of the conjoined constituents are known (Uryson, 2006). In addition, the conjunction *da* is heavily favored by the “preventive” adversative meaning (Sannikov, 1989:177; Serebrjanaja, 1976), as in *I would go but I don’t have the money*. This quantitative meaning-construction association is argued to support the view that the preventive adversative is a distinct semantic subtype of adversativity (Payne, 1985; contra Foolen, 1991:84).

The aim of the present article is to identify factors influencing the choice of the adversative conjunction (*da*, *no*, and *odnako*) in Russian using multivariate analysis of corpus data from the Ogonek Corpus, a collection of magazine articles and interviews written between 1996 and 2000 (Berger, Betsch, & Bremer, 2001). Following Poplack and Tagliamonte (1996) and Torres Cacoullos and Schwenter (2008), I make the case for using quantitative variationist methods to study semantic differences between constructions. Implications for semantics, productivity, and sentence processing are drawn by examining the consequences of the conjunctions’ patterns of distribution for perception and production. Although previous work has attempted to determine “the degree to which each [variant] may appropriately be characterized as marking a given function” (Poplack & Tagliamonte, 1996:72), the aim of the present article is to connect the results of the quantitative analysis to both the extent to which a particular form signals a particular meaning/function for the listener and the extent to which a speaker is likely to use the form to express the meaning/function in question, both of which are essential to “reveal the true relationship between form and function,” as noted by Poplack and Tagliamonte

Many thanks to four anonymous reviewers for helpful feedback. Part of this work was presented at NWA V (New Ways of Analyzing Variation) 33, Ann Arbor, Michigan, and Slavic Linguistics Society 1, Bloomington, Indiana.

(1996:72) as well as to make predictions about form productivity (Schwenter & Torres Cacoullos, 2008). Because semantic factors clearly play a role in conjunction choice in Russian and because the use of the variationist method in cases when such differences do or may exist has been controversial (Cheshire, 1987; Lavandera, 1978; Sankoff, 1988), I review the literature on the semantics of adversative conjunctions and justify the application of variationist methods to the study of semantic influences on syntactic variation in the following section.

The semantics of adversatives

Lakoff (1971) identified two basic meanings for the adversative conjunction *but* in English: the adversative of semantic opposition illustrated by (1) and the denial of expectation adversative illustrated by (2). Unlike in (1), the second conjunct of (2) contradicts certain assumptions that the listener may be expected to make based on the first conjunct. Thus, one might assume that John is good at basketball when hearing that John is tall, an assumption that is contradicted by the conjunct following *but*.

- (1) John has a yacht, but Bill is deep in debt. (Russian: *a*)
- (2) John is tall but he's not good at basketball. (Russian: *no/da/odnako*)

Payne (1985:8–9) identified a third basic meaning of the adversative conjunction, the preventive adversative where the event described by the second conjunct prevents the event described by the first conjunct from taking place (or running its course to completion), as in (3) and (4). In (3), the agent's leaving is prevented by the detention. In (4), the going is prevented by lack of money. Even though Malchukov (2004:182), Payne (1985), and Uryson (2006) considered the preventive to be a separate meaning, Foolen (1991) disagreed, considering it reducible to one of the other meanings.¹

- (3) He wanted to leave but was detained. (Russian: *no/da/odnako*)
- (4) I would go but I don't have the money. (Russian: *no/da/odnako*)

A classic argument for distinguishing two meanings is the existence of a language or languages in which the two meanings are represented by different lexical items (e.g., Malchukov 2004:182). For instance, *but* used in the semantic opposition sense cannot be rephrased using “although” without a change in meaning, as shown in (5) and (6). Similarly, the sentence in (1) cannot be translated into Russian using the adversative conjunctions *da*, *no*, or *odnako*: the conjunction *a* is used instead (Lang, 1984). If the semantic opposition sense and the denial of expectation sense are represented by different lexical items in a certain language, then one can argue that the two senses must be distinguished by the speakers of the language.² Thus, the existence of *a* and *although* supports the hypothesis that the distinction between semantic opposition and denial of expectation is psychologically real for speakers of English and Russian.

The existence of the semantic distinction in Russian or English does not mean that the distinction is also made by speakers of other languages but rather that the distinction is strong enough to be acquired and maintained by speakers of a human language.

- (5) ?Although John has a yacht, Bill is deep in debt.
- (6) Although John is tall, he is no good at basketball.

In the present article, we consider the choice between the adversative conjunctions *da*, *no*, and *odnako* in Russian. These conjunctions can all be used to signal denial of expectation as well as the preventive. Previous researchers disagree on whether there is a distinction between the denial of expectation meaning and the preventive meaning (cf. Foolen, 1991; Payne, 1985). Because *da*, *no*, and *odnako* can carry both meanings, qualitative analysis would consider them irrelevant to this debate. However, the present study shows that *da* has a strong quantitative association with the preventive meaning (as hypothesized by Sannikov, 1989:177; and Serebrjanaja, 1976), supporting Payne's (1985) hypothesis of the preventive/denial of expectation distinction (see also Sannikov, 1989; Uryson, 2006, for Russian). Although quantitative evidence has not been typically used to argue for the psychological reality of semantic distinctions, categorical form-meaning associations are not required by this argument. If we find that speakers of some language are able to use different forms at different rates depending on the intended semantic function, they must distinguish the semantic functions in question in the course of language production. Therefore, quantitative analysis of lexical choice can increase the set of data constraining semantic theory. As Givón (1979:26–31) pointed out, grammatical theory should pay attention to gradient phenomena because categorical rules and gradient tendencies appear to have the same motivations and go in the same direction, so that a structure that is prohibited in some languages is likely to be statistically underrepresented even in languages where it is allowed. For instance, in many languages, the subject of a declarative clause cannot be referential-indefinite (as in *A man in the yard is asking for you.*), but the same tendency is present even in languages in which it is not obeyed categorically (Givón, 1979:26–28).

Aside from stylistic influences, there is no agreement as to why Russian speakers would choose one adversative conjunction over the others. The definitive Russian Academy Grammar (Shvedova, Arutjunova, Bondarko, Ivanov, Lopatin, Uluxanov, Filin, & Institute for Russian Language, 1980:621–625) gives no indication of semantic differences between the conjunctions, noting only that *da* is characteristic of colloquial speech. Shuba, Germanovich, Dolbik, Karaban', Maxon', Pipchenko, Rabchinskaja, Koxno, Tarasevich, Tixomirova, Fedotova, Pleshchenko, Chechet, and Shevchenko (1998) maintained that *da* and *no* are synonymous, not mentioning *odnako* at all. Lekant, Gol'cova, Zhukov, Kasatkin, Klobukov, Malashenko, Tuzova, Fomenko, and Xolodov (1982) wrote that *odnako* has a "more intense" meaning of contradiction than *no*, which is in the middle, and *da*, which has a "softened"

one. Kruchinina (1988) identified *da* with the denial of expectation adversative (Lakoff, 1971): by “VP_X *da* VP_Y,” the speaker is taken to say “given X you might expect that *not* Y, nonetheless Y.” Finally, Serebrjanaja (1976) and Sannikov (1989:178) identified *da* with the preventive adversative, where the event denoted by X fails to run its course due to the event denoted by Y. This variability in proposed semantics suggests that *da*, *no*, and *odnako* are semantically very close and quantitative analysis may be necessary to determine what semantic differences if any hold between them.

In the present study, I use the quantitative variationist method to examine the choice between the conjunctions. Thus I attempt to predict conjunction choice by treating the set of competing conjunctions as the dependent variable and the presence/absence of particular contextual elements as independent variables in a logistic regression analysis. The application of the quantitative variationist method to lexical and syntactic variation has been controversial (e.g., Cheshire, 1987; Lavandera, 1978). Much of the criticism has been leveled against the notion of the linguistic variable as applied to levels above morphophonology. Labov (1966) originally defined the linguistic variable as a set of variants that serve as different ways of “saying the same thing.” Because it is almost always possible to come up with a semantic difference between two constructions, the applicability of the variationist method for the study of lexical and syntactic variation has been disputed (e.g., Cheshire, 1987). Sankoff (1988:153) defended the practice by pointing out that “While it is indisputable that some difference in connotation may, *upon reflection*, be postulated among so-called synonyms whether in isolation or in context ... there is no reason to expect these differences to be pertinent every time one of the variant forms is used” (emphasis in the original). I would like to argue that even if, contrary to Sankoff’s neutralization in discourse hypothesis, speakers are always aware of semantic distinctions when making lexical choices, the putative existence of a semantic distinction between two variants should not prevent the analyst from using multivariate analysis to study the choice between them, although the set of forms studied would not constitute variants of a linguistic variable in the classic sense (Poplack & Tagliamonte, 1996:78, fn. 9).

Even if there is a semantic difference between two variants and the speaker chooses the variant based *exclusively* on the semantics, the context of use may provide information about the semantics of the chosen variant. Therefore, the coding of contextual elements with which a given variant co-occurs provides an objective (and replicable) way of coding the semantic distinction.³

In the present study, I argue that there is a semantic difference between the conjunctions *da* and *no*, such that *da* is strongly associated with the preventive adversative meaning, that is, X *da* Y is much more likely to mean that Y prevented X from going to completion than X *no* Y. This semantic difference is often reflected in the context surrounding the conjunction: if Y prevents X from going to completion, X is realized as an irrealis VP while Y is realized as a realis VP (Payne, 1985). Given that *da* co-occurs with the irrealis-realis construction much more than *no* does, we can conclude that *da* is associated

with the preventive meaning more than is *no*. To the extent that semantic distinctions can be coded in a reliable and replicable manner, the variationist method allows for investigation of semantic distinctions between constructions that would otherwise remain obscure (cf. Aaron & Torres Cacoullas, 2005; Poplack, 1992; Poplack & Tagliamonte, 1996, 1999; Torres Cacoullas & Schwenter, 2007, 2008).

It is important to note that although the contextual reflections of semantic distinctions are coded as independent variables influencing variant choice in the multivariate analysis, they do not have to and perhaps should not be interpreted as such. Thus, even though the irrealis-realism construction is a factor group in the multivariate analysis, it does not have to be interpreted as an independent factor influencing conjunction choice. Upon finding a statistically significant co-occurrence of *da* and the irrealis-realism construction, we do not have to claim that the choice of the irrealis-realism construction precedes and influences the choice of the conjunction. It is quite possible that *da* and the irrealis-realism pattern are part of a single construction (in the sense of Construction Grammar, Goldberg, 1995, 2006) and selected simultaneously based on the semantics that the speaker wishes to express. For instance, Torres Cacoullas and Schwenter (2008:1470–1471) found that the choice of the Spanish marker *se* after verbs *subir* ‘get on’ and *bajar* ‘get off’ is favored when the following prepositional phrase refers to a vehicle and interpret *se* as being a part of the “enter-exit a vehicle” construction of the form {*subir;bajar*} + *se* (50%–83%) + *alde* + ‘vehicle’.

The syntax of adversative conjunctions and sentence production

On the syntactic level, Payne (1985) distinguished between sentential and phrasal adversatives, the former conjoining clauses, and the latter phrases. He identifies (a) adversatives that can only conjoin clauses; (b) adversatives that can conjoin clauses, verb phrases, and adjectival phrases; and finally (c) adversatives that can also conjoin noun phrases and prepositional phrases, yielding the following implicational hierarchy: PP, NP → AP, VP → S. That is, if an adversative can conjoin prepositional phrases and noun phrases, then it can conjoin any of the other phrase types, and if it can conjoin APs or VPs it can conjoin sentences. Payne (1985) proposed that the Russian *a* and *no* are restricted to conjoining APs, VPs, and Ss. However, an interesting feature of Russian adversative conjunction is that the conjoined elements do not have to be of the same type (Sannikov, 1989:20–23; Uryson, 2006:36), as shown in (7)–(11), which are drawn from the Ogonek corpus (Berger et al., 2001).

- (7) Otrezvljus', *da* pozdno.
'I will sober up but too late'.
- (8) Poka zemlja ešče vertitsja: slaboe, *da* utešenie.
'The Earth is still spinning: though weak, it is a reassurance'. (literally: weak, but reassurance)

- (9) Prognala b duraka, *no* nekuda.
 ‘[I] would have thrown the fool out but [there is] nowhere [to throw him out].’
- (10) Kot, *da* ne tot.
 ‘[It is] a cat, but not [the] right [one].’
- (11) Xot’ v redkie, xot’ v svjatye dni, *da* nužna.
 ‘Though [only] in rare, in holy days, but [she is] needed.’

The existence of examples similar to the ones in (7)–(11) presents an interesting opportunity for this corpus investigation of conjunction choice to shed light on the processes of sentence production. It is logically possible for the choice of the conjunction to be influenced only by the type of the preceding constituent, the type of the following constituent, or both. If the two constituents were always of the same type, there would be no way to distinguish between the three possibilities. The existence of asymmetric constructions like the ones in (7)–(11) allows us to investigate this question. In particular, if we find that the choice of the conjunction depends on whether the conjoined constituents are of the same type, we must conclude that the speaker must know the types of both of the conjoined constituents when the conjunction is chosen. Thus, as proposed by Uryson, (2006:35, fn. 10), “the coordinating conjunction is chosen and added to the already assembled construction containing homogeneous constituents at one of the latest stages” of sentence production. More specifically, it would not be possible for the conjunction to be chosen before the type of the second conjunct is known or after the type of the first conjunct becomes inaccessible to the lexical selection system and, if the conjuncts are assumed to be underlyingly of the same type, then the lexical selection of the conjunction would have to occur at a processing stage during which the conjuncts no longer belong to the same syntactic category.

METHODS

Sentences containing *da*, *no*, and *odnako* were obtained from the Ogonek Corpus, a 7,660,000 word collection of articles and interviews from a popular magazine by the same name written or recorded between 1996 and 2002 (Berger et al., 2001)*. The sample included all instances of adversative *da* in the corpus ($n = 238$), about 50% of instances of *odnako* and around 3% of instances of *no*. For *no* and *odnako*, the number of instances was limited to be similar to the number of *da* instances. The sampling was done by taking the first 250 instances satisfying the inclusion criteria that were returned by the corpus’s online concordance engine.

Each instance was coded for topicality of the second conjunct, register (interview vs. article), the irrealis-realis construction, the length of the second conjunct (in words), and the syntactic types of the conjuncts.

* The corpus is available at <http://heckel.sfb.uni-tuebingen.de/cgi-bin/cqp.pl?sprache=en&trans=lat>.

Topicality of the second conjunct has been suggested to disfavor the use of *odnako* (Koolemans Beynen, 1976). In the present study, topicality was operationalized as whether the referent of some noun within the second conjunct was referred to again within the next two sentences. If the second conjunct did not contain any nouns, that token was not coded for topicality. The length of the second conjunct and register can be thought of as imperfect reflections of style and were included because several researchers have remarked that *da* is colloquial (Krilova, 1980; Kruchinina, 1988; Shvedova et al., 1980), whereas *odnako* is formal and characteristic of written discourse (Lekant et al., 1982). The length variable was made binary by splitting the length continuum for each conjunct type (NP, AP, VP, clause) into long and short lengths at the mean length value. Thus long clauses are opposed to short clauses and long NPs to short NPs, avoiding the confound of length and syntactic category. The irrealis-realism construction is included as a reflection of preventive semantics (Payne, 1985).

If a given use of *da*, *no*, or *odnako* is preventive, then the event depicted by the first conjunct is prevented from running to completion by the event depicted by the second conjunct (cf. Sannikov, 1989:157, 177). The formal reflection of the preventive meaning in Russian is that the first conjunct is headed by a verb in the conditional (12), a verb in the past or present imperfective combined with an infinitive (13), or a perfective inchoative verb, with the inchoative prefix *po* (13), or the marker *bylo* (15), and the second conjunct is headed by a perfective (13) and (14), present, and/or future verb, which can be a phonologically null copula (12). If one of the conjuncts was not a VP or a clause, that token was not coded for this factor.⁴

- (12) Ja *by* dal s bol'shim udovol'stvиеm, *da* kurica ne moja.
'I *would* gladly give [it], but the hen is not mine'.
- (13) On *xotel uiti*, *no* ego zaderžali.
'He *wanted to leave* but was detained'.
- (14) *Polez* i Kostylin, *da* zacepil kamen' nogoj, zagremel. (Sannikov, 1989:177)
'Kostylin *started climbing* too but got caught on a stone with his leg and fell'.
- (15) On *bylo zagovoril*, no ego perebili. (Malchukov, 2004:182)
'He *started to speak* but was interrupted'.

Finally, the syntactic type of the two conjuncts was coded as NP, AP, PP, VP, or clause. After the data were analyzed, cases in which the conjuncts were both NPs, As, or PPs were grouped together into the "nominal" category based on similar behavior in respect to conjunction choice. Similarly, cases in which the conjuncts were both VPs or clauses were grouped together into the "verbal" category. The third category was formed by asymmetric constructions, in which the two conjuncts were not of the same type, for example, one was an NP and the other an A, as in (8). As this example indicates, the two conjuncts in an "asymmetric construction" could both be nominal as long as they were of different syntactic categories narrowly defined.

There were four types of exclusions. First, Russian has two *da* conjunctions, only one of which is adversative and can be replaced by *no* and *odnako*. In many cases, as in (16)–(18), *da* has the meaning of *and*, rather than *but*. Such uses of *da* are not the result of competition between adversative conjunctions investigated in the present study and thus were excluded.

- (16) Vokrug zabrošennyj sad, *smorodina da jabloni*.
‘Around [me] is an abandoned garden, *blackcurrant and apple trees*’.
- (17) Pokupajut kogda *sytno da spokojno*.
‘[They] buy when [it’s] *nice and quiet*’.
- (18) A ... kak *povesjat kolokola da udarjat v pervyj raz v samyj bolqshoj*—tak i nastupit Strašnyj Sud.
‘And when they *hang up the bells and hit the biggest one for the first time*, that’s when the Last Judgment will come’.

Second, non-conjunct-initial instances of *odnako* were excluded from the sample. Unlike *no* and *da*, *odnako* may occur after the subject or at the end of the second conjunct. Depending on one’s model of sentence production, non-conjunct-initial instances of *odnako* may or may not compete for selection with the obligatorily conjunct-initial *no* and *da*. For instance, it is possible that non-conjunct-initial cases of *odnako* are selected as an afterthought at a point in language production when the beginning of the second conjunct has already been produced or is about to be produced so that interrupting its production would result in disfluency. If non-conjunct-initial *odnako* is selected after the beginning of the second conjunct has been formulated and cannot be interfered with, it is the only option at that point in sentence production and does not compete with conjunct-initial *da* and *no*. Therefore, non-conjunct-initial cases of *odnako* were excluded from this analysis.

Finally, a number of fixed expressions and a partially lexically specific construction obligatorily selecting *da* were excluded from the analysis. Because the constructions in question categorically favor *da*, they cannot be included in the multivariate analysis. Aside from fixed expressions, such as *mal zolotnik da dorog* ‘the gold coin is small but precious’, these include one more general construction: what one may call the “was X but (all) ended” construction. Central instances of the construction feature a past imperfective form of *byt’* ‘to be’ as the head of the first conjunct and a past form of *splyt’* ‘float off’ or *vyjti* ‘go out’ as the head of the second conjunct but the construction has been extended to other verbs with the same meaning of going away, ending, and disappearing, as shown in (19)–(23).⁵

- (19) Byl, da ves’ vyšel.
Was, but all went out.
‘There was some, but it all evaporated’.
- (20) Byl veter, da splyl.
Was wind, but floated off.
‘There was wind, but no more’.

- (21) Byli koe-kakie sposobnosti, da vse vyšli.
Were some abilities, but all went out.
'[He] used to have some talent, but it has all been lost'.
- (22) Byli vremena bylinnye, da sgnili.
Were times legendary, but rotted away.
'There were legendary times, but they have passed'.
- (23) Byli tri dočeri, da umerli.
Were three daughters, but died.
'[I] had three daughters, but [they] died'.

The data were analyzed using GOLDVARB X (Sankoff, Tagliamonte, & Smith, 2005), a statistical program implementing step-up/step-down logistic regression analysis widely used in variationist research. The same factor groups were selected as significant on both the step-up and step-down runs.

RESULTS AND DISCUSSION

Because logistic regression assumes that the dependent variable is binary,⁶ and there are three possible conjunction choices, I conducted three separate analyses for the choice of *da* versus something else, *no* versus something else, and *odnako* versus something else.

The conjunction *no* is the most common adversative conjunction in Russian, our sample including only 3% of the instances of *no* present in the corpus. Table 1 shows the factor groups that were selected by the regression analysis as having a significant influence on the selection of *no*.

Unlike in the case of the less productive conjunctions, the choice of *no* was significantly affected by only two factor groups. This finding supports the proposal that the default/most productive expression is the one that is preferred in the least specified contexts (Schwenter & Torres Cacoullos, 2008). This proposal is also supported by connectionist simulations reported by Hare, Elman, and Daugherty (1995), who show that expressions that are distributed more or less evenly through the space of possible contexts (thus being less associated with any specific context) are more likely to be extended to novel contexts (i.e., display high productivity) than forms that tend to occur in a restricted range of contexts. The prediction of Hare et al. (1995) is supported empirically by Boudelaa and Gaskell (2002), who demonstrate that the high productivity of the suffixal plural in Arabic, which is unexpected based on its low type frequency, is predicted if evenness of distribution of the competing plurals across (phonological) contexts is taken into account.

Table 1 shows that *no* is the conjunction of choice for noun phrases and adjectives.⁹ This result appears to contradict Payne's (1985) statement that *no* cannot conjoin nominals. It should be noted that the nominals conjoined by *no* are postcopular complements and thus can be considered predicative. Nonetheless, they seem to behave differently than other predicates, which, unlike nominals, do not heavily favor *no* in the present sample. The preventive

TABLE 1. *Factors that significantly favor or disfavor the choice of no*

| Factor group | Range ⁷ | Value | Factor weight | Proportion of cases featuring <i>no</i> ⁸ | <i>n</i> |
|------------------------------|--------------------|------------|---------------|--|----------|
| Conjunct type | 46 | Nominal | .83 | 75 | 52 |
| | | Verbal | .49 | 32 | 579 |
| | | Asymmetric | .37 | 24 | 107 |
| Irrealis-realis construction | 40 | Yes | .18 | 7 | 111 |
| | | No | .58 | 36 | 521 |

Note: Factor groups not selected as significant: length of second conjunct, topicality of the second conjunct, register.

TABLE 2. *Factors that significantly favor or disfavor the choice of odnako*

| Factor group | Range (effect size) | Value | Factor weight | Proportion of cases featuring <i>odnako</i> | <i>n</i> |
|-----------------------------------|---------------------|------------|---------------|---|----------|
| Conjunct type | 39 | Nominal | .30 | 12 | 52 |
| | | Verbal | .58 | 41 | 579 |
| | | Asymmetric | .19 | 7 | 107 |
| Irrealis-realis construction | 25 | Yes | .30 | 16 | 111 |
| | | No | .55 | 43 | 521 |
| Length of second conjunct | 37 | Long | .70 | 54 | 401 |
| | | Short | .33 | 17 | 337 |
| Register | 32 | Interview | .23 | 10 | 106 |
| | | Article | .55 | 38 | 632 |
| Topicality of the second conjunct | 7 | Yes | .48 | 36 | 271 |
| | | No | .55 | 44 | 142 |

meaning is uncharacteristic for *no*, hence *no* is chosen only 7% of the time when the presence of the irrealis-realis construction indicates that the conjunction is used as a preventive.

The conjunction *odnako* is less common than *no*, and therefore our sample contains about 50% of conjunct-initial instances of this conjunction found in the corpus. Table 2 shows the factor groups that were selected by the regression analysis as having a significant influence on the selection of *odnako*. As suggested by Lekant et al. (1982), *odnako* appears to be favored by written discourse and is also associated with long conjuncts, which may be seen as another reflection of formality. In addition, *odnako* is favored by verbal conjuncts, verb phrases, and clauses. Finally, as suggested by Koolemans Beynen (1976), there is a tendency for *odnako* to introduce nontopical information.

The results for *da*, the rarest (but, perhaps, the most interesting) of the three conjunctions, are shown in Table 3. Only factor groups that were selected as significant by the regression analyses are shown. Unlike *no* and *odnako*, *da* tends to introduce short second conjuncts, which could be considered support

TABLE 3. *Factors that significantly favor or disfavor the choice of da*

| Factor group | Range (effect size) | Value | Factor weight | Proportion of cases featuring <i>da</i> | <i>n</i> |
|---------------------------------|------------------------|------------|------------------|--|----------|
| Conjunct type | 60 | Nominal | .23 | 13 | 52 |
| | | Verbal | .45 | 27 | 579 |
| | | Asymmetric | .83 | 69 | 107 |
| Irrealis-realis construction | 48 | Yes | .88 | 77 | 111 |
| | | No | .40 | 21 | 521 |
| Length of second conjunct | 41 | Long | .28 | 13 | 401 |
| | | Short | .69 | 48 | 337 |

Note: Factor groups not selected as significant: topicality of the second conjunct, register.

for *da* being colloquial (Krilova, 1980; Kruchinina, 1988; Shvedova et al., 1980). Although the influence of register is not significant, the range of registers in the present study is small, because only highly formal (and edited) interviews and magazine articles are included in the corpus. Unfortunately, spoken corpora of Russian are not available, making studying differences between speech and writing difficult.

Table 3 indicates that *da* is strongly favored by the irrealis-realis construction, which is a reflection of the preventive adversative meaning. Though we would expect 33% of instances of the irrealis-realis construction to contain *da* by chance, a full 77% of them do. This finding supports the claim that *da* is semantically distinct from the other adversative conjunctions, being much more strongly associated with the preventive meaning. In *X da Y*, *Y* is likely to prevent *X* from running to completion. The meaning of *da* discovered by multivariate analysis is in line with the proposal of Serebrjanaja (1976) and contradicts Kruchinina's (1988) hypothesis that *da* is more closely connected to the denial of expectation meaning than *no*. It also appears to contradict the statement of Lekant et al. (1982) that *da* indicates a "softer contradiction" than the other adversative conjunctions. The preventive adversative appears to be the strongest possible contradiction in that the event expressed by the second conjunct prevents the event expressed by the first conjunct from running its course.

Finally, the presence of associations between the preventive meaning (operationalized by the presence of the irrealis-realis construction) and *da* and the denial of expectation meaning (operationalized by the presence of a realis verb form as the head of the first conjunct) and *odnako* supports Payne's (1985) proposal that the two meanings are distinct, contra Foolen (1991). Psychologically speaking, speakers of Russian seem to categorize preventive and denial of expectation uses into different categories, hence the distinction between prevention and denial of expectation must be learnable.¹⁰ The fact that the semantic effect on conjunction choice is quantitative in nature may suggest that the boundary between the categories is blurry, possibly because the speaker is not always aware of the distinction while producing a sentence (Sankoff, 1988). On the other hand, it is also possible that the speaker is always aware of the

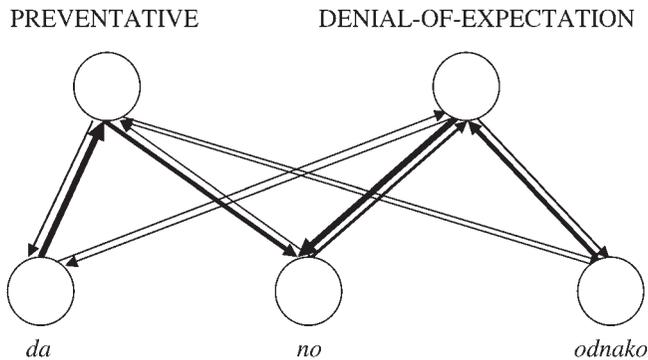


FIGURE 1. Conjunction choice in production (indicated by downward arrows) versus meaning assignment in perception (indicated by upward arrows). Line width indicates likelihood of following a particular arrow. Thus, given *da*, one is likely to arrive at the preventive meaning (the thickest arrow leading upward from *da* leads to the preventive) but given the preventive meaning, one is more likely to produce *no* than *da* as indicated by the thickness of the downward arrow from preventive to *no* relative to the arrow from preventive to *da*.

distinction but the semantic influence on conjunction choice is sometimes overruled by other factors.

At first glance, the fact that *da* is strongly associated with the preventive meaning suggests that *da* must be productive unless Russian speakers find that they never need to talk about situations in which one event prevents another event from running to completion. The literature, on the other hand, is unanimous in claiming that *da* is becoming unproductive (Krilova, 1980; Kruchinina, 1988; Shvedova et al., 1980). Closer inspection of the present data reveals that it is consistent with *da* being relatively unproductive.

Even though *da* is strongly associated with the preventive meaning and thus is a good cue to the preventive meaning for the listener,¹¹ it is highly infrequent. Therefore, even if a speaker needs to express the preventive meaning, *da* is much less likely to be chosen for production than *no*. Given the proportions of *da*, *no*, and *odnako* in the corpus, we can estimate that a sample of preventive adversatives drawn from the corpus randomly would contain 24% *da*, 66% *no*, and 10% *odnako*. Thus, most of the time when a speaker intends to produce a preventive, *no* would be chosen despite *da* being a better cue to the preventive meaning for the listener. Another case of such an asymmetry is presented by Poplack and Tagliamonte (1996:84) who found that one of the past tense markers in Nigerian Pidgin English (*bin*) is used as an anterior 90% of the time while having only a 5% chance of appearing in such a context. Zuraw (2000) presented a similar case of production-perception mismatch in morphophonology. The difference between perception and production is illustrated in Figure 1, where *no* is most likely to be selected for production with either of the intended meanings, resulting in neutralization in discourse (Sankoff, 1988), whereas in perception

da is a cue to the preventive meaning unlike *odnako* and *no*, which are (weaker) cues to the denial-of-expectation meaning.

Recent psycholinguistic evidence (Hudson Kam & Newport, 2005) indicates that children exposed to an artificial grammar in which one rule is obeyed 70% of the time and a competing rule is obeyed 30% of the time tend to overapply the more commonly used rule, using it 100% of the time with novel productions. Given this tendency of children to regularize unpredictable variation, *no* may gradually encroach on *da* as the preventive marker, accounting for statements that *da* is not fully productive in modern Russian (Krilova, 1980; Kruchinina, 1988; Shvedova et al., 1980), although these observations are based on written data, which may lead to underestimating the productivity of *da*, which is suggested to be colloquial (e.g., Shvedova et al., 1980). Ironically, the colloquial nature of *da* may hurt *da*'s chances for survival compared with the chances of *odnako*. Hudson Kam and Newport (2005) show that adults exposed to the same artificial language featuring random variation, unlike the children they studied, tend to reproduce the frequencies of variants found in the input. The conjunction *odnako*, while being almost as rare as *da*, is more formal and characteristic of written language, thus being likely to be learned later in life than *da* is.¹² If unpredictable variation is more likely to be regularized by children than by adults, *da* may lose ground to *no* faster than *odnako* would.

In addition to being favored by the preventive meaning, *da* is favored when the conjuncts are not of the same type. The instances in which the two conjuncts are not of the same type cannot be the same instances as the ones in which *da* is used in the preventive sense because the preventive by definition requires both conjuncts to refer to events and thus be headed by a verb (Payne, 1985). Thus, the high rate of use of *da* when the conjoined units are of different types cannot be due to the association between *da* and the preventive meaning and must be caused by an independent factor. Conjunction choice in Russian thus appears not to be driven exclusively by semantic differences.

The finding that *da* is favored when the two conjuncts are not of the same type has implications for sentence production. If NP conj AP differs in its influence on conjunction choice from both NP conj NP and AP conj AP, the former favoring *da* and the latter two favoring *no*, then the grammatical categories of both conjuncts must be available to speakers when the conjunction is chosen. Thus, speakers must choose the conjunction at a point when they have already decided on the two conjuncts and the information about the type of the preceding conjunct must be available to the lexical selection mechanism at the same time as the information about the type of the following conjunct. This suggests that either the syntactic frame of the sentence is chosen before the selection of (variable) function words or that such words are inseparable from the sentence frame, as argued by Construction Grammar (Goldberg, 1995, 2006). Furthermore, if the asymmetric constructions are assumed to arise from ellipsis (e.g., the sentence in (10) *Kot da ne tot*. 'A cat but not the right [one]'. deriving from *Kot da ne tot kot*. 'A cat but not the right cat'), conjunction choice cannot occur before ellipsis, because before ellipsis, asymmetric and symmetric constructions belong

to the same syntactic category. In the case of (10), both “a cat” and “not the right cat” are noun phrases in the classification adopted for the multivariate analysis. Thus, if conjunction choice happened at the stage that both are noun phrases or clauses, sentences such as (10) would be expected to behave like other sentences consisting of two conjoined noun phrases, favoring *no*.

If the asymmetric constructions are assumed to arise from ellipsis, *da* can be taken to be a marker of ellipsis, serving to alert the listener that some part of the following clause needs to be filled in based on the preceding clause, signaling ellipsis in 31% of its occurrences.¹³ However, it should again be borne in mind that *da* is much less frequent than *no*. Therefore, in a random sample of asymmetric constructions, *da* is expected to be chosen only 8.6% of the time with *no* accounting for 89.7% of tokens, whereas *odnako* almost never occurs in this context, being expected to account for only 1.7% of such instances. Thus, despite being associated with the asymmetric constructions and being potentially helpful to the listener as a marker of ellipsis, it may lose productivity as time goes on.

CONCLUSIONS

Adversative conjunction choice in Russian is influenced by semantic, syntactic, and stylistic factors. Syntactically, the conjunction *no* is favored when the conjuncts are postcopular complements (realized as noun, adjective, or prepositional phrases), *odnako* is favored when they are verb phrases or clauses, and *da* is favored when the conjuncts are of different syntactic types, such as a noun and an adjective. Semantically, the conjunction *da* is favored when the event described by the second conjunct prevents the event described by the first conjunct from running its course (Sannikov, 1989; Serebrjanaja, 1976). Stylistically, *odnako* is favored and *da* is disfavored by formal, written discourse (Krilova, 1980; Kruchinina, 1988; Lekant et al., 1982; Shvedova et al., 1980).

The present study illustrates that quantitative variationist methods can be used to uncover subtle semantic differences between constructions and to constrain semantic theory by providing evidence about the learnability of semantic distinctions. In addition, to draw inferences about differences in productivity between constructions in different contexts, it is important to interpret the results of multivariate analysis with reference to production and perception (something that has also been argued to be necessary for elucidation of relationships between meanings and forms by Poplack & Tagliamonte, 1996; and Schwenter & Torres Cacoullous, 2008). In particular, the presence of an association between a construction and a frequently expressed meaning may not ensure a construction’s survival if the meaning can also be expressed by a more frequent construction, albeit one that is not associated with the meaning as strongly. The most productive expression is shown to be the one that occurs in the least specified contexts, as proposed by Hare et al. (1995) and Schwenter and Torres Cacoullous (2008). Finally, corpus studies of syntactic variation show promise for

constraining psycholinguistic models of the time course of sentence production by specifying which types of information influence a given choice. As previously suggested by Uryson (2006:25, fn. 10), I argue that the choice of the conjunction is made late during sentence production when the types of the conjoined constituents are already known.

NOTES

1. Foolen (1991:83) also identified a third adversative meaning, the correction adversative, as in *The conference is not in Berlin but in Groningen*. As pointed out by Foolen (1991:87), this meaning is realized by *a* in Russian and thus falls outside the scope of this study (see also Malchukov, 2004:183). Though it is possible to think up examples with *no*, such as the fixed expression in (i), this use is clearly archaic and stylistically restricted. Furthermore, *da* and *odnako* cannot be used in this context.

- (i) slova ne mal'čika *no* muža
words not boy's but man's
'words of a man, not a boy'

2. Unless one can find some other factor that is correlated with the semantic distinction and is the real cause of lexical choice for speakers of the language.

3. For instance, Torres Cacoullos and Schwenter (2007) document that the development of *a pesar de* from a phrase meaning "to the regret of" into a concessive meaning *although* in Spanish was accompanied by a change in its patterns of co-occurrence with contextual elements and argue that "semantic change [in general] ... should be manifested in changing distribution and co-occurrence patterns" (Torres Cacoullos & Schwenter, 2007:357).

4. A reviewer notes that Sannikov (1989:157) considers the syntactic realizations of preventive *no* to be much more diverse. Curiously, all of Sannikov's examples of preventive *da* (Sannikov 1989:177) are covered by the irrealis-realis construction. I believe that the examples of "preventive *no*" that do not fit the irrealis-realis construction, two of which are shown in (ii) and (iii), are not preventive, and the event described in the verb phrase preceding the conjunction is not prevented from running to completion by the event described by the verb phrase following the conjunction.

- (ii) Zdes' Severin, *no* ja s nim possorilsja i ne klanjajus'.
'Mr. Severin is here but we fell out and I don't bow [to him anymore]'.
(iii) Kotoryj chas? U menja bez pjati dva, *no* moi speshat.
'What time is it? I have five to two but my [watch] is fast'.

5. Third, conjunction clusters such as *no odnako* and *da i* were excluded because the *no* and *da* appearing in such clusters may not constitute separate lexical items, the whole cluster being selected at once and constituting a separate conjunction type (Shvedova et al., 1980). Even if this is not the case, such cases would have to be excluded from the multivariate analysis because they categorically favor particular conjunctions. For instance, **odnako odnako* and **da odnako* are ungrammatical.

6. Strictly speaking, this is only true of *qualitative* dependent variables, such as conjunction identity. Logistic regression could also be used to predict the values of a *quantitative* dependent variable, which has an unlimited number of possible values, if there was some reason to believe that the dependent variable is a logistic function of the independent variables, for example, if we were predicting accuracy in producing a particular construction as a function of age.

7. Range is the difference between the smallest and largest factor weights in a factor group, which is commonly used as a (rough) estimate of the importance of the factor group.

8. The expected proportion would be one over the number of factors in a factor group, for example, 33% for conjunct type and 50% for length.

9. Predicative and pronominal adjectives were grouped together because there were not enough cases in each subcategory to treat them separately and predicative adjectives appeared to behave more like other adjectives than like other predicates.

10. It is also possible that the distinction is actually along some dimension that covaries with prevention/denial of expectation. One possibility that cannot be ruled out is that the choice of the conjunction is purely syntactic in nature, being triggered by the form of the verb heading the first conjunct (with the use of the conditional or past imperfective triggering the choice of *da*).

11. *Da* signals the preventive meaning in 44% of the cases when it is used, compared with 7% for *odnako* and 4% for *no*.

12. The ages at which children learn the Russian adversative conjunctions are unfortunately not known. My search of the Russian part of the CHILDES corpus (90,000 words) unearthed nine cases of *no* used by a child correctly at 2;4 but no cases of adversative *da* or *odnako*. However, given the size of the corpus and the rarity of *da* and *odnako* relative to *no* in the adult corpus, no cases are expected even by chance.
13. Together, the irrealis-realis construction and the asymmetric construction account for 75% of the uses of *da*.

REFERENCES

- Aaron, Jessi E., & Torres Cacoulos, Rena. (2005). Quantitative measures of subjectification: A variationist study of Spanish *salir(se)*. *Cognitive Linguistics* 16:607–633.
- Berger, Tilman, Betsch, Michael, & Bremer, Bernhard. (2001). Address systems and politeness—Independent or interdependent? Paper presented at the Linguistic Data Structures conference, Tuebingen, Germany. Handout available at: <http://homepages.uni-tuebingen.de/tilman.berger/Handouts/AddressPoliteness.pdf>.
- Boudelaa, Sami, & Gaskell, M. Gareth. (2002). A re-examination of the default system for Arabic plurals. *Language and Cognitive Processes* 17:321–343.
- Cheshire, Jenny. (1987). Syntactic variation, the linguistic variable, and sociolinguistic theory. *Linguistics* 25:257–282.
- Foolen, Ad. (1991). Polyfunctionality and the semantics of adversative conjunctions. *Multilingua* 10:79–92.
- Givón, Talmy. (1979). *On understanding grammar*. New York: Academic Press.
- Goldberg, Adele E. (1995). *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- . (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Hare, Mary L., Elman, Jeffrey L., & Daugherty, Kim G. (1995). Default generalization in connectionist networks. *Language and Cognitive Processes* 10:601–630.
- Hudson Kam, Carla, & Newport, Elissa L. (2005). Regularizing unpredictable variation: The role of adult and child learners in language formation and change. *Language Learning and Development* 1:151–195.
- Koolemans Beynen, G. (1976). Semantic differences between *no* and *odnako*. *Slavic and East European Journal* 20:167–173.
- Krilova, G. (1980). Ruskijat sojuz *da* i negovite funkcionalni ekvivalenti v xudozhestveni prevodi na bolgarski ezik. *Sopostavitelno Ezikoznanie* 5:18–25.
- Kruchinina, I. N. (1988). *Struktura i funkcii sochinitel'noj svyazi v russkom jazyke*. Moscow: Nauka.
- Labov, William. (1966). *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.
- Lakoff, Robin. (1971). 'Ifs', 'ands' and 'buts' about conjunction. In C. Fillmore & D. Langendoen (eds.), *Studies in linguistic semantics*. New York: Holt, Rinehart and Winston. 114–149.
- Lang, Ewald. (1984). *The semantics of coordination*. Amsterdam: Benjamins.
- Lavandra, Beatriz. (1978). Where does the sociolinguistic variable stop? *Language in Society* 7:171–182.
- Lekant, P. A., Gol'cova, N. G., Zhukov, V. P., Kasatkin, L. L., Klobukov, E. V., Malashenko, V. P., Tuzova, M. F., Fomenko, Ju. V., & Xolodov, N. N. (1982). *Sovremennyj Russkij Literaturnyj Jazyk*. Moscow: Vysshaja Shkola.
- Malchukov, Andrej L. (2004). Towards a semantic typology of adversative and contrast marking. *Journal of Semantics* 21:177–198.
- Payne, John. (1985). Complex phrases and complex sentences. In T. Shopen (ed.), *Language Typology and Syntactic Description II*. London: Cambridge University Press. 3–41.
- Poplack, Shana. (1992). The inherent variability of the French subjunctive. In C. Lauefer & T. A. Morgan (eds.), *Theoretical studies in Romance linguistics*. Amsterdam, Philadelphia: John Benjamins. 253–263.
- Poplack, Shana, & Tagliamonte, Sali. (1996). Nothing in context: variation, grammaticization and past time marking in Nigerian Pidgin English. In P. Baker & A. Sylea (eds.), *Changing meanings, changing functions*. Westminister, UK: University Press. 71–94.
- . (1999). The grammaticization of *going to* in (African American) English. *Language Variation and Change* 11:315–342.

- Sankoff, David. (1988). Sociolinguistics and syntactic variation. In F. Newmeyer (ed.), *Linguistics: The Cambridge Survey, Vol. IV*. Cambridge: Cambridge University Press. 140–161.
- Sankoff, David, Tagliamonte, Sali, & Smith, Eric. (2005). Goldvarb X: A variable rule application for Macintosh and Windows. Department of Linguistics, University of Toronto.
- Sannikov, V. Z. (1989). *Russkije sochnitel'nyje konstrukcii. Semantika. Pragmatika. Sintaksis*. Moscow: Nauka.
- Schwenter, Scott A., & Torres Cacoullos, Rena. (2008). Defaults and indeterminacy in temporal grammaticalization: The 'perfect' road to perfective. *Language Variation and Change* 20:1–39.
- Serebrjanaja, F. I. (1976). Nekotorye nabljudenija nad upotrebleniem sojuza *da*. *Russkij Jazyk v Shkole* 4: 76–80.
- Shuba, P. P., Germanovich, I. K., Dolbik, E. E., Karaban', I. A., Maxon', S. V., Pipchenko, N. M., Rabchinskaja, I. A., Koxno, V. F., Tarasevich, T. A., Tixomirova, E. A., Fedotova, N. V., Pleshchenko, T. P., Chechet, R. G., & Shevchenko, L. A. (1998). *Sovremennyj russkij jazyk. Vol. 3. Sintaksis. Puntuacija. Stilistika*. Minsk: Plopress.
- Shvedova, N. Ju, Arutjunova, N. D., Bondarko, A. V., Ivanov, V. V., Lopatin, V. V., Uluxanov, I. S., Filin, F., & the Institute for the Russian Language, The Academy of Sciences of the USSR. (1980). *Russkaja grammatika. Vol. 2: Sintaksis*. Moscow: Nauka.
- Torres Cacoullos, Rena, & Schwenter, Scott A. (2007). Towards an operational notion of subjectification. *Berkeley Linguistics Society* 31:347–358.
- . (2008). Constructions and pragmatics: Variable middle marking in Spanish *subir(se)* 'go up' and *bajar(se)* 'go down'. *Journal of Pragmatics* 40:1455–1477.
- Uryson, E. V. (2006). Semantika sojuza *no*: Dannye jazyka o dejatel'nosti soznaniya. *Voprosy Jazykoznanija* 5:22–42.
- Zuraw, Kie. (2000). *Patterned exceptions in phonology*. Ph.D. dissertation, University of California, Los Angeles.