Maria Zarifyan, Anastasia Melnik,
Anastasia Vyrenkova

A CASE OF USING A MULTILINGUAL DATABASE OF SYNONYMS FOR DESIGNING LEXICAL DRILLS

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: LINGUISTICS

WP BRP 49/LNG/2016

This Working Paper is an output of a research project implemented within NRU HSE’s Annual Thematic Plan for Basic and Applied Research. Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE
A CASE OF USING A MULTILINGUAL DATABASE OF SYNONYMS FOR DESIGNING LEXICAL DRILLS.

This paper presents a set of drills that train students’ competence in using the adjectives that belong to the semantic domain of FAR. The drills are based on and exemplify theoretical results of a study focusing around the comparison of synonymous adjectives across six languages and aim at granting students with a systematic and holistic picture of the domain in question. The training materials highlight the basic semantic parameters governing the correct lexical choice as well as the relevant distributional and grammatical patterns. Exemplified here are the drills for Russian and English languages.

JEL Classification: Z.

Keywords: Multilingual Database of Synonyms, foreign language acquisition, lexical drills.

---

1 National Research University Higher School of Economics, School of Linguistics, BA student, masha.zarifyan@gmail.com
2 National Research University Higher School of Economics, School of Linguistics, BA student, melnik-a-a@mail.ru
3 National Research University Higher School of Economics, School of Linguistics, associate professor, avyrenkova@hse.ru
4 This study was carried out under the auspices of the Basic Research Program at the National Research University, Higher School of Economics in 2016 (research project “Multilingual Database of Synonyms: Theoretical and Computer Models” № 16-05-0054).
1. Introduction

Even these days – when the methodology of foreign language (FL) teaching has achieved impressive results and the repertoire of activities for language learners is very large – one of the most popular types of tasks in FL classrooms are drills of various kind. Obviously, these days the structure and ideology of drills are different from those that were around in the 50s, when in the light of the behavioristic paradigm [Skinner 1957] this type of tasks formed the basis of the language teaching process. Within the framework of Skinner’s operant conditioning theory, drills served a tool for repetition that was supposed to enhance habit formation – the core of an emerging linguistic competence.

Though often criticized for their stimulus-response nature and for the mechanical, repetitive practice they provide, drills survived through the later periods of SLA development, and even under the heavy influence of Stephen Krashen’s theory of SLA [Krashen 1981, 1982] and CLT (Communicative Language Teaching) approach [Canale, Swain 1980; Nunan 1991; Savignon 1997] drills maintained their place in the syllabus of foreign language programs. Now, in the era of digitalization of the learning process and the availability of online teaching tools, the usefulness of drills is being reconsidered and gains more and more importance.

In the present paper we describe a case of designing online multilingual drills that present a practical outcome of a study conducted by the research group “Multilingual Database of Synonyms: Theoretical and Computer Models”. The group conducts a study on sets of synonyms belonging to different semantic domains in six languages (English, German, Russian, Polish, Italian and French) The study is conducted in three stages. First, each item in a set of synonyms is described with a view to juxtaposing their semantic and syntactic features. Second, idiomatic meanings of each word are highlighted (it is especially important for designing language drills because the comprehension of idioms and metaphors usually presents a huge problem in language learning). The final stage of our work is the development of online drills that are based on the descriptions.

2. Why basing drills on a cross linguistic research?

The main purpose of any language task is to develop a learner’s interlanguage [Selinker 1972] into a lexico-grammatical system, maximally similar to the system of the FL being acquired. In the majority of textbooks and exercises the goal is achieved with a sharp focus on the target language. That means that the differences between the lexical systems of the mother tongue and L2 are commonly overlooked. Given this, the first reason for basing drills on a multilingual research stems primarily from the desire to smoothen the transition of a learner’s interlanguage from L1 to L2. And since the theoretical and descriptive accounts developed by our research group pinpoint parallels and contrasts between lexical systems of different languages, they also give a clear opportunity to put these distinctive features into the core of drills for FL students.

The second reason is giving a systematic and structured overview of a specific lexical domain. Most commonly vocabulary descriptions and tasks are presented within relatively broad topics, both concrete (SPORT, SHOPPING, TOURISM etc.) and abstract (COMMUNICATION, POLITICS, SOCIETY etc.). This approach has an obvious advantage: it systematizes learner’s lexical resource at a larger scale grouping words, collocations and idiomatic expressions into domains relevant for communication purposes. At the same it has two important flaws: first, the list of lexical items delivered to students is not exhaustive (usually defined by frequency effects), second, only a limited set of significant semantic features that account for the correct lexical choice in context is explained and trained. These flaws especially impede the acquisition of close

5 A learner acquires a FL through a controlled activity: producing a stimulus evoked utterance (operant) and hearing a response (consequence) to it.
synonyms, because FL learners are not given a holistic and structured picture of narrower semantic domains. Thus, subtle distinctions between close synonyms in the target language can only be acquired by learners through self-training – extensive reading and working with dictionaries at a more advanced level – a process that doesn’t necessarily lead to the best result. The ultimate reason is giving a single systematic ground that enables designing drills for several languages at a time. A cross-linguistic comparison results into a systematic set of features and parameters that are potentially lexicalized in different languages, which substantially simplifies the procedure of building tasks for different FLs.

3. FAR domain in English and Russian.

Within the scope of this paper we take a closer look at the concept FAR and the means of its lexicalization and present the sets of drills for Russian and English languages. The choice of this concept is easily explained. Though the polysemy of spatial words is well-studied, there aren't many works that give a systematized cross-linguistic account of lexical distribution between synonymous words and patterns of their metaphorization. At the same time, many parameters relevant for the domain structure have already been diagnosed. For example, spatial adjectives can be used both in deictic (reference to a speaker position) and non-deictic modes (see E1-E2 respectively):

E1: My school is far away and I get very tired walking there.
E2: My school is very far from the city centre.

Also, lexemes with 'far' semantics are characterized by having two usage types due to their valence of landmark: on the one hand, they can be used in deictic contexts, in which the observer acts as a reference point (see [Paducheva 1993]), on the other hand, they can form constructions with narrative meaning, in which the reference point is an object that is outside the speaker’s universe.

Moreover, it is well-known, that spatial lexemes can develop temporal meanings. “Space-time” metaphor is considered to be one of the language universals of semantics, mostly manifested in spatial prepositions and adverbs [Fillmore 1971; Lakoff , Johnson 1980; Clark 1973], yet also found in Russian adjectives [Bulygina, Shmelev 1997]: the authors observe Russian antonymous adjectives blizkij ‘close, nearby’ and dalekij ‘far’ and point out that blizkij can denote only future events, but not the past ones: blizkoje buduščeje - 'near future'(*blizkoje prošloje - 'near past'). Dalekij most likely denotes events distanced in the past (dalekoje prošloje - 'remote past'), however it can be also used when one is talking about the future (dalekoje buduščeje - 'distant future'). The authors also notice that the expression dalekoje vremja - 'distant time’ is most likely related to the past rather than to the future.

While ‘space-time’ metaphorical mapping is universal, there are also metaphors that are less widespread across languages and thus they are quite complicated for language learners. In order to facilitate language learning, we have highlighted the following meanings of the concept ‘far’ that may vary across the languages.

Thus, relevant is the type of an oriented object, instantiated with animacy/inanimacy parameter which all languages are sensitive. To illustrate this point, we turn the reader's attention to kinship terms: when one denotes kinship between two or more people in English, they would use adjectives distant and remote (cf. distant <remote> relative, but *far relative). This situation also has special means of lexicalisation in Italian (cf. parente lontano), German (cf. entfernter Verwandter), Russian (dal'nij rodstvennik), French (cf. cousin éloigné) and Polish (cf. daleki krewny).

It is worth noting that Russian distinguishes between the concepts of ‘distant relative’ and ‘distant ancestor’ and uses two distinct adjectives for each context - dalekij predok - ‘distant ancestor’, but dal'nij rodstvennik - ‘distant relative’. One can claim that there are two different kinds of ‘distance’ between people bounded by family ties - dalekij denotes the long time period
between the living dates of an ancestor and a descendant, while *dal’niy rodstvennik* indicates that indirect kinship between two people who are contemporaries - for example, a second cousin can be called a *distant* relative.

Another important feature is that two oriented objects can be considered as central and peripheral: for instance, when one says *remote village* or *abgelegen Dorf* (‘remote village’) in German, that means that the village is a peripheral object, and or the town is a central one, but, as a rule, a central object is not expressed explicitly and can be omitted.

The last feature to be mentioned in this section is the lexical potential to express the exact distance. Not all the synonyms have the valency of distance: thus, one can say *This town is 5 km distant <far> from London* in English or *Das Haus ist von dem Stadt 7 kilometer entfernt* (‘The house is 7 km far <distant> from the city’) in German, but not *This town is 5 km remote from London* or *Das Haus ist von dem Stadt 7 kilometer fern*, because the adjectives *fern* and *remote* have no valency for overtly expressing distance between objects.

In chapters 2.1-2.5 we take a closer look at the structure of drills and theoretical principles underlying them. Drills for Russian and English are described in separate sections.

### 3.1 An overview of drills

For every language we identified the means that help distinguish between the synonyms within the FAR domain (e.g. lexical constructions, collocations, idiomatic expressions). Each of such means is further supported by a separate drill.

There are also special tasks that focus students’ attention on lexicalization patterns of particular situations or frames6. Exercises based on frames cover all the meanings of synonyms including both direct and figurative ones (for example, there is a common shift from the meaning ‘far’ to the meaning ‘not knowing or not interested’ in European languages).

We also introduced tasks on word forms because morphological characteristics correlate with lexical choice. Thus, morphological paradigms of adjectives with different meanings are not similar (e.g. Russian adjectives normally have three comparative forms - neutral, comparative and superlative - but the adjective *dal’niy* has no comparative forms at all).

Prior to drills, a learner is provided with a set of rules that describe the use of synonymous words in different contexts.

### 3.2 From the mother tongue to the target language

As we have already mentioned in the introductory section, the majority of exercises and tests from language learning textbooks give the information about foreign language material without considering the structure of a learner’s native language. Our multilingual drills allow a user to establish connections between grammatical and semantic structure of their native and target languages. We have called such type of exercises “From the mother tongue to the target language”.

This goal is achieved in particular by matching lexical items and frames that they denote. In this exercise students should fill in the gaps with the correct word, which contributes to their understanding of the FAR domain in general. The list of meaning types (frames) is given in a student’s native language and covers all the situations that belong to this domain (for more details see sections 2.3 and 2.4).

### 3.3 Drills on FAR in English

---

6 For a broader discussion of the term *frame* in lexical typology, see [Rakhilina, Reznikova 2014]
FAR domain in English is covered by five synonymous adjectives: far, far-away, far-off, remote, distant.

Due to the fact that English adjectives and adverbs have the same form and therefore a beginner might find it difficult to distinguish between them, we have highlighted the typical contexts and meanings for both adjectives and adverbs (the description of meanings for English drills was based on http://www.thefreedictionary.com/far):

**Adverb:**
1. To, from, or at a considerable distance: *a cat that had strayed far from home.*
2. To, from, or at a much earlier or later time: *a movie that takes place far in the future.*
3. To a considerable degree; much: *felt far better yesterday; eyes that seemed far too close together.*
4. To an advanced point or stage: *a brilliant student who will go far.*

**Adjective:**
1. Being at considerable distance (far, distant, remote): *a far country.*
2. Going back a considerable extent in time (far, far-off, distant): *far-off future, distant (remote) past.*
3. Coming from or going to a distance (distant): *a distant sound; a distant telephone call.*
4. More distant than another (usually of two things) (far). *He lives on the far side of the lake.; the far corner.*
5. Extensive or lengthy (far): *a far journey*
6. Distantly related by blood or marriage (distant): *distant cousin.*
7. Minimally similar (distant): *a distant likeness.*
8. Operating or controlled from a distance (remote): *remote sensors, remote job, remote control*
9. Hidden away; secluded (remote): *remote village*

Here is the example of the first task that follows the description:

1. *This ____ town is so out of the way that mail comes only once a week.*

A learner must read the sentence and fill in the gap with one or several suitable adjectives. The context helps to choose the correct lexeme: for instance, from the sentence above, a learner can infer that the town is secluded and difficult to reach, therefore the correct answer is *remote.*

As learners see the differences between the contexts described by each lexeme, they are ready to proceed with more complicated tasks: in the following example one must fill the gap not only with the correct adjective but also with a noun, which can be inferred from the context and used with this adjective:

2. *John is my aunt’s son. He is my _______ _______.*
3. *I can’t work in an office, because I take care of my child. That’s why I’m looking for _______ _______.*

We also take into account all possible correct answers; thus, in the first sentence one can fill the gaps either with the construction distant relative or distant cousin, so we consider both options in the answer key to the tasks.

The next step is training word formation. A learner is provided with the following set of wordforms derived from the lexemes far, remote and distant and is offered to complete the task:

**remote:**
- ADV: remotely
- NOUN: remoteness

**distant:**
ADV: distantly
NOUN: distance

far:
COMPARATIVE-SUPERLATIVE: far, further/farther, the furthest

NB!: further also could be used as 'next, additional' (give us a call if you require further information)

The complexity of the following task is that a learner must keep in mind two features: compatibility of a lexeme with a particular context and the correct derivative form of an adjective:

(4) Our users would log in ___(ADV), do their work, and the data transfer took place locally on that server. #remotely
(5) For ____ (COMPARATIVE) details please contact us at 077734. #further
(6) Due to the _____ of Australia, the flight Moscow-Sydney will take 15 hours. #remoteness

As language learners can distinguish between the direct meanings of the lexemes, they are offered the last task on some idioms and set phrases. As it turned out, only the lexeme far can form the idioms and collocations, thus, the following exercise is focused on comprehension of idioms:

far:
1. Can be used with the names of cardinal directions: far West
2. To an advanced point or stage: a brilliant student who will go far.
3. Everywhere: far and wide
4. Something that is very different from something else: This food is a far cry from what we got in the cafeteria.

(7) Instead of returning by way of the Atlantic we made arrangements to cross the Pacific Ocean and voyage via the ____ East. #far
(8) The hospital in Mbale is a ______ from what we would expect to find in a UK hospital. #far cry
(9) We were looking _________ for the lost diamond ring. #far and wide
(10) We all need to take some time and focus on something we are eager to learn. Dave - you will do fine and since we all need to learn another language, I know you will ____! #go far

3.4 Drills on FAR in Russian

As mentioned above, there are several ways to distinguish between the synonyms with the common sense 'far', such as comparing of forms, constructions, frames and collocations. All of them seem to be relevant for the Russian language. Here we will take a quick look at drills for Russian learners whose dominant language is English.

Russian adjectives with meaning ‘far’ differ in several ways. First, some adjectives do not have the full set of forms (e.g. the adjective dal'nij 'distant' does not have a comparative form and the adjective otdalennyj 'far-off' is the only one which has a synthetic superlative form otdalennejšij). This peculiarity is trained with the help of gap filling tasks like (11):

(11) On ležal i smotrel, kak zažigajuts'a dalekie ogni goroda i eščjo __________ zvezdy. (dalekij, comp.) (lit. 'He was lying and looking at the lightening of far-off city lights, and also __________ stars')
As we have already mentioned in chapter 2.3, we use semantic frames to introduce the student to the set of situations lexicalized with FAR adjectives. Below we present some of the rules that familiarize the student with the domain:

“We use dal'nij if the situation describes a) two objects/sites belonging to the same semantic class – one being at a greater distance than the other (cf. Iz dal'nix i bližnix mest priexali mastera, lit. ‘From distant and near places masters came’) or b) a long way that the object is to go or has gone (cf. Ja uezžaju v dal'nij put', žena i deti doma ždut, lit. ‘I’m setting off on a long way, my wife and children are waiting at home’).”

The student should be aware of the frame type described by the task sentence and match the frame with the adjective or set of adjectives if more than one lexeme can be used, see (12), where the correct answer is dal'nij ’distant’:

(12) Zavtra my otpravljaemsja v __________ put’. (lit. ‘Tomorrow we are setting off on a __________ way’)

After the person is presented with morphological differences and relevant distinctions in meaning between the synonyms, they train some special constructions with these lexemes, as well as collocations and idioms. Therefore we represent a list of idioms (the only adjective which can form idioms with itself is dal’nij ‘distant’) and ask a student to remember them. Then we offer them to match each sentence in the task with the appropriate answer which should be selected from a set of given options semantically connected with each other (see (13) where the right answer is Bocmok ’East’):

(13) ______________ i Sibir’ bogaty poleznymi iskopaemymi, ryboj i lesami. (lit. ’_________________ and Siberia are rich in fossil fuels, fish and woods’)
    Dal’nij/Dal’naja: a) Zapad b) Avstralija c) Okean d) Vostok

Finally, we propose a task based on both semantic and grammatical properties of an adjective. In the theoretical description given before the task we list semantic valences for the trained adjectives (object, landmark and exact distance) and the types of constructions where these valences can be expressed (e.g. in Russian we use the preposition ot ‘from’ and a noun phrase to express the landmark, and the preposition na ‘on’, a quantitative phrase and a noun phrase to express exact distances). Thus, this task includes instances with gaps that should be filled with one adjective and one preposition (see (14), where the adjectives udalennyx or otdalennyx should be used and the preposition ot ‘from’, as gorodov ‘cities’ is landmark):

(14) Učenye provodjat eksperimenty v gluxix rayonax, ____________(participle) (preposition) gorodov. (lit. ‘Scientists carry out experiments in desolate regions, ___________ towns’)

4. Analysis of common errors

Since the main purpose of language drills is to predict and minimize the number of mistakes that students make in their speech, before designing the tasks, we had analyzed the data presented in Learner Corpora (Russian Learner Corpus (RLC) and Russian Error-Annotated Learner English Corpus), so that we could focus on the most problematic areas for students of Russian and English as FLs.

In chapters 3.1-3.3 we will discuss some frequent errors attested in these corpora.
4.1 Adverb vs Adjective

In Russian, unlike many other European languages, adjective and adverb as morphological categories differ greatly. So we couldn’t unite them into one group and we had to point out the differences between them.

In many contexts we can use an either adjective or an adverb without violating standard grammatical rules, but they are not always interchangeable. As the corpora show, it is not absolutely clear to Russian learning students, whether it is preferable to use an adjective or an adverb in a particular context. They use adverbs more often than adjectives, which may lead to an error (cf. (15) where the speaker used the construction *daleko ot lingvistiki*, lit. 'far.adv from linguistics', instead of *dalekij ot lingvistiki*, lit. 'far.adj from linguistics').

(15) Čelovek, kototyj *daleko* ot lingvistiki, ispol’zuet svoju etimologiju slov. (lit. "A man who is far.adv from linguistics uses his own etymology of words")

Actually, both adverbial and adjectival constructions occur in standard Russian, however their distribution is not arbitrary. As the main source of data on the Russian language is the Russian National Corpus, we checked it to see which factors underlie this distinction. The results showed that adverbs are more often used in spatial contexts7 (but not always, see the example (3)) and the second one - in a figurative sense (see the examples (16) and (17)).

(16) 'Moj syn 8 let učitsja v škole, kotoraja daleko ot doma. [Naši deti: Doškolyata i mladšie škol’niki (forum) (2005)]. (lit. 'My sun has been studying for 8 years at a school that is far from home')

(17) Ja *dalek* ot politiki, hotja mne nebezrazlično proishodjaščije v našej strane. [Galina Stepanova. Neispravimiy Al’demaro // « Soveršenno sekretno», 2003.07.04] (lit. 'I'm far from politics although I care about what happens in our country')

4.2 Adjective forms

The other challenge in lexical choice for students may stem from the subtle semantic differences underlying the use of various grammatical forms of the same adjective.

These differences may be found between short and long forms of an adjective or between its singular and plural forms. For the latter case cf. examples (18 a-b), where both collocations *poslednee vremja* 'latest time' and *poslednie vremena* 'latest times' (for more information on the adjective *poslednij* 'last/latest' see [Apresjan, Shmelev 2016]) are grammatically valid and have the same meaning 'times when the speaker (= a reference point) lives', and (19), where the expression *poslednie vremena* 'latest times' is idiomatized and means 'doomsdays'. In (19) the use of the adjective *poslednij* 'last/latest' and the noun *vremja* 'time' in the singular form is unacceptable.

(18a) Pravda, v *posledneje vremja* dela ego kompanii na birže šli nevažno. [Elmar Gusejnov. Čisto francuzskoe samoubijstvo. Pokončil s soboj odin iz samyx izvestnyx kulinarov strany (2003) // «Izvestija», 2003.02.25] (lit. 'However, in latest time his business was nor quite successful on stock exchange')

(18b) Ne vezet mne v *poslednie vremena* na ličnuju žizn’. [Andrey Beljanin. Svirepyj landgraf (1999)] (lit. 'I'm not lucky in love affairs in the latest times')

(19) Nastuplenije poslednih vremen / *poslednego vremeni i okončatel’nogo Strašnogo Suda Božija na padošim mirom neotvratimo... [mitropolit Vladimir (Ikim). Slovo v den' pamjati apostola i evangeliasta Ioanna Bogoslova (2004) // «Zurnal Moskovskoj Patriarxii », 2004.05.24] (lit. 'The advent of the latest times / *latest time* and the ultimate Great God’s Judgment over the fallen world is inevitabe')

---

7 Such a statement also has its exceptions, cf. Otvetstvenno zayavljaju: vsjo eto daleko ot istiny. [Anatolij Sakov. Detjam nužen rejniŋ! (2004) // «64 — Shaxmatnoje obozrenie», 2004.09.15], lit. 'I'm postulating with the whole responsibility: all of this is far from the truth'. A deeper analysis of this problem, however, is not the central topic of the present paper.
5. Conclusion

In this paper we outlined the template and theoretical accounts of online drills developed on the basis of a Multilingual Database of Synonyms – a project aimed at contrasting sets of synonyms across languages. The analysis of these sets highlights parameters relevant for different lexicalization strategies in six languages (English, German, Russian, Polish, Italian and French) and enables incorporating those of them that operate in the mother tongue and the target language into tasks for foreign language students. The drills also provide students with a systematized training for better acquisition of lexical domains characterized by subtle semantic and grammatical distinctions between their members (exemplified in this paper, are words belonging to the domain of FAR in English and Russian). The tasks are based on semantic features and grammatical patterns used for each synonym and its meanings. Particular attention is given to cases that are frequently misused by foreign learners – such cases are detected and analysed with the help of learner corpora and presented within the scope of this paper for the Russian language.

References

Contact details and disclaimer:

Anastasia Vyrenkova
National Research University Higher School of Economics, School of Linguistics, associate professor, avyrenkova@hse.ru

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

© Zarifyan, Melnik, Vyrenkova 2016