

Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2018”

Moscow, May 30—June 2, 2018

LINEARLY ORDERED GROUPS OF ADVERBS

Vvedensky V. L. (VictorLvo@yandex.ru)

NRC Kurchatov Institute, Moscow, Russia

We analyzed parallel lists of Russian and English adverbs. They are linked together by acceptable translations. Many adverbs have multiple translations which we use for the ordering of both lists. The lists can be subdivided into domains of meanings, which include groups of semantically close linearly ordered adverbs. We have identified 34 such groups, distinct in basic meaning. Small fraction of adverbs may have several different meanings. These outliers indicate links between semantically different groups of words. We believe that these links reflect cortical representation of the elements of language similar to ocular dominance pattern in the visual cortex. It consists of parallel bands of cortical columns and emerges before visual experience. The adverbs are organized into similar pattern and those with multiple meanings presumably indicate bifurcations in the cortical representation of words.

Keywords: multiple translations of word, semantic space structure, Russian and English adverbs

ЛИНЕЙНО УПОРЯДОЧЕННЫЕ ГРУППЫ НАРЕЧИЙ

Введенский В. Л. (VictorLvo@yandex.ru)

НИЦ Курчатовский институт, Москва, Россия

Introduction

Everybody knows that most of words make sense or have meaning, though “the notions of *word* and *word meaning* are problematic to pin down” [1]. Meaning seems at once the most obvious feature of language and the most obscure aspect to study. Researchers in computer science try to allocate words in a metric semantic space based on their meaning [2]. In this space the words acquire geometric properties [3]. Usually a tree or a network is built, where the words are nodes connected by certain relationships. There are several approaches to generate the network of words [4], most often the words are nouns and the emerging structures are very large.

The language and its components are somehow packed into the human brain. Neuroimaging studies have identified regions that seem to represent meaning of words and it looks like these regions cover nearly the whole cortex [5] or considerable part of it [6]. However, not only the meanings of words should be represented in the brain. The pronunciation of words requires a repository of standard subroutines for each word. They activate proper sequence of contractions of articulatory muscles. This component of language may have a different distribution in the cortex than the “semantic system”. Semantic maps created with distance metrics that emphasize different aspects of words may have different properties [7]. Various similarity relations between words can be used to construct the embedding space—in this work we adopted dictionary approach.

Many, many words have several translations into other language. Translation of a particular word depends strongly on the context. Dictionaries assess the frequency of word use and present several most common translations. One can build proximity measure between words using this data. We analyze adverbs because they form a distinct group of appropriate size which is both small enough to handle with, using available software, and large enough to make inferences. Our results show that it is important that the majority of existing words of a certain type should be included into analysis.

Results

We used translation services Google Translate, Reverso Context and Translate.academic.ru [8–10] to construct a table which contains the most essential part of translations between Russian and English adverbs. If necessary, we used Oxford Dictionaries [11] to clarify the uncertainty of the use of a particular translation. Single-word translations from one language to another were basically considered. Starting with any common adverb one can find majority of important adverbs of both languages using repetitive translations from Russian to English and back. To date we accumulated 1979 Russian and 1985 English adverbs which form an interconnected array. One can find more adverbs in both languages, for example, the dictionary of the ETAP linguistic processor (ca. 130,000 words altogether for each language) has over 9,200 Russian adverbs and about 6,700 English ones [12]. Scrutiny of the smaller scale frequency dictionary for Russian [13] shows, that nearly all extra adverbs have only single translation or form small isolated groups. We believe that our selected core of the thesaurus of adverbs is fairly representative for inferences.

Russian and English adverbs were accumulated in an Excel spreadsheet, as shown in Fig. 1. A set of custom-made mathematical tools was used for manipulations with

the emerging parallel lists of adverbs. Repetitive Rus-Eng-Rus translations produced initial small table of common adverbs where acceptable translations were indicated as 1. The scatter of units in the raw table looks quite random. Reshuffling of the collected adverbs minimizes the scatter of units around the diagonal of the table. This procedure is carried out automatically after the addition of several new words. Inspecting emerging ordered Russian and English lists one can easily detect discrepancy between semantic meanings of some adjacent words in the lists. This is a consequence of the polysemy of words. Majority of adverbs tend to display gradual change of meaning along the ordered lists, though some words pop out clearly. One can start accumulation of a new group of adverbs around this outlier. Two groups start to grow up when new adverbs are added. Later on more independent groups of adverbs emerge. The growth of each group eventually stops when no new translation can be found. Adding considerable amount of adverbs into an existing group can result in the splitting of the group into two parts, densely interwoven inside, but with a weak connection between them. These parts can be either nearly equal, or one is much smaller. In the latter case the smaller part sometimes can find its natural place in another existing group. In this way accumulation of adverbs turns into development of an array of groups, which grow up, become more accurately ordered, split, exchange clusters of words and eventually get their final shape. In some cases, when it was difficult to attribute certain adverb to an existing group, we used similar procedures for Russian and French adverbs.

Fig. 2A visualizes our results. Two domains of meanings (different from shown in Fig. 1) are represented as parallel lists of Russian and English adverbs, where acceptable translations are indicated as lines connecting words in different languages. Each domain is “stitched together” into a linear structure by a dense array of lines (translations). To make easier the perception of the image we present parallel lists of adverbs in both domains as a text.

Самонадеянно, вульгарно, крупно, самоуверенно, безразлично, грубо, отчужденно, поодаль, сипло, вопиюще, равнодушно, хрипло, высокомерно, свежо, надменно, важно, заносчиво, бессердечно, нагло, провокационно, вызывающе, беззастенчиво, дерзко, демонстративно, нахально, бесстыдно, размашисто, нахрапом, развязно, ослепительно, бесцеремонно, позорно, постыдно, подсознательно, неосторожно, беззаботно, бессознательно, халатно, блаженно, небрежно, безмятежно, неосознанно, беспечно, легкомысленно, поверхностно, несерьезно, необдуманно, нерадиво, бездумно, неволью, мельком, вскользь, кстати, попутно, мимоходом, невзначай, нечаянно, непреднамеренно, ненамеренно, ненароком, неумышленно, произвольно, неровно, случайно, неодинаково, по-разному, неравномерно, по-иному, иначе, различно, по-другому, наудачу, набум, беспорядочно, выборочно, вразброс, произвольно, наугад, хаотично, нестройно, смущенно, хаотически, суммарно, упрощенно, слепо, вслепую, беспомощно, рассеянно, комично, забавно, смешно, иронично, смехотворно, абсурдно, нелепо, безучастно, невыразительно, глупо, сдуру, наивно, бессмысленно, неостроумно, тупо, бесцельно, бесполково; conceitedly, vulgarly, presumptuously, coarsely, grossly, roughly, rudely, indifferently, perkily, aloof, huskily, hoarsely, blatantly, raucously, freshly, loftily, superciliously, flagrantly, coolly, importantly, arrogantly, insolently, callously, provocatively, pertly, acock, brazenly, impertinently, defiantly, boldly, shamelessly, unashamedly, saucily, impudently,

glaringly, forwardly, unceremoniously, dazzlingly, shamefully, ignominiously, unwarily, offhandedly, subconsciously, airily, unconsciously, negligently, carelessly, blissfully, blithely, serenely, jauntily, perfunctorily, unknowingly, frivolously, lightly, superficially, thoughtlessly, mindlessly, unwittingly, casually, incidentally, inadvertently, accidentally, unintentionally, unevenly, variously, fecklessly, differently, haphazardly, otherwise, irregularly, at random, randomly, indiscriminately, arbitrarily, disorderly, confusedly, chaotically, summarily, blindly, simplistically, helplessly, absentmindedly, distractedly, absently, diffusedly, comically, amusingly, funnily, ironically, ridiculously, ludicrously, vacantly, farcically, absurdly, blankly, apathetically, foolishly, fatuously, naively, aimlessly, pointlessly, puproselessly, senselessly, dully, stupidly, muzzily.

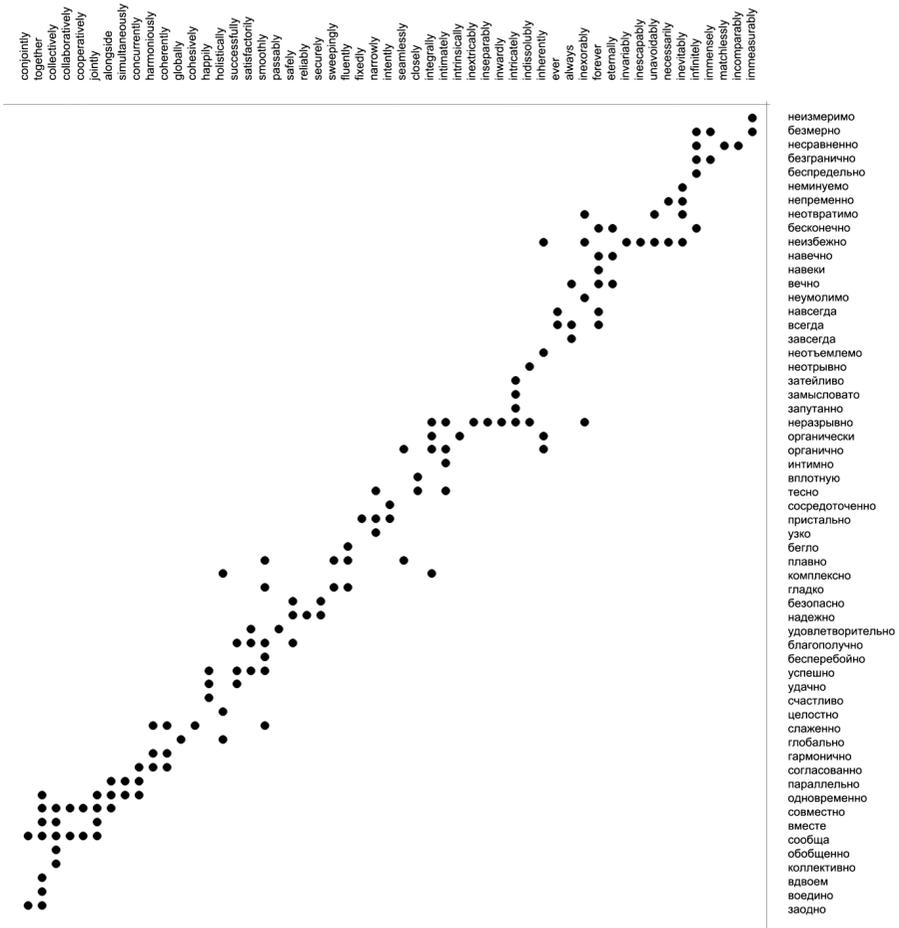


Fig. 1. Part of the whole table of translations between Russian and English adverbs, corresponding to a single domain of meanings. The points mark acceptable translations. Gradual shift of the meaning of adverbs in Russian and English lists can be seen

Покойно, неторопливо, неспешно, не спеша, методично, потихоньку, медленно, непрерывно, методически, беспрестанно, непрестанно, непрерывно, планомерно, понемногу, надолго, исправно, постепенно, регулярно, периодически, постоянно, ритмично, исподволь, систематически, поэтапно, прогрессивно, подряд, неизменно, поочередно, последовательно, попеременно, устойчиво, взаимозаменяемо, неотступно, стабильно, навязчиво, деловито, неуклонно, цепко, неустанно, непреклонно, уверенно, упорно, назойливо, упрямо, неумолимо, стойко, густо, компактно, плотно, твердо, туго, намертво, наглухо, затруднительно, сложно, крепко, жестко, накрепко, прочно, трудно, настойчиво, решительно, настоятельно, категорически, наотрез, тяжело, безапелляционно, энергично, сильно, повелительно, усиленно, агрессивно, властно, грузно, напряженно, мощно, могущественно, активно, интенсивно, силой, напористо, насильственно, деятельно, честолюбиво, масштабно, поневоле, инициативно, упреждающе, насильно, вынужденно, принудительно; *easy, unhurriedly, leisurely, methodically, unceasingly, slowly, ceaselessly, systematically, incessantly, continuously, continually, permanently, regularly, constantly, recurrently, intermittently, rhythmically, gradually, serially, progressively, successively, consecutively, sequentially, consistently, alternately, sustainably, interchangeably, steadily, persistently, stably, obtrusively, unswervingly, busily, adamantly, tenaciously, relentlessly, confidently, tirelessly, stubbornly, doggedly, obstinately, steadfastly, indefatigably, perseveringly, thickly, densely, compactly, rigidly, staunchly, importunately, tightly, difficult, arduously, firmly, resolutely, hard, solidly, insistently, strongly, decisively, sternly, emphatically, flatly, categorically, strictly, strenuously, gravely, peremptorily, vigorously, aggressively, imperiously, heavily, forcefully, mightily, overbearingly, authoritatively, powerfully, tensely, energetically, intensely, intensively, constrainedly, actively, proactively, ambitiously, forcibly, involuntarily, forcedly, compulsorily.*

Scrutiny of these lists reveals gradual change of meaning of the subsequent words. It is possible that some words do not fit perfectly and would be more properly attributed to another group (since our procedures are semi-automatic), though we did not see many such cases. Small changes of meaning throughout the list result in drastically different words on both ends. This holds for all domains which incorporate majority of important Russian and English adverbs. We conjecture that this reflects the principle of word storage in the cortex.

In addition to translations which cluster together in a single domain, there are few, which connect groups of words that differ greatly in their semantic content. The word *involuntarily* can be translated as *принудительно* and *непреднамеренно*. These words fall into different groups of Russian adverbs created by our procedure of selection. Links between different domains emerge as shown in Fig. 2A.

The array of translations (lines) inside each linear domain looks quite uneven. Few words have many translations, while majority of adverbs have only one or two. The number of adverbs with the same number of translations falls down exponentially with the number of translations per word. This is shown in Fig. 2B. The number of adverbs falls 4 times with increment 3 in the number of translations—1, 4, 7, 10, 13. The excellent exponential fit and characteristic numerical relationship leads to the conclusion that this is hardly an accidental coincidence. Some basic and robust mechanism of cortical function is hidden behind this accurate interrelation.

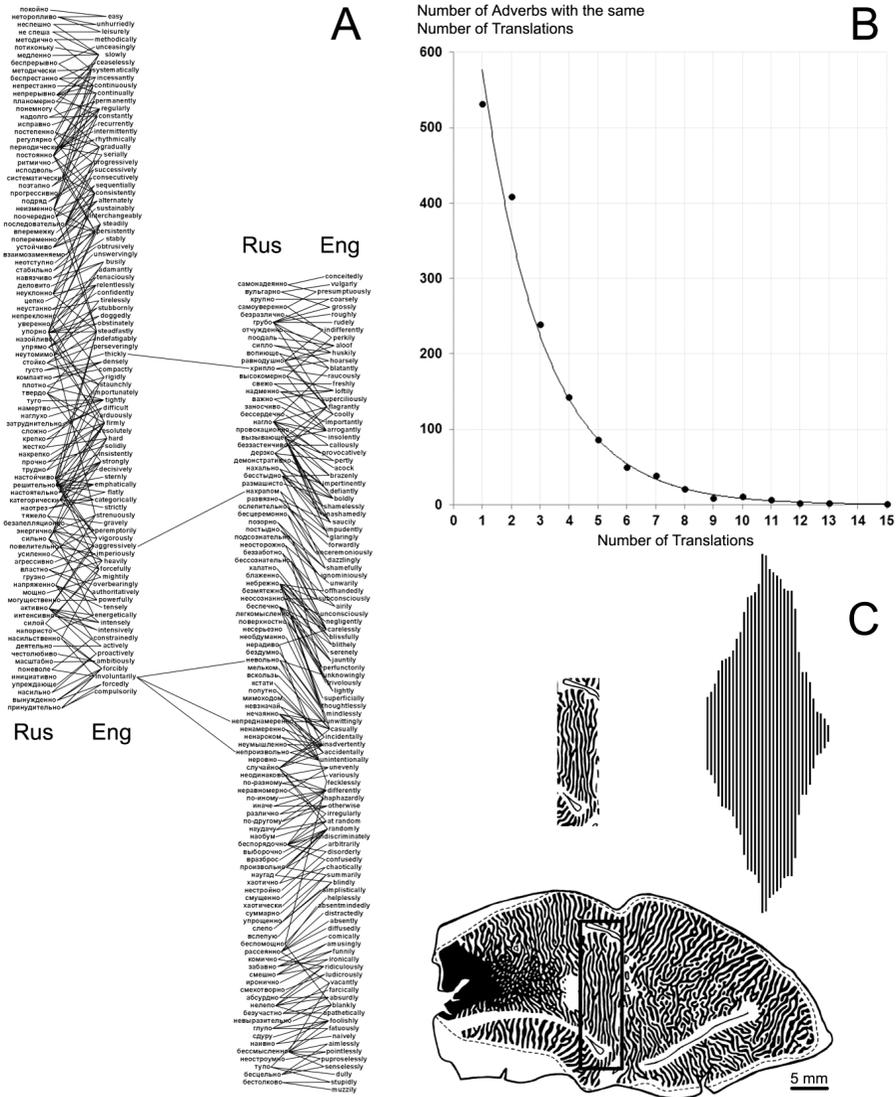


Fig. 2. A) Visualization of translations as lines connecting adverbs in Russian and English parallel lists. Two different domains of meaning are presented. Translations inside each list stitch together the adverbs into the linear structure. Rare translations-outliers link the words with multiple meanings in different domains. B) Vertical axis shows the number of Russian adverbs which have the same number of translations into English. The latter number is indicated on the horizontal axis. The trend line shows exponential dependence. C) Pattern of ocular dominance in the striate cortex of newborn monkey [14]. It consists of bands of cortical columns. A portion of the cortex is shown above for comparison with the array of domains of meaning for adverbs. Two patterns may have common background.

Some hints on the structure of the array storing subroutines of words in the cortex come from the comparison of the domain size (or length, or number of adverbs). We have found that the length falls down nearly linearly when the domains are ordered by size. In Fig. 2C the domains are arranged so, that those having more interconnections (Fig. 2A) are closer to each other. Each band is one-word thick, while the length of the band equals the number of adverbs in the domain.

Discussion

The spindle-like structure incorporates majority of important adverbs and one can make conjectures on its size on the cortical surface. Linearity, the number of elements in the group, and the number of emerging groups, suggest that the thesaurus of adverbs resembles the structure well known to brain researchers—zones of ocular dominance in the striate cortex. The columns, containing several thousand neurons each, line up into the rows on the cortical surface, which look like the pattern of dark and light stripes shown in Fig. 2C. It is important to mention that these bands develop in the cortex of a newborn before visual experience. Such periodic structures are observed in different areas of the cortex, including those that are usually associated with the function of speech [15]. It is not yet understood how the system of ocular dominance and similar banded structures work in the brain. The spindle-like area of parallel bands, shown in Fig. 2C, looks too perfect in comparison with the real ocular dominance pattern in the cortex. The latter contains bifurcations or loops forming intricate maze, though preserving general layout of parallel bands. We believe that the multiple meaning words, connecting different domains in Fig. 2A, provide information on the real structure of the “maze of adverbs”. It occupies a small patch of cortex and hence the number of adverbs is limited. We already see that other words (nouns, verbs, adjectives) are not much different from adverbs in the structure of the embedding space.

Strict localization holds for the subroutines of words used for vocalization. In contrast, meaning is imbedded into running context which is distributed throughout the cortex. Selection of an appropriate word to be perceived or uttered implies convergence of streams of neural pulses from many cortical locations into the patch where the subroutines of words are stored. Similar to the light focused on the retina of the eye. Some words (*резко, безусловно, ужасно, решительно, completely, badly, nicely, immediately*) have large receptive fields or many translations, while the vast majority of cortical columns storing articulation subroutines respond to just single meaning in the stream of perceived words or events. We do not understand yet how this system works, though it looks clear that the interleaved bands provide more dense storage of linguistic material, like folded book pages do.

Conclusions

Analysis of multiple translations of words from language to language provides important information on the structure of the space embedding different types of words. The structures emerging from this analysis can be compared with those

obtained in human subjects with the help of neuroimaging. This can provide insight into language functioning in the neural tissue.

The author is supported by the Russian Fund for Basic Research, grant 15-29-03814-ofi_m.

References

1. *Gasparri, L., Marconi, D.* (2016). Word Meaning, The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), Edward N. Zalta (ed.), [Online]. Available: <https://plato.stanford.edu/archives/spr2016/entries/word-meaning/>.
2. *Samsonovic A. V., Ascoli G. A.,* (2010). Principal Semantic Components of Language and the Measurement of Meaning. *PLoS One*. 2010; 5(6): e10921.
3. *Gärdenfors, P.* (2014). *Geometry of meaning: semantics based on conceptual spaces*. Cambridge, Mass.: MIT Press. ISBN 9780262026789.
4. *Masucci A. P., Kalampokis A., Eguíluz V. M., Hernández-García E.* (2011). Wikipedia information flow analysis reveals the scale-free architecture of the semantic space. *PLoS One*. 2011; 28;6(2):e17333.
5. *Huth A. G., de Heer W. A., Griffiths T. L., Theunissen F. E., Gallant J. L.* (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532 (7600), 453–8. PMID: 27121839.
6. *Petrides, M.* (2014). *Neuroanatomy of Language Regions of the Human Brain*: Elsevier, 186 pages.
7. *Samsonovich A. V., Goldin R. F., Ascoli G. A.* (2010). Towards a semantic general theory of everything. *Complexity* 15 (4): 12–18.
8. Google Translate, available at: <https://translate.google.com/>
9. Reverso Context , available at: <http://context.reverso.net/>
10. Translate.academic.ru, available at: <https://translate.academic.ru/>
11. Oxford Dictionaries, available at: <https://en.oxforddictionaries.com/>
12. *Apresjan Ju., I. Boguslavsky, L. Iomdin, A. Lazurskij, L. Mitjushin, V. Sannikov, L. Cinman.*(2003) “ETAP-3 LinguisticProcessor: a Full-Fledged NLP Implementation of the Meaning Text Theory,” Proceedings of the 1st International Conference on Meaning-Text Theory. Paris, Ecole Normale Supérieure, June 16–18; 279–288, 2003.
13. *Sharoff S.* (2001) The frequency dictionary for Russian, available at: <http://artint.ru/projects/frqlist/frqlist-en.php>
14. *Horton J. C., Hocking D. I.* (1996) An Adult-Like Pattern of Ocular Dominance Columns in Striate Cortex of Newborn Monkeys prior to Visual Experience. *The Journal of Neuroscience*, 76(5):1791–1807.
15. *Horton J. C., Adams D. L.* (2005) The cortical column: a structure without a function. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* Vol. 360. p. 837–862.