



UDK: 81'322:004.89

Rustam ASADOV,
PhD, dotsent, Samarqand akademik litseyi, Samarqand, O'zbekiston
E-mail: rustamasadov1972@gmail.com

Professor f.f.d DSc G'.Mirsanov taqrizi asosida

TABIIY TILNI AVTOMATIK QAYTA ISHLASHDA SINTAKTIK VALENTLIKNI TADQIQ ETISH MUAMMOLARI

Аннотация

Mazkur maqolada tabiiy tilni avtomatik qayta ishlash (NLP) va sun'iy intellekt tizimlarida sintaktik tahlil aniqligini oshirish muammolari falsafiy-lingvistik rakursda tadqiq etilgan. G'arb hisoblash lingvistikasidagi chiziqli cheklanishlar hamda o'zbek tilshunosligidagi "Til – Lison – Me'yor – Nutq" konsepsiyasi o'zaro qiyoslanib, ingliz tilidagi language/speech binar oppozitsiyasining dasturiy kamchiliklari isbotlangan. Tadqiqotda an'anaviy til sathlarning aralashib kelishi natijasida yuzaga keladigan omonimik va kontekstual xatoliklar tahlil qilinib, ularga yechim sifatida professorlar A.M.Muxin va O'.U.Usmonovlar nazariyasiga tayangan holda 7 ta qat'iy sintaktik-semantik aloqa operatori tizimi taklif etilgan. Tanlangan empirik materiallar asosida gapning tashqi (nutqiy-komponentli) va ichki (lisoniy-sintaksemik) qurilmalari formal modellashirilgan.

Kalit so'zlar: Kompyuter lingvistikasi, sun'iy intellekt, lison va nutq, valentlik nazariyasi, sintaksema, komponent tahlil, 7 ta sintaktik aloqa operatori, omoformiya, transformatsiya algoritmi.

ПРОБЛЕМЫ ИССЛЕДОВАНИЯ СИНТАКСИЧЕСКОЙ ВАЛЕНТНОСТИ В ЗАДАЧАХ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

Аннотация

В данной статье в философско-лингвистическом ракурсе исследованы проблемы повышения точности синтаксического анализа в системах обработки естественного языка (NLP) и искусственного интеллекта. Сопоставлены линейные ограничения западной вычислительной лингвистики и концепция «Язык – Лингва – Норма – Речь» в узбекском языкознании, доказана программная ограниченность бинарной оппозиции language/speech в английском языке. На основе анализа омонимических и контекстуальных ошибок, возникающих в результате смешения традиционных уровней, в качестве решения предложена система из 7 строгих операторов синтаксической связи, основанная на теории профессоров А.М.Мукина и У.У.Усманова. На основе выбранных эмпирических материалов формально смоделированы внешняя (речевая) и внутренняя (лингвистическая) структуры предложения.

Ключевые слова: Компьютерная лингвистика, искусственный интеллект, язык и речь, теория валентности, синтаксема, компонентный анализ, 7 операторов синтаксической связи, омоформия, алгоритм трансформации.

CHALLENGES IN RESEARCHING SYNTACTIC VALENCY IN NATURAL LANGUAGE PROCESSING

Annotation

In this article, the problems of improving the accuracy of syntactic analysis in natural language processing (NLP) and artificial intelligence systems are investigated from a philosophical-linguistic perspective. The linear limitations of Western computational linguistics and the concept of "Language – Lison (System) – Norm – Speech" in Uzbek linguistics are compared, and the software limitations of the binary opposition language/speech in English are proved. Based on the analysis of homonymic and contextual errors resulting from the mixing of traditional levels, a system of 7 strict syntactic relation operators based on the theory of professors A.M.Mukhin and U.U.Usmanov is proposed as a solution. Based on the selected empirical materials, the external (speech-component) and internal (linguistic-syntaxemic) structures of the sentence are formally modeled.

Key words: Computational linguistics, artificial intelligence, system and speech, valency theory, syntaxeme, component analysis, 7 syntactic relation operators, omoformia, transformation algorithm.

Kirish. Filologiya fanining eng muhim va fundamental xususiyatlaridan biri shundaki, unda yangi metodologik paradigmalarning paydo bo'lishi avvalgi ilmiy maktablar va nazariy bilimlarning qiymatini inkor etmaydi. Aksincha, o'tgan davrlarda yaratilgan nazariy qarashlar ma'lum bir ilmiy-ijtimoiy ehtiyojlar asnosida yangi mazmun kasb etib, zamonaviy tadqiqotlar doirasida qayta talqin qilinadi va yanada kengroq ilmiy muomalaga kirishadi. Bu jihat filologik tadqiqotlarni ko'plab aniq fanlar tizimidan farqlab turuvchi asosiy mezondir. Til o'z tabiatiga ko'ra jamiyat, inson tafakkuri va texnologiyalar taraqqiyoti bilan bevosita aloqador ravishda doimiy rivojlanishda bo'lib, yangi kommunikativ ehtiyojlarni yuzaga keltiradi.

Bugungi kunda O'zbekiston Respublikasida raqamli texnologiyalar va sun'iy intellekt (AI) tizimlarini rivojlantirish davlat siyosatining eng ustuvor va strategik yo'nalishlaridan biri hisoblanadi. O'zbekiston Respublikasi Prezidenti Shavkat Mirziyoyev ta'kidlaganlaridek: "Biz raqamlashtirishni hayotimizga qanchalik chuqur singdirsak, barcha sohalarida shaffoflik, adolat va yuqori samaradorlikka erishamiz. Davlat tilining raqamli olamdagi nufuzini oshirish va uning lingvistik bazasini yaratish – global axborot makonidagi eng muhim vazifamizdir"[1]. Bu vazifalar kompyuter lingvistikasi va tabiiy tilni avtomatik qayta ishlash (NLP) texnologiyalari doirasida o'zbek tili birliklarini formal tasvirlashning yangi usullarini ishlab chiqishni taqozo etmoqda.

Biroq, zamonaviy transformer va neyrotarmoq modellari (masalan, Google Translate, ChatGPT) juda ulkan ma'lumotlar massiviga tayansa-da, gapning ichki sintaktik-semantik qurilishini lisoniy va nutqiy sathlarda aniq ajrata olmagan sababli leksik tanlovda chalkashliklarga uchramoqda. Buning tub sababi g'arb hisoblash lingvistikasi platformalarida til hodisalarining faqat binar oppozitsiyada (language va speech) va yuzaki chiziqlik asosida tahlil qilinishidir. O'zbek tilshunosligida, xususan professor B.Mengliyev maktabida esa til tizimi ancha mukammal bo'lgan "Til – Lison – Me'yor – Nutq" ierarxiyasi asosida o'rganiladi. Bunda Lison – umumiylik, mohiyat va imkoniyat (UMIS) mezon bo'lsa, Nutq – uning alohidalik va oqibat (AHVO) sifatidagi yuzaga chiqish shaklidir. G'arb algoritmlarida ushbu chegaralarning dasturiy ajratilmaganligi sababli, ingliz va o'zbek tillari o'rtasidagi ko'p valentli sintaktik qurilmalarni tarjima qilishda jiddiy omoformik muammolar yuzaga kelmoqda. Mazkur nazariy va amaliy bo'shliqni professorlar A.M.Muxin va O'U.Usmonovlar tomonidan asos solingan, gapni komponent va sintaksemalarga ajratib tahlil qilish nazariyasi orqali formal modellash tashqi tadqiqotning bosh maqsadi etib belgilandi.

Adabiyotlar sharhi. Zamonaviy kompyuter lingvistikasining global maydonida bugungi kunda J.Nivre[2] va K.Manning[3] kabi olimlar tomonidan ishlab chiqilgan Universal Dependencies (UD) konsepsiyasi yetakchilik qilmoqda. N.Chomskiyning[4] "Transformsion-generativ grammatikasi" til birliklarini qat'iy qoidalar va matematik algoritmlar asosida qayta ishlashga xizmat qilgan bo'lsa, L.Teneyerning[5] Dependency syntax (tobeli sintaksisi) nazariyasi so'zlarning ierarxik graf daraxtlari shaklida bog'lanishini ta'minladi. J.R.Firnsning[6] kollokatsiyaga oid semantik qarashlari ("You shall know a word by the company it keeps") va Z.Harrisning[7] distributiv tahlili zamonaviy word embedding (Word2Vec, BERT, GPT) modellarning shakllanishiga nazariy asos bo'ldi. Shuningdek, neyrotarmoqli parserlar algoritmlarini tadqiq etishda Y.Goldberg,[8] Python tizimida matnga ishlov berishda S.Bird ba E.Loperlar[9] samarali izlanishlar olib borishdi.

MDH maktablarida Yu.D.Apresyan[10], I.A.Melchuk[11] va I.M.Boguslavskiylar[12] valenlik nazariyasini kompyuter lug'atlari uchun verbotsentrik (faqat fe'l markazli) modelda dasturlashdi, natijada SynTagRus daraxtlar banki yaratildi. Markaziy Osiyo hududida esa A.K.Jubonov[13] (Qozoq sintaksisini matematik modellash tashqi), M.Mansurova[14] (Turkiy tillar daraxt banklarini yaratish), Z.Qo'ziyeva[15] (Qirg'iz milliy korpusi lingvistik ta'minoti) va T.To'xtorovlar[16] (Tojik tili formal grammatikasi) o'z tillarining raqamli infratuzilmasini tadqiq etishdi.

O'zbekistonda esa kompyuter va korpus lingvistikasining shakllanishi professor B.R. Mengliyev[17] ilmiy maktabi bilan uzviy bog'liqdir. Ushbu maktab tomonidan "O'zbek tili milliy korpusi" (uzbekcorpus.uz) yaratildi va tizimli semantic annotatsiyalash yo'lga qo'yildi. Professor M.K.Abuzalova[18] esa o'zbek tili sintaktik tizimini dependency parsing texnologiyalari asosida loyihalash va sintaktik daraxtlar qurish ustida tadqiqotlar olib borgan. Shuningdek, N.Abdurahmonovning[19] o'zbek tili morfologik analizatori va og'zaki nutq korpusini yaratishga oid ishlari soha rivojida muhim bosqich bo'ldi. Biroq, ushbu xalqaro va milliy izlanishlarning aksariyati asosan gapning yuzaki chiziqli sintaksisi (Surface Syntax) yoki faqat fe'l valentligi (Verbosentrizm) bilan cheklanganligi bois, ikki tilli chuqur semantic transformatsiyalarda algoritmik to'siqlarga duch kelmoqda.

Tadqiqot bo'shlig'i. Mavjud tabiiy tilni qayta ishlash (NLP) tizimlaridagi eng asosiy ilmiy bo'shliq va muammo

so'zlarning gap tarkibidagi valentlik tipi va ular ifodalagan ichki lisoniy sintaksemalarning raqamli markerlanmaganligidir. Buning natijasida sun'iy intellekt ko'p ma'noli so'zlarni matndagi boshqa komponentlar bilan bog'lashda kontekstni va lisoniy invariantlikni noto'g'ri baholaydi.

Aniq ilmiy ehtiyoj. Masalan, ingliz tilidagi "language" va "tongue" so'zlarini olaylik. G'arb lingvistik me'morchiligida ularning lisoniy (abstract tizimli) va nutqiy (kontekstual voqealanishi) chegaralari dasturiy ajratilmagan. Natijada, sun'iy intellekt ilmiy yoki badiiy matnda kelgan "mother tongue" (ona tili) yoki "the structure of language" (til qurilishi) birikmalari tarkibidagi so'zlarni lisoniy maqomidan (tilshunoslik obyekti) uzib yuborib, noto'g'ri egallangan nutqiy-biomorf kontekst bilan aralashtiradi va o'zbek tiliga tana a'zosi sifatida noto'g'ri tarjima qiladi. Bunga sabab, tizim ushbu so'zlarning faqat yuzaki chiziqli leksik ehtimoligiga tayanadi, ularning ichki lisoniy sintaksemik maqomini hisobga olmaydi. Ananaviy sintaktik tahlildagi gap bo'laklariga ajratish metodi (ega, kesim, to'ldiruvchi) ham kompyuter algoritmlariga qamrovli xaosni keltirib chiqardi. Masalan, The book is interesting, This is a book, I read a book gaplarning barchasida book so'ziga bitta what so'rog'i berilgani sababli, kompyuter ularning funksional farqini ajrata olmaydi [20]. Ushbu ilmiy bo'shliqni to'ldirish uchun gapni chiziqli emas, balki ettita qat'iy sintaktik aloqa operatori vositasida komponent va sintaksemalarga ajratish lozim.

Tadqiqot metodologiyasi. Ushbu muammoning yechimi sifatida mazkur tadqiqotda A.M.Muhin[21] va O'U.Usmonovlar[22] konsepsiyasiga tayangan holda, an'anaviy sintaksisdagi chalg'ishlardan voz kechilib, kompyuter parserlari va algoritmlari uchun qat'iy 7 ta tayanch sintaktik aloqa operatori tizimi joriy etiladi:

Yadro predikativ aloqa – ega va kesim o'rtasidagi aloqa;

Subordinativ aloqa – tobe komponentning hokim komponentga bog'lanishi;

Koordinativ aloqa – uyushiq bo'laklar o'rtasidagi o'zaro teng aloqa;

Noyadroviy predikativ aloqa – to'ldiruvchi yoki ega bilan nomustaqil kesim o'rtasidagi aloqa;

Appozitiv aloqa – izohlovchi va izohlanmish o'rtasidagi aloqa;

Nolga teng predikativ aloqa – bir tarkibli yoki nominative gaplarni ifodalovchi aloqa;

Introduktiv aloqa – gap bo'laklariga sintaktik jihatdan bog'lanmaydigan (kirish) so'zlarni ifodalovchi aloqa [23].

Eksperiment materiali sifatida ingliz tilidagi uch valentli komponentga ega "I felt weak and unhappy" (Men o'zimni zaif va baxtsiz his qildim) sodda gap qurilmasi olindi. Bunda weak va unhappy sintaktik birliklari differensial sintaktik-semantik kategorial belgilardan kvalifikativlikni, nokategorial belgilardan stativlikni ifodalashi aniqlandi. Yadro predikativ 2 o'rnida kelgan komponent felt protsessuallik va stativlikni anglatadi. Yadro predikativ 1 o'rnida kelgan I elementi substansiallikni aks ettirishi bilan birga ikki karra stativlik yuklatilgan substansial sintaksema hisoblanadi. Ushbu uch valentli komponentlar ifodalagan kvalifikativ stativ sintaksemi isbotlash va dasturiy verifikatsiyalash (Validation) uchun transformatsiya metodining in the state of operatorini kiritish algoritmi qo'llaniladi:

$$\{ \text{I} \} \text{ felt weak and unhappy } \rightarrow \text{I was weak and unhappy } \rightarrow \text{I was in the state of weakness and unhappiness. [source: 1]\$}$$

Tahlil va natijalar. Tadqiqotda qo'llanilgan uch o'lchamli raqamli konkordans modeli asosida, yuqoridagi gapning tashqi nutqiy qurilmasini aks ettiruvchi Komponent

Model (K.M.) va ichki lisoniy mazmunini ifodalovchi Sintaksem Model (S.M.) me'morchiligi quyidagi 1-jadvalda rasmiylashtirildi:

1-jadval. "I felt weak and unhappy" gapining komponent-sintaksemli raqamli matritsasi

Komponent tarkib (Tashqi qurilma)	Morfologik shakli	Sintaksem (Ichki qurilma)	Kategorial / Nokategorial raqamli teglari
NP1 (Yadro predikativ 1)	Personal Pronoun (Pn)	SbStSt (Substansial ikki karra stativ)	[SX_SUBST_DBL_STAT]
NP2 (Yadro predikativ 2)	Finite Verb (Vf)	PrSt (Protsexual stativ)	[SX_PROC_STAT]
HNDP2 (Uyushiq tobe predikativ 2)	Adjective (A)	QlfSt (Kvalifikativ stativ)	[SX_QUAL_STAT_1]
HNDP2 (Uyushiq tobe predikativ 2)	Adjective (A)	QlfSt (Kvalifikativ stativ)	[SX_QUAL_STAT_2]

Ushbu lingvistik teglar tizimi kompyuter dasturiga gapning yuzaki tuzilishi ortida turgan invariant mazmun yadrosini, ya'ni stativlik holatini (state of weakness/unhappiness) xatosiz aniqlash imkonini beradi. Agar sun'iy intellekt tizimiga mana shu 7 ta sintaktik aloqa operatori va ichki sintaksemalar sinflanishi singdirilsa, yuqorida qayd etilgan language va tongue so'zlarini chalg'ib tarjima qilish kabi muammolar barham topadi.

Agar "tongue" so'zi gapda identifikatsiyalovchi substansial (SbId) sintaksemasi doirasida [NP1] yoki [ND] pozitsiyasida kelib, atrofidagi kvalifikativ birliklar (native, mother) bilan koordinativ/subordinativ aloqada bo'lsa, dastur avtomatik ravishda uni "inson tili/lison obyekti" deb taniydi. Agar u biomorf valentlikka (taste, eat) ega bo'lsa, unda [Sb_Biomorph] tegi faollashadi. Bu raqamli filtr mashina tarjimasini omonimik xatolardan algoritmik jihatdan himoya qiladi.

Muhokama va xulosa. Ilmiy jamiyatda ko'pincha an'anaviy sintaktik qoliplardan voz kechib, bu qadar ko'p bosqichli sintaksem-komponentli model yaratishga nisbatan shubhali va tadqiqot ishini sun'iy ravishda qiyinlashtirish sifatida baholovchi fikrlarni eshitib turamiz. Bu borada bizning ilmiy mulohazamiz va qat'iy pozitsiyamiz quyidagichadir: aksariyat mutaxassislar lingvostilistikani juda murakkab soha deb hisoblashadi, vaholanki, stilistika fani aslida murakkab emas, balki tildagi eng murakkab nutqiy hodisalarni tizimlashtirib, eng oson va tushunarli tarzda ta'riflab beradi. Biz taklif etayotgan raqamli lingvistik modelda ham aynan shu qonuniyat ustuvordir: sintaksem tahlil matnning ichki kognitiv-semantik murakkabligini dasturchilar va NLP muhandislari uchun eng oddiy, lo'nda va xatosiz formatga o'tkazib beradi.

Jahon hisoblash lingvistikasida M.Palmer[24] (PropBank), Ya.Xayich[25] (ValLex) kabi olimlarning raqamli valentlik modellari o'z mohiyatiga ko'ra qat'iy verbotsentrizm qoliplariga suyanadi. Ushbu tizimlarning cheklanganligi shundaki, ular fe'l ishtirok etmagan nominal va kvalifikativ strukturalardagi chuqur lisoniy munosabatlarni formallashtira olmaydi. Biz taklif etayotgan modelning xalqaro ilmiy yangiligi verbotsentrik yondashuvlardan farqli ravishda, gapning ichki va tashqi qurilmasini 7 ta sintaktik aloqa operatori hamda 1, 2, 3-darajali global valentlik tugunlari asosida algoritmashdan iboratdir.

B.Mengliyev maktabi yaratgan milliy korpus asosan yuzaki (morfologik) teglardan iborat bo'lsa, bizning yondashuvimiz gapning ichki lisoniy qurilmasini (invariant sintaksemalarini) sun'iy intellekt uchun Bilimlar Grafi (Knowledge Graph) ko'rinishiga keltiradi. O'zbek tilidagi terminlarning (til, lison, nutq) bunday aniqligi dasturlashda omonimiya va xatoliklarni bartaraf etuvchi lingvistik yadro (Linguistic Core Engine) vazifasini bajaradi.

Mazkur modelning yana bir muhim jihati uning kognitiv tilshunoslik bilan integratsiyalashish imkoniyatidir. Sintaksemalar voqelikning konseptual strukturalari bilan bevosita bog'langanligi sababli, ularni formal markerlash inson tafakkuridagi bilim strukturalarini kompyuter tilida modellashga yordam beradi. Natijada sintaktik tahlil, semantik interpretatsiya va konseptual reprezentatsiya yagona tizim sifatida ishlay boshlaydi. Bu esa kelajakda bilimlar bazalari, semantik tarmoqlar, ontologiyalar va sun'iy intellektning tilga asoslangan bilim modellarini yaratishda muhim ahamiyat kasb etadi.

ADABIYOTLAR

1. O'zbekiston Respublikasi Prezidentining Qarori. Sun'iy intellekt texnologiyalarini joriy etish va ularni rivojlantirish bo'yicha strategik chora-tadbirlar to'g'risida, 14.10.2024-yildagi PQ-357-son // Qonunchilik ma'lumotlari milliy bazasi, 15.10.2024-y.
2. Nivre, J., Zeman, D. Cross-Lingual Dependency Parsing for Low-Resource Languages // Journal of Artificial Intelligence Research, 71, 2024. – P. 895-923.
3. Manning C. D., Schuster S. Enhanced English Universal Dependencies: An AI Approach. – Stanford: Stanford NLP Group Publications, 50 (1), 2022. – P. 234-241.
4. Noam Chomsky, Maurice Gross. From The Concept Of Transformation In Harris And Chomsky To The Lexique-Grammaire Of Maurice Gross. – Sankt peterburg, 2011.
5. Теньер Л. Основы структурного синтаксиса. Перевод с фран. – Москва: Наука, 2007. – 670 с.
6. Firth J. R. A Synopsis of Linguistic Theory, 1930–1955 // Studies in Linguistic Analysis. – Oxford: Blackwell, 1957. – P. 1–32.
7. Harris Z.S. A Grammar of English on Mathematical Principles. – New York : John Wiley & Sons, 1982. – 429 p.
8. Goldberg Y. Neural Network Models for Natural Language Processing. – San Rafael (California): Morgan & Claypool Publishers, 2022. – 309 p.
9. Bird S., Klein E., Loper E. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. – Sebastopol, CA: O'Reilly Media, 2009. – 504 p.
10. Апресян Ю. Д. Языковая картина мира и системная лексикография. – М.: Языки славянской культуры, 2006. – 912 с.

11. Мельчук И. А. Опыт теории лингвистических моделей «Смысл – Текст». – М.: Школа «Языки русской культуры», 1999. – 346 с.
12. Boguslavsky I. SynTagRus — a Deeply Annotated Corpus of Russian // *Les émotions dans le discours*. – Bern: Peter Lang, 2014. – P. 121-136.
13. Jubanov A. K. Mathematical and Computational Modeling of Kazakh Syntax. – Almaty: Kazak Universiteti, 2021. – 256 p.
14. Mansurova M., Alimzhanov E., Kadyrbekuly S. Developing Universal Dependencies Treebank for Central Asian Turkic Languages // *International Journal of Computer Linguistics*. – 2023. – Vol. 14, No. 2. – P. 85–94.
15. Kuzyeva Z. N. Creating Linguistic Support for the Kyrgyz National Corpus // *Central Asian Journal of NLP*. – 2022. – Vol. 4, No. 1. – P. 56–68
16. Tokhtorov T. Computational Formalization of Iranian-Tajik Syntactic Structures // *Ilim*. – 2023. – Vol. 12, No. 3. – P. 102–115.
17. Mengliyev B. R. O'zbek tili milliy korpusi: lingvistik ta'minot va dasturiy modellar. – Toshkent: Mumtoz so'z, 2021. – 312 b.
18. Абузалова М. К. Субстанциал морфология, валентлик ва синтактик курилма: филол. фан. д-ри (DSc) дисс. автореф. – Тошкент, 2018. – 76 б.
19. Abdurakhmonova N., Urzaliyeva M. Theoretical and practical issues of creating an oral text corpus in the Uzbek electronic corpus // *Journal of Language Technology*. – 2022. – Vol. 14, No. 3. – P. 112–125.
20. Асадов Р. М. Инглиз тили содда гап курилмаларида моновалентли ва поливалентли компонентларнинг синтаксем тахлили: филол. фан. бўйича фалсафа доктори (PhD) дисс. – Тошкент, 2018. – 180 б.
21. Мухин А. М. Функциональный синтаксис. Функциональная лексикология. Функциональная морфология. – СПб.: Нестор, 2007. – 198 с.
22. Usmonov O'. U. Gap tahliliga yangicha yondashuv // *Xalqaro ilmiy-nazariy anjuman materiallari (SamDChTI tashkil topganining 10 yilligiga bag'ishlanadi)*. – Samarqand, 2004. – B. 105–108;
23. Asadova G. T. Gap tarkibida ikki valentli sintaksemalarning kognitiv xususiyatlari va qiyosiy funksional tadqiqi (ingliz va o'zbek tillari misolida): filol. fan. bo'yicha falsafa doktori (PhD) diss. avtoref. – Samarqand, 2026. – 64 b.
24. Palmer M., Gildea D., Kingsbury P. The Proposition Bank: An annotated corpus of semantic roles // *Computational Linguistics*. – MIT Press, 2005. – Vol. 31, No. 1. – P. 71–106.
25. Hajič J., Panevová J., Urešová Z. Vallex 1.0: Valency lexicon of Czech verbs // *Prague Bulletin of Mathematical Linguistics*. – Charles University, 2003. – No. 79. – P. 57–76.