UDK 81'37:81'322.2

O‘ZBEK VA INGLIZ TILLARIDAGI FALSAFIY ILMIY DISKURSDA KO‘PMA’NOLI LEKSEMALARNING KONTEKSTUAL-SEMANTIK XUSUSIYATLARI: PARALLEL KORPUS ASOSIDA CHOG‘ISHTIRMA TADQIQ

Azizova Robiya
robiyasobitova2001@gmail.com
DOI:https://doi.org/10.5281/zenodo.23209474

Annotatsiya. Ko‘pma’nolilikni o‘rganishda leksemaning nechta ma’nosi borligini aniqlash — masalaning faqat birinchi yarmi; ikkinchi yarmi esa kontekstdagi qaysi aniq belgi u yoki bu ma’noni tanlashini ko‘rsatishdir. Maqolada o‘zbek va ingliz tillaridagi falsafiy-konseptual mazmunli ko‘pma’noli leksemalarning ma’nosi parallel korpusda qanday formal belgilar orqali ajralishi tekshiriladi. Material — Ch. Dikkensning “Oliver Tvist” asari VII–VIII boblaridan tuzilgan, gap darajasida moslashtirilgan 254 birlikli ikki tilli korpus. 10 ta leksemaning 70 ta qo‘llanishi qo‘lda annotatsiya qilinib, ma’no tanlovchi belgilarning beshta turi — leksik-sintaktik kolokat, qo‘shma birlik, frazeologik birikma, grammatik qolip va so‘z yasovchi affiks — ajratildi. Qo‘llanishlarning 98,6 % ida ma’no gap doirasidagi lokal formal belgi bilan ajraladi. Tillar o‘rtasida sezilarli farq aniqlandi: o‘zbek tilida ma’noni asosan fe’l-kolokat va frazeologik birikma (71,8 %), ingliz tilida esa grammatik qolip va qo‘shma birlik (51,6 %) tanlaydi.

Kalit so‘zlar: ko‘pma’nolilik, parallel korpus, kontekstual semantika, ma’no tanlovchi belgi, kolokatsiya, frazeologik birikma, chog‘ishtirma tilshunoslik, so‘z ma’nosini aniqlash, mashina tarjimasi.

Abstract. Establishing how many senses a lexeme has is only half of the problem; the other half is identifying which formal cue in the context selects a given sense. This paper examines, on a sentence-aligned Uzbek–English parallel corpus of 254 units from Chapters VII–VIII of Dickens’s Oliver Twist, the cues that disambiguate polysemous lexemes with philosophical-conceptual content. Seventy occurrences of ten lexemes were hand-annotated, yielding five cue types: lexico-syntactic collocate, compound, phraseme, grammatical construction and derivational affix. In 98.6 % of the occurrences the sense is fixed by a cue local to the sentence. The two languages differ markedly: Uzbek relies mainly on verbal collocates and phrasemes (71.8 %), English on grammatical constructions and compounds (51.6 %).

Kirish

Ko‘pma’nolilik tabiiy tilni qayta ishlashda noaniqlikning eng ishonchli manbai bo'lib qolmoqda [6]. Tilshunoslik leksikografiyasi leksik birlikning ma'nolarini ro'yxatlaydi, lekin A. Kilgarriff ta'kidlaganidek, lug'atdagi ma'no chegaralari ko'pincha leksikografning ichki mulohazasi mahsulidir va har doim ham so'zning matndagi haqiqiy qo'llanilishini aks ettirmaydi [2]. J. Fertning so'zni kontekstdan tushunish mumkinligi haqidagi distributiv printsipi ushbu muammoni hal qilishning boshqa yondashuvining nazariy asosidir. J. Sinclair tahliliy nuqtai nazardan matnlarda "idiom printsipi"ni "ochiq tanlov printsipi"dan ustun qo'yishimiz kerakligini

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

ta'kidlagan [7]. P. Hanks ham bu nuqtai nazarni davom ettiradi; u ma'no leksik birlikning o'zida emas, balki uning qo'llanilishida ekanligini ta'kidlaydi [9].Parallel korpus bu yerda alohida imkoniyat beradi. H. Dyvikning "semantik ko'zgu" metodiga ko'ra, A tilidagi leksemaning ma'no tuzilishi uning B tilidagi tarjima ekvivalentlari orqali aks etadi: leksema turli kontekstlarda turlicha tarjima qilingan bo'lsa, bu tarjimon amalda ajratgan ma'no farqining obyektiv dalilidir [3]. Yondashuv U. Geyl, K. Chyorch va D. Yarovskiy ishlarida asoslangan [4], N. Ayd, T. Eryavets va D. Tufish tomonidan ko'p tilli materialda sinovdan o'tkazilgan [5].

Biroq mazkur ishlarning aksariyati ma'nolarni ajratish (sense discrimination) bilan cheklanadi. Mashina tarjimasi uchun esa bundan keyingi qadam zarur: ma'no ajratilgandan so'ng, uni tanlaydigan kontekstual belgini formal tarzda ko'rsatib berish kerak[8]. Shunga ko'ra maqolaning maqsadi — o'zbek va ingliz tillaridagi falsafiy-konseptual mazmunli ko'pma'noli leksemalar ma'nosini parallel korpusda qanday formal belgilar tanlashini aniqlash, bu belgilarning tipologiyasini tuzish va ikki tilni shu jihatdan chog'ishtirishdir. Ilmiy yangiligi — ma'no tanlovchi belgilarning o'zbek-ingliz materialidagi tipologiyasi birinchi marta parallel korpus dalillari asosida tuzilganligida.

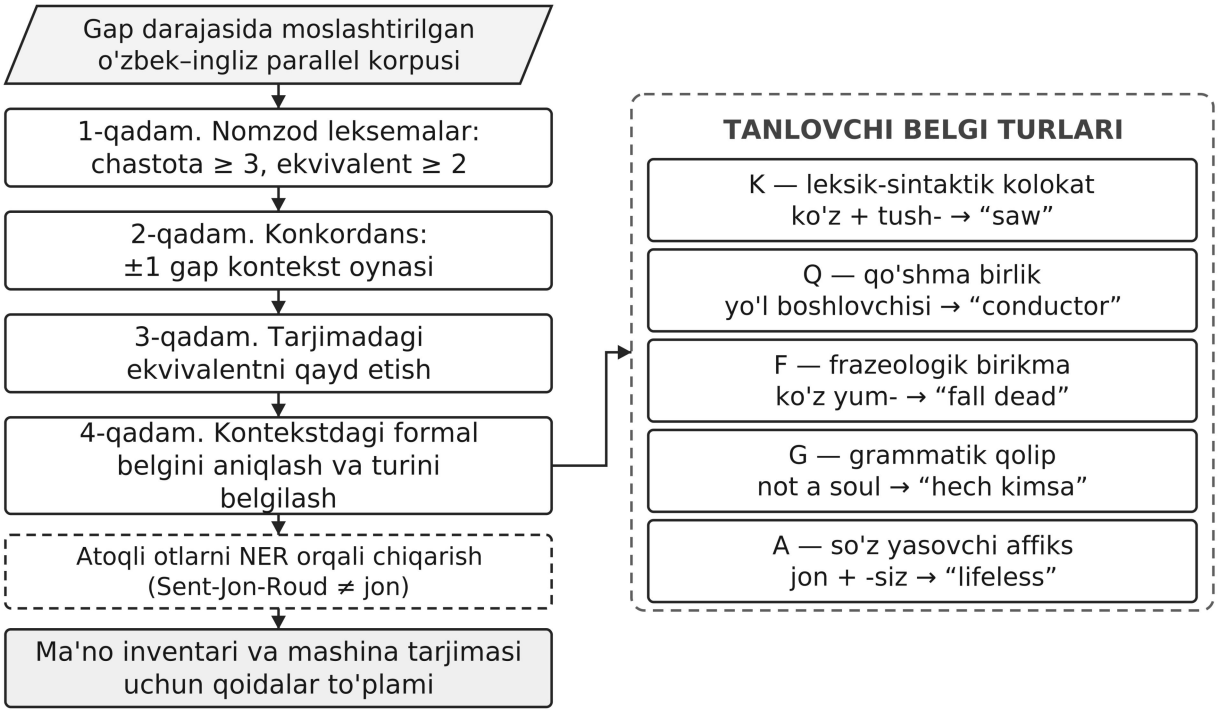
1. Material va metod

Material sifatida Ch. Dikkensning "Oliver Tvist" asari VII–VIII boblaridan tuzilgan, gap darajasida moslashtirilgan o'zbek-ingliz parallel korpusi olindi (254 juftlik). Matn badiiy-falsafiy diskursning namunasi bo'lib, unda rahm-shafqat, adolat, inson tabiati, qashshoqlik va ma'naviyat bilan bog'liq konseptual leksika zich uchraydi, shu bois ko'pma'nolilikni kuzatish uchun qulay.

Tahlil tartibi 1-rasmda keltirilgan. Dastlab chastotasi uchtadan kam bo'lmagan va kamida ikkita turlicha ekvivalent olgan leksemalar nomzod sifatida ajratildi; ularning ichidan konseptual mazmunga ega bo'lgan 10 tasi tanlandi: o'zbekcha ko'z, yo'l, jon, qalb hamda inglizcha good, light, man, heart, soul, spirit. Har bir qo'llanish uchun ±1 gap kontekst oynasida konkordans tuzilib, tarjimadagi ekvivalent qayd etildi va kontekstdagi qaysi formal element shu ekvivalentni taqozo etgani aniqlandi. Jami 70 ta qo'llanish qo'lda annotatsiya qilindi.

Atoqli otlar alohida ko'rib chiqildi: masalan, "Sent-Jon-Roud" tarkibidagi Jon leksemasi o'zbekcha jon so'zining ma'nosi emas, balki shakliy tasodufdir. Bunday holatlar nomlangan obyektlarni aniqlash (NER) bosqichida chiqarib tashlanishi lozim.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.



1-rasm. Ma'no tanlovchi kontekstual belgilarni parallel korpusdan ajratish tartibi

2. Natijalar

Annotatsiya natijasida ma'no tanlovchi belgilarning beshta asosiy turi ajratildi. Ularning tillar bo'yicha taqsimoti 1-jadvalda berilgan.

1-jadval. Ma'no tanlovchi belgi turlarining taqsimoti

Belgi turi	o'zbek, n	%	ingliz, n	%	jami	%
K — leksik-sintaktik kolokat	20	51,3	13	41,9	33	47,1
Q — qo'shma / leksikalizatsiyalangan birlik	8	20,5	5	16,1	13	18,6
F — frazeologik birikma	8	20,5	1	3,2	9	12,9
G — grammatik qolip (konstruksiya)	1	2,6	11	35,5	12	17,1
A — so'z yasovchi affiks	1	2,6	0	0	1	1,4
N — atoqli ot (NER orqali chiqariladi)	1	2,6	0	0	1	1,4

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Belgi turi	o'zbek, n	%	ingliz, n	%	jami	%
W — lokal belgi yo'q, keng kontekst zarur	0	0	1	3,2	1	1,4
Jami	**39**	**100**	**31**	**100**	**70**	**100**

Qo'llanishlarning 98,6 % ida (69 tadan 68 tasida) ma'no gap doirasidagi lokal formal belgi bilan ajraladi; faqat bitta holatda — imitation of a man birikmasida — ma'noni aniqlash uchun kengroq baho konteksti talab qilindi. Bu natija Hanksning qo'llanish qolipi haqidagi qarashini [9] ikki tilli materialda tasdiqlaydi.

O'zbek leksemalarining korpusdagi ma'no inventari va ularni tanlovchi belgilar 2-jadvalda keltirilgan. Eng boy tuzilishni ko'z (5 ma'no, 18 qo'llanish) va yo'l (7 ma'no, 14 qo'llanish) leksemalari namoyish etadi.

2-jadval. O'zbekcha ko'z va yo'l leksemalarining ma'no inventari

Leksema	Ma'no	Tanlovchi belgi (turi)	Ingliz ekvivalenti	n
ko'z	a'zo	ko'zlari + chaqna-, tik-, o'tkir (K)	eyes	4
	idrok hodisasi	ko'zi + tush- (K)	saw, observed	3
	ko'rinish, namoyon bo'lish	ko'zga + tashlan-, chalin- (K)	peeped out, to see	3
	yosh	ko'z yoshi (Q)	tears	4
	iboraviy ma'nolar	ko'z yum-, ko'zi pish-, ko'zdan qochirma-, ko'z yugurtir- (F)	fall dead, experienced, keep sight of, a look	4
yo'l	ob'ekt — qatnov yo'li	katta yo'l (K)	the road, high-road	2
	bosib o'tilgan masofa	yo'l + bos-, yur- (mil bilan) (K)	walk, come a long way	5
	safar	yo'lga tush- (K)	set forward on his journey	1
	chekka, yon	yo'l yoqasida (K)	by the side of	1
	o'tib ketayotib	yo'l-yo'lakay (Q)	as they hurried by	1
	yetaklovchi shaxs	yo'l boshlovchisi (Q)	conductor, leader	3
	duo	oq yo'l (F)	blessing	1

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Jadvaldan ko‘rinadiki, o‘zbek tilida ma’noni tanlovchi element ko‘p hollarda leksemadan keyin keladigan fe’l hisoblanadi: ko‘z + tush- idrok hodisasini, ko‘zga + tashlan- ko‘rinishni, yo‘l + bos- esa masofani bildiradi. Bu holat o‘zbek tilidagi analitik fe’l birikmalarining keng tarqalganligi bilan izohlanadi. Shu bilan birga jon leksemasi uchta mexanizmni bir vaqtda namoyish etadi: frazeologik birikma (jon-jahdi-la → “loudly”, jon-jon de- → “I do, indeed”), so‘z yasovchi affiks (jon + -siz → “lifeless”) va NER yordamida chiqarilishi lozim bo‘lgan atoqli ot (Sent-Jon-Roud).

Ingliz tilida manzara boshqacha: ma’noni ko‘pincha grammatik qolip yoki defisli qo‘shma shakl tanlaydi (3-jadval).

3-jadval. Inglizcha good, light va heart leksemalarining ma’no inventari

Leksema	Ma’no	Tanlovchi belgi (turi)	O‘zbek ekvivalenti	n
good	shaxsga ijobiy baho	good + [+inson] ot (K)	ajoyib	1
	mavhum tushunchaga baho	good + mavhum ot (K)	yaxshi	1
	xayrlashuv	good-b’ye — defisli shakl (Q)	xayr	2
	insoniy xislat	good-hearted — defisli shakl (Q)	rahmdil	1
	kuchaytiruvchi	a good + ot + of (G)	(tus ber-)	1
	noaniq miqdor	a good deal / many of (G)	sanog‘i yo‘q, ∅	3
light	tarqoq yorug‘lik	light of the candle (K)	yorug‘	1
	nur dastasi	ray of light, quyosh manbayi (K)	nur	2
	xira yolqin	light of a feeble candle (K)	shu’la	1
heart	his-tuyg‘u markazi	hissiyot oti + in his heart (K)	qalb	1
	a’zo	heart + beat (K)	yurak	1
	iboraviy	heart into his mouth (F)	(qo‘rquvdan yuragi orqasiga tortdi)	1
	jasorat	no heart to + fe’l (G)	jur’at	1

Shu qatorda soul leksemasi inkor qolipida butunlay boshqa ma’no oladi: not a soul birikmasi “hech kimsa”, ya’ni shaxs ma’nosini bildiradi va “ruh” bilan bog‘liq emas. Xuddi shunday, soul va spirit teng bog‘lovchi bilan birga kelganda (soul or spirit) tarjimon ularni farqlashga majbur bo‘ladi va “qalb, ruh”

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

juftligini tanlaydi; man leksemasi esa turnpike-man qo‘shmasida kasbni (“qorovul”), a man’s coat birikmasida yosh belgisini (“kattalar”), the man shaklida esa jinsni (“erkak”) ifodalaydi.

3. Muhokama

1-jadvaldagi taqsimot ikki til o‘rtasida tipologik farqni ko‘rsatadi. O‘zbek tilida ma’noni tanlovchi belgilarning 71,8 % i leksik darajaga — fe’l-kolokat va frazeologik birikmaga — to‘g‘ri keladi, grammatik qolip esa atigi bir holatda uchraydi. Ingliz tilida, aksincha, grammatik qolip va qo‘shma birlik birgalikda 51,6 % ni tashkil etadi: artikl, miqdor qolipi (a good deal of), inkor qolipi (not a soul), defisli birikma (good-bye, good-hearted) ma’noni bevosita belgilaydi. Bu farq o‘zbek tilining agglutinativ-analitik, ingliz tilining esa analitik-konstruksion tuzilishi bilan izchil bog‘lanadi.

Amaliy jihatdan bu natija mashina tarjimasi uchun bevosita ahamiyatga ega. Agar qo‘llanishlarning 98,6 % ida ma’no gap doirasidagi lokal belgi bilan ajralsa, demak qoidaviy komponentda ma’no tanlash uchun butun hujjat konteksti emas, cheklangan oyna yetarli bo‘ladi. Bundan tashqari, tanlovchi belgi turi qoidaning texnik shaklini ham belgilaydi: kolokat uchun fe’l-ot bog‘lanishi jadvali, qo‘shma birlik va ibora uchun ko‘pso‘zli birliklar lug‘ati, grammatik qolip uchun esa sintaktik shablon zarur bo‘ladi. Atoqli otlarni oldindan chiqarib tashlash — alohida va majburiy bosqich.

Tadqiqotning cheklovlari ochiq ko‘rsatilishi lozim. Birinchidan, qoidalar o‘sha 70 ta qo‘llanishdan induktiv tarzda chiqarilgan, shu bois 98,6 % raqami ularning bashorat qilish aniqligi emas, balki shu korpusdagi qamrovi hisoblanadi; mustaqil sinov to‘plamida baholash keyingi bosqich vazifasidir. Ikkinchidan, material bitta asar va bitta tarjimon matniga tayanadi, shu bois tarjimon individual uslubining ta’siri istisno qilinmaydi. Uchinchidan, annotatsiyani bitta tadqiqotchi bajargan; annotatorlararo kelishuv koeffitsienti o‘lchanmagan. Bu cheklovlarni bartaraf etish uchun korpusni kengaytirish, bir necha tarjima variantini jalb qilish va mustaqil annotator bilan ishlash rejalashtirilgan.

Xulosa

Parallel korpus ko‘pma’noli leksemaning ma’nolarini ajratibgina qolmay, ularni tanlovchi kontekstual belgini ham ko‘rsatib bera oladi. O‘zbek va ingliz tillaridagi 10 ta konseptual leksemaning 70 ta qo‘llanishi tahlili asosida ma’no tanlovchi belgilarning beshta turi ajratildi va ularning tillar bo‘yicha taqsimoti aniqlandi. Qo‘llanishlarning 98,6 % ida ma’no gap doirasidagi lokal belgi bilan ajraladi. O‘zbek tilida bu vazifani asosan fe’l-kolokat va frazeologik birikma, ingliz tilida esa grammatik qolip va qo‘shma birlik bajaradi. Olingan ma’no inventari o‘zbek-ingliz mashina tarjimasi tizimining qoidaviy komponenti uchun ma’no tanlash qoidalarini shakllantirishda bevosita qo‘llanilishi mumkin.

Adabiyotlar

1. Firth J.R. A synopsis of linguistic theory, 1930–1955 // Studies in Linguistic Analysis. Oxford: Blackwell, 1957. – P. 1–32.
2. Kilgarriff A. I don’t believe in word senses // Computers and the Humanities. 1997. Vol. 31, No. 2. – P. 91–113.
3. Dyvik H. Translations as semantic mirrors: from parallel corpus to WordNet // Language and Computers. 2004. Vol. 49. – P. 311–326.
4. Gale W., Church K., Yarowsky D. A method for disambiguating word senses in a large corpus // Computers and the Humanities. 1992. Vol. 26. – P. 415–439.
5. Ide N., Erjavec T., Tufiş D. Sense discrimination with parallel corpora // Proceedings of the ACL-02 Workshop on Word Sense Disambiguation. Philadelphia, 2002. – P. 61–66.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

6. Navigli R. Word sense disambiguation: A survey // ACM Computing Surveys. 2009. Vol. 41, No. 2. – P. 1–69.
7. Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991. – 179 p.
8. Atkins B.T.S., Rundell M. The Oxford Guide to Practical Lexicography. Oxford: Oxford University Press, 2008. – 540 p.
9. Hanks P. Lexical Analysis: Norms and Exploitations. Cambridge, MA: MIT Press, 2013. – 462 p.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.