

Manzura ABJALOVA,

Toshkent davlat o'zbek tili va adabiyoti universiteti dotsenti, Filologiya fanlari doktori

E-mail: abjalova.manzura@gmail.com

ORCID: 0000-0002-1927-2669

Fil.f.n., dotsent Abdiraxmanova M. taqrizi asosida

WORDNET – LINGVISTIK ONTOLOGIYALAR UCHUN TAYANCH BAZA

Аннотация

XX asrda axborot tizimlari sohasida ma'lumot qidiruv imkoniyatini oshirish, avtomatik tarjima tizimlarini takomillashtirish, matnlarni komponent tahlil qilishga erishish, til leksikonini to'liq aks ettirish maqsadida semantik tizimlar yaratila boshlandi. Jumladan, keng ko'lamli predmet sohaslarida bilimni tavsiflashga qaratilgan uch paradigma: axborot-qidiruv tezauruslari, WordNet va uning asosida yaratilgan lingvistik resurlar, ontologiyalar yaratildi. E'tiborlisi, WordNet ingliz tili leksik ma'lumotlar bazasi boshqa tillar leksikografik bazalari uchun tayanch manba vazifasini bajardi. Mazkur maqolada aynan Princeton WordNet (PWN) imkoniyatlari xususida so'z yuritildi.

Kalit so'zlar: Wordnet, Princeton WordNet (PWN), lingvistik ontologiyalar, leksik ma'lumotlar bazalari, sinsetlar.

WORDNET – A BASE FOR LINGUISTIC ONTOLOGIES

Annotation

In the 20th century, in the field of information systems, semantic systems were created in order to increase the possibility of information search, improve automatic translation systems, achieve component analysis of texts, and fully reflect the language lexicon. In particular, three paradigms aimed at describing knowledge in a wide range of subject areas were created: information-search thesauri, WordNet and linguistic resources created on its basis, and ontologies. It is noteworthy that the WordNet English lexical database served as the main source for lexicographic databases of other languages. This article is about the possibilities of Princeton WordNet (PWN).

Key words: Wordnet, Princeton WordNet (PWN), linguistic ontologies, lexical databases, synsets.

WORDNET – БАЗА ДЛЯ ЛИНГВИСТИЧЕСКИХ ОНТОЛОГИЙ

Аннотация

В XX веке в области информационных систем были созданы семантические системы с целью повышения возможности поиска информации, совершенствования систем автоматического перевода, достижения компонентного анализа текстов, полного отражения языковой лексики. В частности, были созданы три парадигмы, направленные на описание знаний в широком спектре предметных областей: информационно-поисковые тезаурусы, WordNet и созданные на его основе лингвистические ресурсы, а также онтологии. Примечательно, что лексическая база данных английского языка WordNet послужила основным источником для лексикографических баз других языков. В данной статье представлены возможности исходной лексикографической базы данных – системы Princeton WordNet (PWN).

Ключевые слова: Wordnet, Princeton WordNet (PWN), лингвистические онтологии, лексические базы данных, синсеты.

Kirish. So'zlar va ularning izohlari haqidagi ma'lumotlar azaldan an'anaviy ravishda lug'atlarda saqlanadi. Tabiiy tilni qayta ishlash bo'yicha tadqiqotlarning rivojlanishi bilan mashina (P.S. kompyuter) o'qiy oladigan lug'atlarga ehtiyoj paydo bo'ldi [Miller, 1995:39-41]. Ushbu ehtiyojlarni qondirish maqsadida leksikografik ma'lumotlarni zamonaviy hisoblash mashinalari uchun moslashtirilgan formatda so'zlar tarmoqlari yuzaga keldi. WordNet mana shunday ingliz tilidagi keng qamrovli so'zlar bazasiga ega va ularning turli izohlari hamda tavsiflariga asoslangan dastlabki mukammal, yirik lug'at hisoblanadi.

Ingliz tili imkoniyatlarini o'zida namoyon etish maqsadida yaratilgan WordNet leksik ma'lumotlar bazasi ayrim manbalarda ochiq elektron ontologiya, ayrim manbalarda tezaurus [8] deyiladi. WordNet Internet tarmog'ida foydalanish imkoniyati mavjud, leksik ontologiyalar sinfiga kiruvchi baza hisoblanadi. Shu bois uning asosidan axborot qidiruv sohasida minglab tajribalar o'tkaziladi.

Asosiy qism. Princeton WordNet (PWN) leksik ma'lumotlar bazasi (LMB)ni yaratish ishlari 1984-yilda Jorj Miller va Kristian Fillbaumlardan tomonidan boshlangan bo'lib, 1995-yildagina WordNet'dan Internet tarmog'ida erkin foydalanish imkoniyati paydo bo'ldi va u matnlarni avtomatik qayta ishlashga mo'ljallangan dasturiy ta'minotlar bo'yicha tadqiqotlarning jadallashishiga turtki berdi.

Dastlabki vaqtlarda WordNet'dan foydalanish davomida natijalar unchalik ijobiy bo'lmagan, ammo WordNet ulkan lingvistik resurslarni rivojlantirish uchun yangi davrni ochib berdi. Turli mamlakatlarda o'z milliy va davlat tillari uchun ko'plab "Wordnet"larning paydo bo'lishiga hamda kompyuter lingvistikasining mashina tarjimasini, kompyuter leksikografiyasi yoxud kiberleksikografiya [Abjalova, Erkinov, 2021:240-243], korpus lingvistikasi kabi yo'nalishlarida katta lingvistik resurslarni yaratish uchun izlanishlar hamda amaliy natijalarni yuzaga keltirishga bois bo'ldi. Aslida, WordNet inson xotirasi modeli sifatida Jorj Miller tomonidan yaratilgan. Shu o'rinda savol tug'iladi, nega aynan tilshunos emas, psixolingvistlar WordNet'ning tamal toshini qo'ydi? Javobi shunday: so'z tavsiflarini taqdim etish yuzasidan chiqarilgan ko'plab xulosalar psixolingvistik eksperimentlar bilan bog'liq bo'lgani bois inson xotirasi va miya neyronlari tarmoqlari imitatsiyasi sifatida ingliz tili uchun WordNet tarmoqli leksik ma'lumotlar bazasi ishlab chiqiladi. Ammo WordNet psixolingvistlardan ko'ra kompyuter lingvistlari qiziqishlarini uyg'otdi.

WordNet'ning unikal semantik tarmog'i foydalanuvchiga muayyan so'zning leksik munosabatlari, sinonimlari va grammatik ma'lumotlarini topish imkonini beradi.

J.Miller WordNet rivojlanishi omillarini quyidagi 3 farazda mujassamlashtirdi [8]:

ajraluvchanlik farazi: tabiiy tilning leksik tarkibiy qismi tavsifini ajratish va alohida o'rganish mumkin. Muayyan yo'nalish, masalan, mashina tarjimasini lingvistik bazasi uchun so'z turkumlarini teglash maqsadida barcha turkumlar bazasi alohidalanilishida mana shunday yirik leksik ma'lumotlar bazasi qo'l keladi;

“namuna” farazi (patterning hypothesis): tilda o‘z formal izohiga ega so‘zlar borki, bunday izohlarni tildagi aksariyat so‘zlarga qo‘llash imkoniyati mavjud. Bunday tavsiflar, asosan, ma‘nodosh so‘zlarga muvofiq keladi. Shu bois ham WordNet asosini tashkil etadi;

qamrab olish farazi (comprehensiveness hypothesis): lug‘aviy birliklar qamrovi keng elektron lug‘at. Matnlarni avtomatik qayta ishlash dasturlarida kompyuter lug‘atlaridan samarali foydalanish uchun lug‘at juda katta hajm va qiymatga ega bo‘lishi zarur hisoblanadi.

J.Millerning ushbu uch farazi lingvistik ontologiyalarni yaratish uchun asosiy tamoyillar vazifasini o‘tagan. Ammo ta‘kidlash o‘rinliki, mazkur uch tamoyil umumiy xarakterga ega bo‘lib, har bir til tabiatiga muvofiq ravishda yanada aniqroq tamoyillarni ilgari surish mumkin. Jumladan, KeNet – turk tili ontologik resursi ingliz tili leksik ma‘lumotlar bazasi asosida yaratilgan bo‘lsa-da, J.Millerning uch tamoyiliga qo‘shimcha ravishda, asosan, sinsetlarni yaratishda asosiy muammo sinset tarkibiga kiruvchi birliklarning semantik aloqasida yuzaga kelganligi sababli sinsetlardagi sinonimik munosabatlarni boshqarish uchun ikkita tamoyil qo‘shimcha qilingan: birlashish jarayoni va ajratish jarayoni [Özge Bakay, 2021:166].

E‘tiborli jihati, J.Miller o‘z farazida semantik munosabatlar dolzarbligini asosiy masala tarzida ko‘tarmagan, va holanki, tabiiy tildagi mavjud leksik birliklar qamrab olingan tizimning o‘zi lingvistik ontologiya hisoblanmaydi. Leksik birliklar o‘rtasida semantik munosabatlarning o‘rnatilishi va tizimda siklning mavjudligi leksik ma‘lumotlar bazasiga lingvistik ontologiya maqomini beradi, aks holda tizim yirik elektron leksik bazaligicha qoladi va o‘zida til hamda dunyo bilimlari simmetriyasini eks ettirmaydi. Shu bois o‘zbek tili ontologiyasini yaratishda WordNet, EuroWordNet, BabelNet, KeNet, RuTez, RussNet tizimlari imkoniyatlari hamda biznes-jarayoni tadqiq etilib, quyidagi to‘rt tamoyilni belgilab oldik [Abjalova, 2021:33]:

keng qamrovlilik: o‘zbek tilidagi mavjud leksik birliklar qamrab olinadi. Buning uchun barcha sohalardagi leksikografik manbalarga tayanish maqsadga muvofiq. Shuningdek, qamrov imkoniyatini oshirish uchun o‘zbek tili korpuslaridagi matnlarga asoslanilsa, maqsadga muvofiq. Natijada tildagi sinkretik shakllar, kvazisinonimlar, enantiosemlar, polifunksional so‘zlar va polisemantik so‘zlarni aniqlash imkoni oshadi, pragmatik tahlil natijasida o‘zbek tili ontologiyasining lingvomadaniy ma‘lumotlarni qamrab olish xususiyati yuzaga keladi.

formal til aspekti: tizimga qamrab olingan leksik birliklarning har biri o‘z turkumiga ega bo‘ladi, so‘z turkumlari teglari aniq belgilanadi, lemmatizatsiya jarayoni [Abjalova, 2020:25] amalga oshiriladi, tildagi istisnoli holatlar bartaraf etilib o‘zbek tilining formal shakli yaratiladi.

muvofiqlik: lingvistik ontologiyaga qo‘yilgan talab asosida mutloq sinonimlarda ularning izohlari muvofiqdigi belgilanadi. Kvazisinonimlar va matniy sinonimlar esa muayyan mutloq sinonim sinsetlariga kiritilgansagina shun sinset izohiga aynan keladi.

semantik simmetriya: LO va tezauruslarni boshqa lingvistik tizim va leksikografik manbalardan farqlovchi asosiy xususiyat va ularning qiymatini oshiruvchi omil semantik munosabatlar sanaladi. Shu bois lingvistik ontologiyada har bir sinset tarkibida jins-tur, butun-bo‘lak, antonimiya, omonimiya kabi munosabatlarning o‘rnatilishi leksik birliklar o‘rtasida bog‘lanishni ta‘minlaydi. Natijada tizimda yirik tarmoqlanish vujudga kelib, axborot-qidiruv imkoniyatini oshiradi.

XX asr oxirlarida semantik komponent tahlil qilishning nazariyasi mashhur bo‘lgan va unda so‘zlar ma‘nosini gap mazmuni singari sodda semantik to‘plam asosida ifodalash mumkin, deb taxmin qilingan. Lekin ko‘p yillik tadqiqotlar davomida tabiiy tilni qayta ishlash uchun maqbul resurslarda foydalanishga yaroqli sodda semantik to‘plam aniqlanmadi. Muqobil sifatida relyatsion ma‘noviy yondashuv tanlab olindi. Bunda so‘zlarning ma‘nosi ularning ayrim sememalari yordamida beriladi, turli so‘zlarning ma‘nolari o‘rtasidagi semantik munosabatni tavsiflashga asoslanilmaydi.

WordNet‘da sinsetlar o‘rni. WordNet‘ning asosiy qurilish bloklari sinset (synset)lar sanaladi. Sinset – ma‘nodosh so‘zlardan tashkil topgan, shu bois unda semantik munosabatning sinonimiya turiga asoslaniladi. Synsetlar – so‘z tarmoqlarining alohida birliklari, barcha ichki va tillararo bog‘lanishlar ham ushbu to‘plamlar asosida tuzilgan. Leksik semantikada so‘zlarni ular orasidagi bog‘liqlik asosida belgilash mumkinligi ta‘kidlanadi [Özge Bakay, 2021:166]. Mazkur fikrga tayangan holda, vordnetlarda giperonim, meronim, antonim kabi ma‘noviy munosabatlar sinsetlar orqali yuzaga chiqadi.

Sinonimlar to‘plami – sinsetlar (synsets) – WordNet‘ning asosiy tarkibiy elementlari hisoblanadi. Shu bois ingliz tili uchun ishlab chiqilgan WordNet LMB dunyoda o‘nlab boshqa tillar uchun xuddi shu kabi elektron manba yaratishga katta qiziqish uyg‘otdi. Jumladan, fincha WordNet – FinnWordNet [Linden, 2010:119-140], polyakcha WordNet [Derwojedowa, 2008:162-177], norvegcha WordNet [Fjeld, 2009:13-16], Dat tili WordNet‘i [Pedersen, 2009:269-299], fransuzcha WordNet (WOLF) [Sagot, 2008] kabi bir necha WordNet tarmoqlari yaratilgan.

Bundan tashqari, bir qator tillardagi lug‘at tarmoqlarini bog‘laydigan ko‘p tilli lug‘at tarmoqlari yaratilgan. EuroWordNet (EWN) loyihasi [Vossen, 2007] ana shunday multilingval formal ontologiyani yaratish maqsadida amalga oshirildi. Mazkur loyiha doirasida turli tillar uchun WordNetlar yaratish ikki bosiqni o‘z ichiga oldi. Birinchi bosqich (1996-1999)da golland, ispan va italyan tillari WordNet‘lari yaratildi. Ikkinchi bosqichda esa fransuz, chex, nemis va eston tillari uchun leksik ma‘lumotlar bazalari shakllantirildi. EWNda so‘z tarmoqlari har bir til uchun alohida-alohida yaratilgan, shundan so‘ng PWNga asoslangan tillararo indeks orqali bog‘langan. Loyiha mualliflari EuroWordNet va WordNet tizimi resurslarini formal ontologiya va lingvistik ontologiya sifatida farqlashadi.

So‘zlarning ma‘nosiga ko‘ra guruhlanishi jihatidan WordNet tezaurusga o‘xshaydi, shu bois ko‘p manbalarda unga nisbatan ingliz tili tezaurusi, deb baho beriladi [10]. Biroq bir nechta omillar ushbu tizimning lingvistik ontologiya ekanligini isbotlaydi [11]:

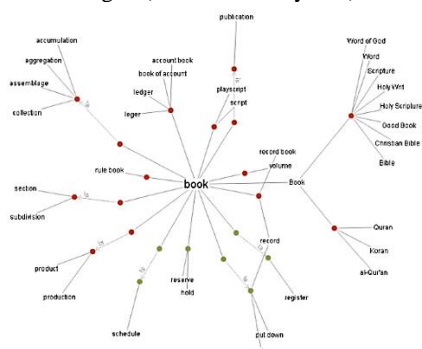
Birinchidan, WordNet nafaqat so‘z shakllari – harflar zanjirini bog‘laydi, balki so‘zlarning muayyan semalarini ham birlashtiradi. Natijada, tarmoqda yonma-yon uchraydigan so‘zlar semantik jihatdan ko‘p ma‘nolilikni bartaraf etadi.

Ikkinchidan, WordNet so‘zlar o‘rtasidagi semantik munosabatni belgilaydi, holbuki, tezaurusda hech bir shablona amal qilinmaydi, balki so‘zlar shakl va ma‘no munosabatiga ko‘ra guruhlanadi, xolos. Aniqroq aytganda, tezaurusda muayyan so‘zning ifoda va ma‘no munosabatlari belgilab beriladi, so‘zlararo ko‘ptarmoqli munosabat, ya‘ni bog‘lanish e‘tiborga olinmaydi.

WordNet bazasi ot, fe‘l, sifat va ravishlar “sinset” deb nomlangan kognitiv sinonimlar to‘plamida guruhlashtirilgan bo‘lib, ularning har biri alohida tushunchani ifodalaydi. Masalan 1.1-rasmda ko‘rganimizdek, “book” so‘zining rule book (qoida kitobi) / record book (yozuv kitobi) / volume (jild, tom) / Book (Kitob (atoqli ot)) / record (yozuvlar) / reserve (zaxira) tushunchalari mavjud. Ular, o‘z navbatida, tarkibiy ma‘nodoshlariga ega. Jumladan: accumulation (to‘plash), aggregation (jamlash), assemblage (yig‘ish), collection (to‘plash, kolleksiya) – ot turkumi; section (bo‘lim), subdivision (bo‘linma) – ot

turkumi; product (mahsulot, production (ishlab chiqarish) – ot turkumi; schedule (dastur, jadval) – ot turkumi; record book (yozuvlar kitobi) – record (yozuvlar) – ot turkumi; account book (kirim-chiqim kitobi), book of account (hisob kitobi), ledger (ro‘yxatga olish kitobi), leger – ot turkumi; playscript (ssenariy), script-publication (qo‘l yozma) – ot turkumi; reserve (zaxiraga olmoq), hold (ushlab turish) – fe‘l turkumi; put down (qo‘ymoq), enter (kiritmoq) – fe‘l turkumi; record (yozmoq) – put down (yozib qo‘shmoq), enter (kiritish) – fe‘l turkumi.

Mazkur tarmoqda “record” so‘zi ham ot turkumi, ham fe‘l turkumiga mansub bo‘lib, ot va fe‘l sinonimik qatorlarini birlashtiruvchi tugun (AKTda sinf deyiladi) hisoblanadi.



1.2.1-rasm. WordNet tizimida [9] “book” so‘zi qidiruvining tarmoqli natijasi

WordNet terminologiyasida har bir sinonimik guruhda ma‘nodosh so‘zlar qatori mavjud bo‘lib, ular bir tushunchaning lug‘aviy variantdoshlari hisoblanadi. Mazkur fikrni “book” so‘zi orqali yanada ochiqalaymiz (1.1-rasm). Rasmda ko‘rinib turganidek, WordNet tizimida sinsetlar o‘zaro yuqoridagi kontseptual-semantik va leksik munosabatlar bilan bog‘lanadi. Tarmoqda ot turkumiga oid “book” so‘zining atoqli ot shakli (“Book”), ya‘ni sinsetining 1-sinset guruhida Quran, Koran, al-Qur‘an (Qur‘on / Qur‘oni Karim) so‘zlarining har biri Koran tushunchasining leksik variantdoshlari hisoblansa, 2-sinsetdagi Word of God (Xudoning Kalomi), Word (Kalom), Scripture (Muqaddas Bitik), Holy Writ (Muqaddas Bitik), Holy Scripture (Muqaddas Bitik), Good Book (Yaxshi Kitob), Christian Bible (Nasroniyalar Injili), Bible (Injil) so‘zlarining har biri “Bible (Injil)” tushunchasining lug‘aviy variantdoshlari sanaladi.

Ingliz tilida book so‘zi ot va fe‘l turkumiga mansub shakldosh so‘z bo‘lib, kitob, qayd varaqalari to‘plami (ot) va buyurtma bermoq, zaxiraga olmoq (fe‘l) kabi tarjimalarga ega. Yuqorida ushbu so‘zning sinsetlariga binoan aytish mumkinki, book so‘zining qo‘llanilish ko‘lami keng bo‘lib, har bir turkum doirasida sinkretiklik hodisasi seziladi.

Tadqiq manbalariga tayanib aytish mumkinki, tildagi sinkretiklik hodisasi [Abjalova, 2022:193-195] polisemantik va omonimlik hodisasiga yondosh bo‘lib, sinkretik shakllar til taraqqiyoti mobaynida yuzaga keladi va uning zamiridagi qarama-qarshi ma‘nolar enantiosemiyaning shakllanishiga omil bo‘lgan.

Xulosa. Wordnet‘lardan foydalanish afzalligi shundaki, ular yordamida ma‘lumotlarni qidirish yoxud information qidiruv, matnlarni yaratish, mashina tarjimasini, matnlarni avtoreferatlash va annotatsiyalashda muayyan leksik birlikni uning lug‘aviy yoki ma‘noviy yoxud kvazisinonimiga almashirish imkoniyatini beradi.

O‘rganilgan manbalar asosida kengaytirish va birlashtirish yondashuvlari taqqoslanganda [Abjalova, 2021:44-45], kengaytirilgan yondashuvning amaliy ahamiyatga egaligi hamda vaqtni tejashi ma‘lum bo‘ldi. Buning asosiy sababi PWN bazasida bir tili munosabatlarni namoyon qiluvchi mukammal tayanch manba hisoblanadi. Chunki PWN semantik munosabatlar aniq berilgan tarmoqdir. Ushbu tarmoqdan ko‘p tillardagi semantik munosabatlar to‘plami – sinsetlarni yaratish mumkin bo‘ladi. PWNdan munosabatlarning avtomatik ravishda olinishi semantik assotsiatsiyalari ingliz tiliga o‘xshash modelni ko‘rsatadigan tillar uchun, ayniqsa, foydalidir. Shu o‘rinda tillar tipologiyasini yodga olamiz: ingliz tili hind-yevropa tillar oilasining german guruhiga mansub flektiv til hisoblanadi. Hind-yevropa tillar oilasiga, shuningdek, hind, eron, slavyan, boltiq, roman, kelt, yunon, alban, arman til guruhlari ham kiradi. Demak, PWNda kengaytirish yondashuvi yordamida, asosan, hind-yevropa tillar oilasiga mansub ko‘plab (hammasi emas, chunki har bir tilning o‘z milliy tabiati hatto ingliz tilidan ham farq qilishi mumkin) tillar uchun yaratish samarali natijalarni beradi.

ADABIYOTLAR

1. Abjalova M. O‘zNet lingvistik ontologiyasini yaratish loyihasining konsepsiyasi. // O‘zbekiston: til va madaniyat. Amaliy filologiya masalalari, – Toshkent: ToshDo‘TAU, 1/2022. – B. 123-139.
2. Abjalova M. Sinkretizm – lingvistik hodisa sifatida. O‘zbekiston Milliy universiteti xabarleri, 2022, 1/3. ISSN 2181-7324. – B.193-195.
3. Abjalova M. Tahriri va tahlil dasturlarining lingvistik modullari. [Matn] : monografiya / M.A.Abjalova. – Toshkent: Nodirabegim, 2020. – B. 25.
4. Abjalova M.A. O‘zbek tili ontologiyasi: yaratish texnologiyasi va konsepsiyasi. [Matn] : monografiya / M.A. Abjalova. – Toshkent: Nodirabegim, 2021. – 215 b. ISBN 978-9943-7804-5-3
5. Abjalova M., Erkinov F. Elektron lug‘at va kiberleksikografiya. – T.: O‘zMU xabarleri // Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti ilmiy jurnali. – Toshkent, 1/2021. – B. 240-243.
6. Derwojedowa M., Piasecki M., Szpakowicz S., Zawislawska M. and Broda B. 2008. Words, Concepts and Relations in the Construction of Polish Word-Net. In Proceedings of GWC 2008, pages 162-177.
7. Fjeld R.V. and Nygaard L. 2009. Nornet – a monolingual wordnet of modern Norwegian. In NODALIDA 2009 workshop: WordNets and other Lexical Semantic Resources – between Lexical Semantics, Lexicography, Terminology and Formal Ontologies, – P.P. 13-16.
8. http://db4.sbras.ru/elbib/data/show_page.phtml?20+1531_Тезауус WordNet
9. <http://wordnetweb.princeton.edu/perl/webwn?s=book>
10. <https://nlpub.ru/Ресурсы>
11. <https://towardsdatascience.com/%EF%B8%8Fwordnet-a-lexical-taxonomy-of-english-words-4373b541cfff>

- Linden K. and Carlson L. 2010. Construction of a FinnWordNet. *Nordic Journal of Lexicography*, 17:119-140.
- Miller G.A. 1995. WordNet: a lexical database for English. *ACM Communications*, 38: 39-41.
- Özge Bakay and others. TurkishWordNet KeNet. *Global Wordnet Virtual Conference*. 2021. January. – P. 166. https://www.researchgate.net/publication/348264475_Turkish_WordNet_KeNet
- Pedersen B.S., Nimb S., Asmussen J., Sørensen N.H., Trap-Jensen L., and Lorentzen H. 2009. DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. *Language resources and evaluation*, 43:269–299.
- Sagot B. 2008. Building a free French wordnet from multilingual resources. page 24. *ACM*.
- Vossen. P. 2007. EuroWordNet: A multilingual database for information retrieval. In *DELOS workshop on Cross-Language Information Retrieval*.
- Лукашевич Н. В. Тезаурусы в задачах информационного поиска – М., 2010. – 396 с.