

Дәріс №14

Дәріс тақырыбы: Корпус зерттеулері

Сұрақтар:

1. Корпусқа негізделген лексикографиялық зерттеулер
2. Корпусқа негізделген грамматикалық зерттеулер
3. Корпусқа негізделген дискурс зерттеулері

1. Лексикографиялық зерттеулер корпус лингвистикасы пайда болғанға дейін жүргізілді. Дегенмен корпусардың пайда болуы мұның аясын кеңітті. Осы тұста Тони МакЭнери мен Эндрю Уилсон берген дерекке көңіл бөлейік: Аткинс пен Левин 1995 жылы шыққан зерттеуінде ағылшын тіліндегі етістіктердің, соның ішінде *shake* етістігінің семантикалық тобын қарастырған (McEnery T., Wilson A. *Corpus Linguistics: An Introduction*. Second Edition. Edinburgh University Press, 2001). Atkins, B. T. S. and Levin, B. 1995. Building on a corpus: A linguistic and lexicographical look at some near-synonyms. *International Journal of Lexicography*. 8. 85–114). Мысалы:

quake - to shake because you are very frightened or find something very funny, or to feel or show great fear:

quiver - to shake slightly, often because of strong emotion:

Сонда *quake*, *quiver* етістіктері екі сөздікте те - Longman Dictionary of Contemporary of English пен Collins COBUILD Dictionary - непереходные глаголы (тура объектіні қажет етпейтін) деп көрсетілген. Ал Оксфорд сөздігі де *quake* етістігін непереходные глагол деп тіркеген, ал *quiver*-ді переходной глагол деп те тіркеген. Алайды Аткинс пен Левин 50 000 000 сөзден тұратын корпустың осы етістіктердің кездесуін/қолданысын қарай отырып, екі етістіктің де салт және сабақты етістіктер болатынына мысалдар тапқан. Тони МакЭнери мен Эндрю Уилсон сонда былайша қорытынды жасайды: «осы кішкентай мысал жеткілікті үлкен және репрезентативті корпустың қалай лексикографтардың ішкі түйсіктерін (интуицияларын) толықтыруы немесе теріске шығаруы және болашақта сөздік жазбаларын дәлдірек етуде ақпаратпен қамтамасыз етуі мүмкін екенін айқын көрсете алады».(107-б)

these verbs in a corpus of some 50,000,000 words, Atkins and Levin were able to discover examples of both *quiver* and *quake* in transitive constructions (for example, *It quaked her bowels; quivering its wings*). In other words, the dictionaries had got it wrong: both these verbs can be transitive as well as intransitive. This small example thus shows clearly how a sufficiently large and representative corpus can supplement or refute the lexicographer's intuitions and provide information which will in future result in more accurate dictionary entries.

Лексикографтар сөздердің мағынасы мен қолданысына көңіл бөледі. Дәстүрлі лексикографиялық зерттеулер сөздердің мағыналары мен синонимдерді қарастырған болса, кейіннен мұндай зерттеулер корпусқа негізделген құралдармен кеңі түсті. Сондықтан кейінгі зерттеушілер мынадай сұрақтарға көңіл бөледі:

- How common are different words?
- How common are the different senses for a given word?
- Do words have systematic associations with other words?
- Do words have systematic associations with a particular registers or dialects? (Дуглас Бибер, Сюзан Конрад, Рэнди Реппен авторлығымен шыққан «Corpus Linguistics. Investigating language structure and use» 21стр)

Лексикографиялық зерттеулер, ең алдымен, сөздіктерді құрастыру үшін қажет, сонымен қатар сипаттамалы және қолданбалы тіл білімінің қажеттіліктері үшін қажет. Мысалы,

Орыс тілінің түсіндірмелі академиялық сөздігі авторының сұрау салуының негізгі типтері мыналар:

- новое слово по времени его появления,
- исходную форму слова;
- цитаты к уже известным значениям;
- цитаты к тем значениям, которые в словаре не проиллюстрированы цитатами (чаще всего это грамматически обусловленные значения, например, страдательные формы русских глаголов или речевые употребления);
- дополнительные новые цитаты к тому или иному значению;
- новые типы лексической и синтаксической сочетаемости;
- новые фразеологизмы;
- новые современные научные толкования специальных терминов

• [9].

Д. Байбер, С. Конрад, Р. Реппен авторлығымен шыққан «Corpus Linguistics. Investigating language structure and use» [40] (далее Corpus Linguistics) еңбегінде лексикограф зерттеушілер үшін корпус тәсілдемесін ұстанғанда олардың алдарында тұратын негізгі сұрақтардың алтауын атап өткен:

1. Какие значения ассоциируются с конкретным словом?
2. Какова частотность слова относительно других близких к нему слов?
3. Какие нелингвистические модели имеет данное слово (по отношению к регистрам, историческим периодам, диалектам и т.д.)?
4. Какие слова обычно встречаются вместе с данным словом, и каково распределение этих сочетаемостных последовательностей в разных регистрах?
5. Как распределены смыслы и типы использования слова?
6. Как используются и по-разному распределяются слова, кажущиеся синонимичными?

В.Захаров пен С.Богданова оқулығында осы кітаптан **лексикографиялық зерттеуде корпусты пайдаланудың бір мысалы** алынып, көрсетілген. Соған назар аударайық. Жақсы түсіну үшін. Ол мысал ағылшын тіліндегі deal сөзінің мағыналарын анықтаумен байланысты.

Сөздердің мағынасын талдау ағылшын тіліндегі көптеген сөз формаларының бірнеше грамматикалық қызмет атқаратындығымен қиындайды. Осылайша, deals 3-жақ жекеше етістік ретінде және зат есімнің көпше түрдіне қолданылуы мүмкін. Deal мен dealing етістік ретінде де, зат есім ретінде де қолданылады. Аннотацияланбаған корпустар деректерінен құрастырылған жиілік тізімдері пайдалылығы жағынан шектеулі болады, себебі олар сөздердің қай грамматикалық қолданыстары жиі кездесетінін және қайсысы сирек кездесетінін көрсетпейді.

Керісінше аннотацияланған корпустар сөздер қолданысының жиілігін анықтауда көп пайдалы. Неге? Себебі аннотацияланған корпустарда әрбір сөз грамматикалық тұрғыдан кодталған. Сондықтан мұндай корпустарда сөздің әрбір грамматикалық формасы үшін жеке автоматты санау жүргізуге болады. Мәселен, Deal сөзінің зат есім ретінде неше рет, етістік ретінде неше рет кездесетінін анықтау үшін сөз тұлғаларды контекстке қарап, олардың грамматикалық категорияларын анықтап, содан кейін ғана санауды жүргізу керек. Мұндай шешім LOB корпусындағы deal сөзінің 182 рет қайталануы үшін өте көп уақытты қажет етеді және миллион сөзге шамамен 500 рет келетін look сияқты өте кең таралған сөздер үшін одан да көп уақытты қажет етеді. Бұл жағдайда неғұрлым дұрыс шешім - әр сөз өзінің грамматикалық категориясымен кодталған аннотацияланған корпусты пайдалану. Мұндай корпуста автоматты есептеулер сөздің әрбір грамматикалық формасы үшін жеке жүргізіле алады.

Кестеге қарайық: TACT (Text-Analysis Computing Tools) бағдарламасы арқылы алынған DEAL сөзінің жиілік тізімі берілген. Корпус - Ланкастер-Осло-Берген. Бұл аннотацияланған корпус. Сондықтан аталған сөздің әр грамматикалық тұлғасының жиілігін жеке есептеу санап береді. Deal сөзі зат есім ретінде жекеше тұлғада (граммема nn) 115 рет кездесе, жалқы есім ретінде (np) ретінде – 1 рет, етістік ретінде (vb) – 66 рет т.б. кездескен.

Таблица 5

**Частотность форм DEAL в аннотированном корпусе
Ланкастер-Осло-Берген (LOB)**

deal_nn	115
Deal_np	1
deal_nns	5
deal_vb	66
dealing_vbg	51
deals_vbz	20
dealt_vbd	14
dealt_vbn	17

DEAL-дің регистрлер бойынша бөлінісі

Сөздер көбіне әртүрлі регистрлерде әртүрлі қолданылады, сондықтан сөздің жалпы сипаттамалары тілдегі нақты жағдайды көрсетпеуі мүмкін. Сіздер білесіздер корпустарда түрлі мәтіндер жинақталады. Баспасөз, ғылыми әдебиет, көркем әдебиет, т.т. сияқты әр түрлі регистрден алынған мәтіндер бар. Сонда талдауға DEAL лексемасының Ланкастер-Осло-Берген аннотацияланған корпусындағы зат есім ретіндегі қолданысын аламыз, оның жекеше және көпше түрлерін (deal және deals) ғана қарастырамыз. Ланкастер-Осло-Берген корпусында әртүрлі регистрдегі мәтіндер бар.

6-кесте тікелей санаулар (непосредственные (сырые) (raw counts) мен нормаланған санауларды (нормированные подсчеты (normed counts) қамтиды. **Тікелей** санау сөздің әрбір регистрдегі кездесу жағдайларының санын көрсетеді: deal зат есімінің 14 кездесу жағдайы баспасөз репортаждарында, 4 кездесу жағдайы баспасөзде шолуларда, 16 кездесу жағдайы ғылыми әдебиетте және т.б. кездескен.

Таблица 6

**Частотность существительного DEAL в определенных регистрах,
нормированная на 100 тыс. слов**

Регистр	Примерное количество слов в подкорпусе	Непосредственные (сырые) подсчеты для DEAL	Нормированные подсчеты для DEAL (на 100 тыс. слов)
Репортажи прессы	88000	14	15,9
Обзоры прессы	34000	4	11,8
Передовицы	54000	4	7,4
Религиозная литература	34000	5	14,7
Научная литература	160000	16	10,0
Научно-популярная литература	88000	11	12,5
Беллетристика	154000	24	15,6
Художественная проза	58000	5	8,6

Дегенмен, бұл ішкі корпустар сөз саны жағынан бірдей емес. Мысалы, баспасөз репортаждарында - 88 мың сөз, баспасөзге шолуларда - 34 мың сөз, ал ғылыми әдебиетте - 160 мың сөз. Осы себепті, тікелей санауларды екінші регистрмен салыстырғанда сөздің бір регистрдегі көп немесе аз жиілігі туралы қорытынды шығаруда критерий ретінде пайдаланылуға болмайды. Өйткені бір регистрдегі мәтіннің үлгісін үлкейтіп талдай бастасаңыз, ізделетін сөзді қолданудың басқа жағдайлары да ашылуы әбден мүмкін. Баспасөз материалдарындағы DEAL үшін тікелей санаулар ғылыми әдебиеттегіден төмен болады деп күтуге болады, себебі баспасөзде репортаждарынан әлдеқайда аз ішкі корпус талданған және бұл DEAL баспасөзде репортаждарында сирек кездеседі дегенді білдірмейді. Керісінше, оқырманның зат есім ретінде DEAL-ді қаншалықты жиі кездестіретінінің өлшемі қажет.

Мұндай өлшемдер **нормаланған жиілік** деп атайды. **Нормаланған санаулар** қолданылу жағдайларының санын стандартты шкалаға сәйкес түрлендіреді, әдетте 1 млн сөзқолданысты пайдалануды санау тұрғысынан немесе мына талдау жағдайда 100 000 сөз қолданысты санаумен. Санаулар нормаланған болғанда, (кестенің соңғы бағаны) баспасөз репортаждарында 100 мың сөзге 15,9 кездесу жағдайы, ал ғылыми әдебиетте 100 мың сөзге 10 жағдай. Сәйкесінше тек нормаланған санаулар регистрлер бойынша салыстыруда сенімді негізін қамтамасыз етеді.

DEAL зат есімінің кездесу жағдайларын регистрлер бойынша бөліп қарағанда, лексикографиялық жұмыстар үшін корпустың көлемі мәселесі шығады. 6-кестеден 8 ішкі корпустың 4-уінде барлығы 4-5 кездесу жағдайы анықталған. Ешбір регистрде DEAL зат есімінің өте үлкен қолданыс саны жоқ, максималды сан көркем прозада – 24. Ланкастер-Осло-Берген корпусы DEAL лексемасының зат есім ретінде қолданысын егжей-тегжейлі талдауға өте аз, сондықтан авторлар бұдан гөрі көлемі үлкен Лонгман-Ланкастер корпусын қарайды (оқулықтан өздеріңіз оқисыздар).

Таблица 7

Частотность существительного и глагола deal в подкорпусах из двух регистров корпуса Лонгман-Ланкастер

	Примерное количество слов в подкорпусе	Нормированные подсчеты на 1 млн. слов (IPM)	
		существительное	глагол
Всего	4000000	90	119
Художественная проза	2000000	107	63
Научная литература	2000000	74	176

Таблица 7 показывает, что в корпусе Лонгман-Ланкастер deal и deals встречаются намного чаще, и это обеспечивает более солидную базу для анализа их употребления. Благодаря данным этой таблицы частотности становятся очевидными несколько интересных моделей.

2. Корпусқа негізделген грамматикалық зерттеулер.

Оқуға: Douglas Biber (2005). Corpus Linguistics and the Study of English Grammar // Indonesian Journal of English Language Teaching, Volume 1/Number 1

Үлкен корпустар мен компьютер құралдарының бар болуы сөйлеушілердің тілдің грамматикалық ресурстарын қолдануының үлгілерін зерттеуге мүмкіндік берді. Олар - әртүрлі құрылымдардың таралу жиілігін, грамматикалық құрылымдар мен басқа тілдік және тілдік емес факторлардың арасындағы байланыс үлгілерін, құрылымдық варианттар арасындағы таңдауға әсер ететін факторларды зерттеу мүмкін болды. Корпусқа негізделген зерттеу грамматикада сөз деңгейінде, сөйлем деңгейінде және дискурс деңгейінде

қолданыла алады. (Д.Бибер, С.Конрад және Р.Реппен «Corpus Linguistics: Investigating Language Structure and Use» 1998).

Корпусқа негізделген грамматикалық зерттеулерде регистрлер бойынша морфологиялық сипаттамалардың таралуын талдау арқылы олардың қолданылу және қызметі мәселесі қарастырылады (В.Захаров пен С.Богданова). Д.Бибер бойынша сөздердің қолданысы (мысалы, ағылшын тіліндегі ең жиі қолданылатын етістіктер), грамматикалық формалардың қолданысы мен бөлінісі/таралуы (мысалы, the relative frequency of simple, progressive, and perfect aspect in English) және үшінші, лексика мен грамматикалық құрылымның күрделі жолдармен өзараәрекеттесуі (етістіктің валенттілігі) қарастырылады. Корпустың көмегімен морфологиялық сипаттаманың таралуын оны қолдану контекстімен салыстыруға және оның атқаратын қызметтерін жақсырақ түсінуге болады. «Corpus Linguistics» оқулығында бұл мәселені шешу жолдары номиналданудың (туынды зат есімдердің) үш регистрге таралуы арқылы талданған.

Номиналдану деп орыс тіл білімінде етістіктен абстрактылы зат есімнің және осындай жолмен зат есімнің өзінің жасалу процесін ұғынса, еуропалық тіл білімінде бұл түсінік кеңірек сын есімнен зат есімнің жасалуы да номиналдану болады. Мысалы, *civilization* номиналдану *civilize* етістігінен жасалған, *kindness* – номинализацией, производной от прилагательного *kind*.

«Корпус лингвистикасы» оқулығында төрт өнімді жұрнақ талданды: -tion/-sion, -ness, -ment, -ity (және олардың көпше түрі). Арнайы бағдарламалардың көмегімен корпус мәтіндерін автоматты түрде өңдеуден басқа, форма бойынша іздеу үлгісіне сәйкес келетін, бірақ номинализацияға жатпайтын (*mansion, nation, city*) бірліктерді таңдау үшін қолмен өңдеу де қажет болды.

Номиналдануды талдау үш регистрде жүргізілді. Алғашқы екеуі - ғылыми әдебиет және көркем проза - Лонгман-Ланкастер корпусының қосалқы корпустарымен ұсынылған. Үшінші регистр – ауызша сөйлеу – Лондон-Лунд корпусымен (500 мың сөз қолданысының көлемі) ұсынылған. Мұндағы барлық жиіліктер мәтіннің 1 миллион сөзіне дейін нормаланған. (бұл санаудың қай түрі?).

60 Investigating the use of language features

Table 3.1 Frequency distribution of nominalizations across three registers (academic prose and fiction from the Longman–Lancaster Corpus; speech from the London–Lund Corpus)

	Academic prose (2.7 mill. wds)	Fiction (3 mill. wds)	Speech (.5 mill. wds)
Nominalizations per million words	44,000	11,200	11,300

the advantages and disadvantages of concordancing packages in comparison to individually designed computer programs.

The analysis of nominalizations was conducted on three regis-

Частотные распределения номинализаций по трем регистрам

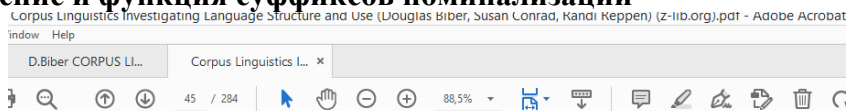
	Научная литература (2,7 млн. с/у)	Художественная проза (3 млн. с/у)	Устная речь (0,5 млн. с/у)
Количество номинализаций на 1 млн. с/у	44 000	11 200	11 300

Таблица 11 показывает частотные распределения для номинализаций по всем трем регистрам. Очевидно, что в текстах художественной прозы и устной речи отмечаются близкие частотности, а в текстах научной литературы частотность номинализаций в четыре раза больше. Можно попытаться объяснить, почему регистры имеют такие разные распределения, исследуя наиболее частые формы в контексте. Авторами учебника *Corpus*

Регистре жиі кездесетін нақты номинациялар корпус мәтіндерінде қарастырылған тақырыптарға байланысты. Мысалы, ғылыми әдебиеттерде (academic prose) жиілігі миллион сөзге 500-ден асатын алты номинализация басқаларға қарағанда әлдеқайда жиі кездеседі: movement (миллион сөзге 900-ге жуық кездесу), activity, information, development, relation және equation. Керісінше, номинациялардың ешқайсысы көркем әдебиетте де, ауызша сөйлеуде де бұл жиілікпен кездеспейді. Мысалы, movement көркем әдебиетте миллион сөзге шамамен 100 рет және ауызша сөйлеуде шамамен 60 рет кездескен, көркем әдебиетте development миллион сөзге 10 рет қана жүреді және ауызша сөйлеуде жоқ.

Осы алты номинализацияның конкорданстарын талдау ғылыми әдебиеттегі номинациялар әрекеттер мен процестерді адамның қатысуынсыз абстрактілі субъектілер ретінде сипаттайтынын көрсетеді. Бұл үлгі movement номинализациясының мысалында көрінеді: The legs and hips, or arms and shoulders, may be used to initiate movement in any direction. Ғылыми әдебиет мәтіндерінде қандай да бір субъектінің қозғалысы емес, жалпылама қозғалыс әрекеті талқыланады. Көркем әдебиет пен ауызекі сөйлеу тілі, керісінше, адамға бағытталған, сондықтан бұл регистрлерде адамның мінез-құлқын сипаттау үшін етістіктер мен сын есімдер жиі қолданылады. Осылайша, бұл регистрлерде көбінесе іс-әрекет субъектілері ретінде нақты адамдар болады және оларда move етістігі жиі қолданылады: Garth whistled breathily to himself and moved his hand crabwise along the table (fiction); It's how much they move it that counts (spoken).

Распределение и функция суффиксов номинализаций



Grammar 63

Table 3.2 Proportion of nominalizations formed with each suffix

	Academic prose	Fiction	Speech
-tion/-sion	68%	51%	56%
-ment	15%	21%	24%
-ness	2%	13%	5%
-ity	15%	15%	15%

Таблица 12

Пропорции номинализаций с каждым суффиксом			
Суффикс	Научная литература	Художественная проза	Устная речь
-tion/-sion	68%	51%	56%
-ment	15%	21%	24%
-ness	2%	13%	5%
-ity	15%	15%	15%

Таблица 12, не показывая действительные частотности, демонстрирует относительные пропорции номинализаций с каждым суффиксом в каждом регистре.

1. Хотя суффикс -tion/-sion встречается в большинстве номинализаций во всех трех регистрах, его пропорция намного выше

1. -tion/-sion жұрнағы барлық үш регистрдегі атаулы атаулардың көпшілігінде кездесе де, ғылыми әдебиеттерде оның үлесі әлдеқайда жоғары (68%).

2. -ment жұрнағы ғылыми әдебиеттерге қарағанда ауызша сөйлеу мен көркем әдебиетте көбірек кездеседі.

3. Көркем әдебиетте басқа екі регистрдің екеуіне қарағанда - ness жұрнағы жиі кездеседі.

1.3.4.2. Распределение грамматических категорий (өздеріңіз оқысыздар)

Biber D. 2005. Corpus Linguistics And The Study Of English Grammar // Indonesian Journal of English Language Teaching Volume 1/Number 1, 1-22

In the present paper, I take a strong position on the importance of register for studies of grammar and use, arguing that most functional descriptions of a grammatical feature will not be valid for the language as a whole. Rather, characteristics of the textual environment interact with register differences, so that strong patterns of use in one register often represent only weak patterns in other registers. Thus, a complete functional analysis must consider the patterns of use across registers.

Бұл мақалада мен грамматикалық белгілердің функционалдық сипаттамаларының көпшілігі жалпы тіл үшін жарамсыз деп есептей отырып, грамматиканы үйрену және қолдану үшін регистрдің маңыздылығын берік ұстаным ретінде ұстанамын. Керісінше, мәтіндік ортаның сипаттамалары регистрлік айырмашылықтармен өзара әрекеттеседі, осылайша бір регистрде күшті пайдалану үлгілері көбінесе басқа регистрлерде әлсіз үлгілерді көрсетеді. Осылайша, толық функционалдық талдау регистрлерді пайдалану үлгілерін ескеруі керек.

Д.Бибер осы мақаласында өзінің 1999 жылғы еңбегіне сілтеме жасай отырып грамматиканың, қолданудың және регистрдің өзара әрекеттесуін Лонгманның ауызша және жазбаша ағылшын тілінің грамматикасынан бейімделген корпус негізіндегі талдаулармен суреттейді (Biber, et al., 1999). Үш кейс зерттеу ұсынған, олардың барлығы етістіктерді қолдануға бағытталған: бірі лексикалық үлгілерге қатысты (яғни, ағылшын тіліндегі ең көп таралған етістіктер, 2-бөлім); екіншісі грамматикалық формалардың қолданылуы мен таралуына назар аударады (яғни, ағылшын тіліндегі simple, progressive және perfect aspect-лердің салыстырмалы жиілігі, 3-бөлім); ал үшінші лексика мен грамматикалық құрылымның күрделі тәсілдермен өзара әрекеттесуін сипаттайды (яғни, бірдей валенттілік потенциалы бар етістіктердің белгілі валенттіліктерге таңқаларлық әр түрлі қалаулары болуы мүмкін екенін көрсетеді, 4-бөлім).

Талдаулар төрт регистрдегі мәтіндерге негізделген: әңгіме/conversation, көркем әдебиет/ fiction, газет тілі және академиялық проза. Бұл жалпы регистрлер болғанымен,

олар бір-бірінен маңызды жолдармен ерекшеленеді (мысалы, типі, интерактивтілігі, өндірістік жағдайлары, мақсаты және мақсатты аудиториясы бойынша). Талдаулар жалпы 40 миллион сөзден тұратын мәтіндері бар Longman Spoken and Written English (LSWE) корпусында жүргізілді. Осы төрт регистрдің әрқайсысында 4-5 миллион сөз бар (1-кестені қараңыз). Төменде келтірілген барлық жиілік санаулары жалпы қалыпқа/негізге - мәтіннің 1 миллион сөзіне есептелу - келтірілді, осылайша олар регистрлер бойынша тікелей салыстырылады.

	Number of texts	Number of words
Conversation (BrE)	3,436	3,929,500
Fiction (AmE and BrE)	139	4,980,000
News (BrE)	20,395	5,432,800
Academic prose (AmE and BrE)	408	5,331,800

THE MOST COMMON LEXICAL VERBS ACROSS REGISTERS

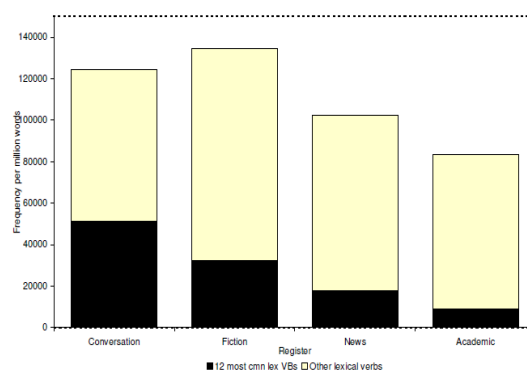
Регистрлер бойынша ең кең таралған лексикалық етістіктер

Ағылшын тілінде ең жиі таралған ондаған лексикалық етістіктер бар. Мысалы, LSWE корпусында 400-ге жуық әртүрлі етістік формалары бір миллион сөзге 20-дан астам рет кездеседі (қараңыз: Biber et al., 1999: 370-371). Оларға *pull*, *throw*, *choose*, and *fall* сияқты көптеген күнделікті етістіктер жатады.

Салыстырмалы түрде кең таралған етістіктердің осы үлкен тізімін ескере отырып, жекелеген етістіктердің ешқайсысы ерекше кең таралған деп болжам жасау оңай сияқты. Алайда бұл мүлде олай емес: LSWE корпусында регистрде миллион сөзге 500-ден астам рет кездесетін 63 лексикалық етістік бар, ал миллион сөзге 1000-нан астам рет кездесетін 12 етістік қана бар (Biber et al., 1999:367-378). Бұл ең көп таралған 12 етістік: *say*, *get*, *go*, *know*, *think*, *see*, *make*, *come*, *take*, *want*, *give*, and *mean*.

Осы 12 етістіктің маңыздылығын көрсету үшін 1-сурет олардың жиынтық жиілігін барлық басқа етістіктердің жалпы жиілігімен салыстырады. Топ ретінде қарастыратын болсақ, бұл 12 етістік сөйлесуде/ conversation ерекше маңызды, мұнда олар барлық лексикалық етістіктердің көрінуінің 45%-ын құрайды.

Figure 1: Distribution of the most common lexical verbs vs. other verbs, across registers (based on Biber et al., 1999, Figure 5.8)



Бұдан әрі бұл 12 етістіктің өз ішінде жалпы және олардың регистрлік таралуларында үлкен жиілік айырмашылықтары бар екені белгілі болды. Мысалы, 2 және 3-суреттер

сөйлесудегі/ conversation және газет тіліндегі әрбір етістіктің жиілігін көрсетеді (Cf Biber et al., 1999:374-376). *Say* етістігі бұл суреттерде бірінші болып берілген, себебі ол ауызша және жазбаша регистрлерде жиі кездеседі және сондықтан ең жоғары жиілікке ие. Бөгделердің сөзін хабарлаудың барлық жерде қажеттілігін ескерсек, бұл таңқаларлық емес; шешендер де, жазушылар да бұл мақсат үшін, әдетте, өткен шақта төл сөз немесе бөгде сөзді беруде, жалғыз *say* етістігіне көп сүйенеді.

For example:

You said you didn't have it. (conversation, henceforth CONV)

He said this campaign raised 'doubts about the authenticity of free choice'. (news)

Сөйлесудегі *get* етістігінің өте жоғары жиілігі көбірек таңқаларлық. Бұл етістік негізінен байқалмайды, бірақ әңгімеде ол кез келген регистрдегі ең көп таралған жалғыз лексикалық етістік болып табылады. *Get* сөзінің кең тарағанының басты себебі - оның өте жан-жақты болуы, кең мағынада қолданылуы. Мысалға:

Obtaining something:

See if they can *get* some of that beer.

(CONV)

Possession:

They've *got* a big house. (CONV)

Moving to or away from something:

Get in the car. (CONV)

Causing something to move or happen:

It *gets* people talking again, right? (CONV)

Understanding something:

Do you *get* it? (CONV)

Changing to a new state:

So I'm *getting* that way now. (CONV)

Figure 2: Frequencies in conversation of the most common lexical verbs (based on Biber et al, 1999, Figure 5.9)

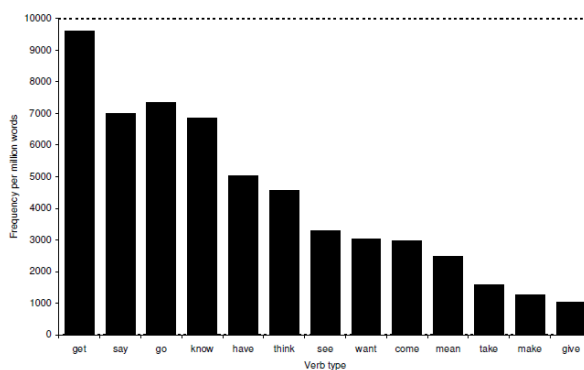
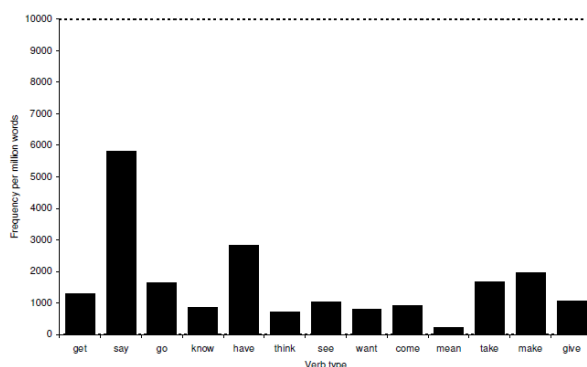


Figure 3: Frequencies in Newspapers of most common lexical verbs (based on Biber et al, 1999, Figure 5.11)



Басқа бірнеше етістіктер де сөйлесуде//conversation өте жиі кездеседі: *go*, *know* және аз дәрежеде *think*, *see*, *come*, *want*, and *mean*. Жаңалықтар, керісінше, мүлдем басқа үлгіні көрсетеді, тек *say* етістігі өте жиі болады. Дегенмен, осы 12 етістіктің барлығы ағылшын тіліндегі көптеген етістіктермен салыстырғанда екі регистрде де жиі кездесетінін атап өткен жөн. Мысалы, жоғарыда атап өткеніміздей, *pull*, *throw*, *choose*, және *fall* сияқты етістіктер миллион сөзге шамамен 50-100 рет қана кездеседі. Басқа сансыз етістіктердің жиіліктері одан да төмен. Керісінше, ең көп таралған 12 етістіктің көбі conversation де, жаңалықтарда да миллион сөзге 1000-нан астам рет кездеседі.

Осылайша, етістіктерді қолдануда клин бар: аздаған етістіктер өте жоғары жиілікте кездеседі; бірнеше етістік орташа жоғары жиілікте кездеседі; ал етістіктердің көпшілігі салыстырмалы түрде төмен жиіліктерде кездеседі. Сонымен қатар, әртүрлі регистрлер белгілі бір етістіктерге арналған таңқаларлық әртүрлі артықшылықтарды көрсетеді. Мысалы, *get*, *go*, *know*, және *think* етістіктері жаңалықтарға қарағанда сөйлесуде әлдеқайда жиі кездеседі (2 және 3-суреттерді қараңыз). Керісінше, *add*, *spend*, *claim*, және *continue* сияқты етістіктер сөйлесуге қарағанда жаңалықтарда әлдеқайда жиі кездеседі.

3. Корпусқа негізделген дискурс зерттеулері

Оқуға:

- 1) Corpora and Discourse. The challenges of different settings. Edited by Annelie Ddel and Randi Reppen. Volume 31. John Benjamins Publishing Company. Amsterdam / Philadelphia
- 2) McEnery T., Wilson A. Corpus Linguistics: An Introduction. Second Edition. Edinburgh University Press, 2001

4	The Use of Corpora in Language Studies	103
4.1	Corpora as Sources of Empirical Data	103
4.2	Corpora in Speech Research	104
4.3	Corpora in Lexical Studies	106
4.4	Corpora and Grammar	109
4.5	Corpora and Semantics	112
4.6	Corpora in Pragmatics and Discourse Analysis	114