Дәріс №13
Дәріс тақырыбы: Корпус зерттеулері
Сұрақтар:

1. Корпустарды қолданушылар
2. Корпустарды қолдану жолдары

1. Корпустарды тілді зерттеулерде қолдану эмпирикалық деректерді қолданумен тығыз байланысты. Тони МакЭнери мен Эндрю Уилсон корпустарды «эмпирикалық деректердің көзі» деп таниды. (Tony McEnery, Andrew Wilson 'Corpus Linguistics'). «Тілді зерттеуде корпустың маңыздылығы жалпы алғанда эмпирикалық деректердің маңыздылығымен тығыз байланысты. Эмпирикалық деректер лингвиске тілге қатысты өзінің жеке когнитивті қабылдауына негізделген субъективті мәлімдемелер жасағаннан гөрі тілге негізделген объективті мәлімдемелер жасауына мүмкіндік береді».

Корпустарды пайдаланушылар, ең алдымен, лингвистер, әдетте, нақты мәтіндердің мазмұнына емес, олардың метамәтіндік ақпаратына және белгілі бір тілдік элементтер мен конструкцияларды қолдану мысалдарына қызығушылық танытады (В. Захаров, С.Богданова. Корпусная лингвистика). Корпустар көмегімен жүргізілген алғашқы лингвистикалық зерттеулер әртүрлі тілдік элементтердің кездесу жиілігін санаумен шектелген. Статистикалық әдістер күрделі лингвистикалық міндеттерді шешуде қолданылады, ол қандай міндеттер: машиналық аударма, сөйлеуді тану және синтездеу, емлені және грамматиканы тексеру құралдары т.б. сияқты. Осылайша, корпус материалы негізінде статистикалық әдістерді пайдалана отырып, қандай сөздердің тұрақты түрде кездесетінін және осылайша, оларды тұрақты сөз тіркестеріне жатқызуға болатынын анықтауға болады. Корпустар лексикография мен грамматикадағы зерттеулер үшін бай деректер көзі болып табылады. Семантика саласындағы зерттеулер лексикографиядағы зерттеулермен тығыз байланысты. Корпустағы белгілі бір тілдік бірліктердің ортасын бақылай отырып, осы бірлікті сипаттайтын белгілі бір мағыналық белгілерді белгілеуге болады.

Теоретик лингвистер өз теорияларының гипотезаларын тексеру және дәлелдеу үшін корпусты эксперименттік жүйе ретінде пайдаланады. *Қолданбалы лингвистер* (оқытушылар, аудармашылар және т.б.) компьютерлік корпустарды тілдерді үйрету және өздерінің кәсіби міндеттерін шешу үшін пайдаланады. *Компьютер лингвистері* тілдің компьютерлік үлгілерін жасау үшін мәтіндерде бар статистикалық және тілдік заңдылықтарды анықтауға және пайдалануға тырысады. Корпусты пайдаланушылар осылайша жалғаса береді. Тілшілер ғана емес, әдебиетшілер де пайдаланады.

Жалпы теориялық лингвистикаға корпус деректері не береді деген сұраққа В.Захаров пен С.Богданова былай дейді:

> «Корпусы могут в принципе дать три типа данных, которые могут быть использованы в ходе лингвистических исследований: эмпирическая поддержка, информация по частотности, экстралингвистическая информация (метаинформация). Рассмотрим эти типы данных более подробно». (96-б) (толық оқулықтан өздеріңіз қарайсыздар)

2. Корпустарды қолдану жолдары
Корпустарды магистрлік зерттеу жұмыстарына қатысты қолдануға үйрету

1. Сөздердің қолданылу жиілігі
2. Конкорданс
3. Коллакация

1. Зерттеу тақырыбыңызға қатысты кез-келген сөзді іздеуге салып, қолданылу жиілігін анықтаңыз.

2. Конкорданс - Бағыттау материалы

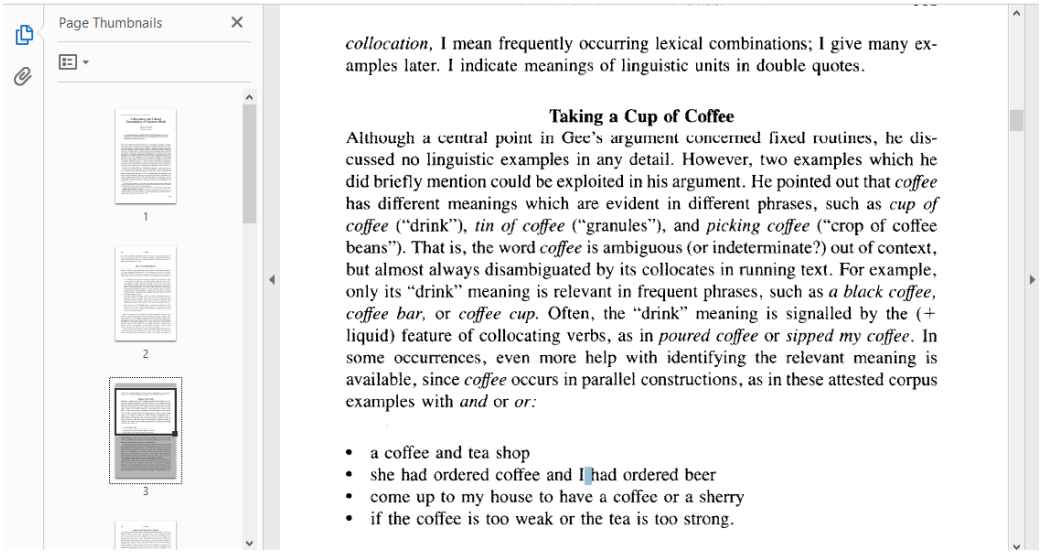
Конкорданс - контекстегі кілт сөз (KWIC) деп те аталады, конкорданс - бұл белгілі бір іздеу терминінің корпустағы барлық кездесулерінің тізімі, олар контексте орын алады, әдетте іздеу терминінен солға және оңға бірнеше сөзбен ұсынылады. Іздеу термині көбінесе жеке сөз, бірақ көптеген конкорданс бағдарламалары пайдаланушыларға бірнеше сөз тіркестерін, тегтерді немесе сөздер мен тегтердің комбинацияларын іздеуге мүмкіндік береді. Конкорданстарды әдетте іздеу терминінің өзіне немесе іздеу терминінің сол немесе оң жағындағы х орындарына әліпби бойынша сұрыптауға болады, бұл лингвистикалық үлгілерді адамдарға оңайырақ байқауға мүмкіндік береді.

Конкорданспен қалай жұмыс жасау керектігіне кеңес:

Көптеген конкорданс іздеулер жүздеген немесе мыңдаған жолдарды шығара алатындықтан, Синклер (1999) 30 кездейсоқ жолды таңдауды және олардағы үлгілерді белгілеуді, содан кейін басқа 30 жолды таңдауды, жаңа үлгілерді атап өтуді және т.т., 30 жолды одан әрі таңдау жаңа ештеңе ашпайынша талдауға кеңес береді. (Paul Baker, Andrew Hardie And Tony McEnery. A Glossary of Corpus linguistics. Edinburgh University Press, 2006, 42-43)

3. **Коллакация -Бағыттау материалы**

Michael Stubbs. Collocations and Cultural Connotations of Common Words // Linguistics and Education 7, 379-390.

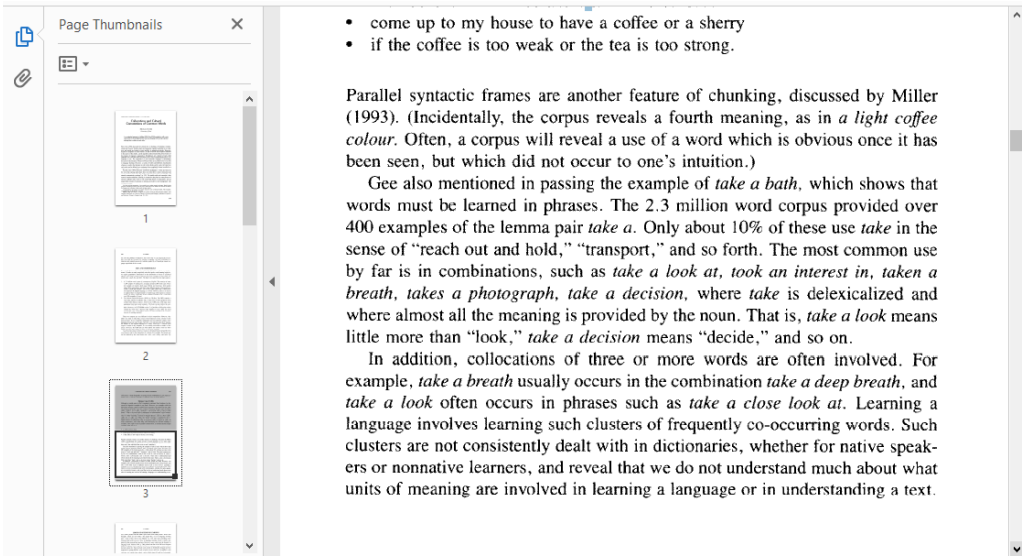


collocation, I mean frequently occurring lexical combinations; I give many examples later. I indicate meanings of linguistic units in double quotes.

Taking a Cup of Coffee

Although a central point in Gee's argument concerned fixed routines, he discussed no linguistic examples in any detail. However, two examples which he did briefly mention could be exploited in his argument. He pointed out that *coffee* has different meanings which are evident in different phrases, such as *cup of coffee* ("drink"), *tin of coffee* ("granules"), and *picking coffee* ("crop of coffee beans"). That is, the word *coffee* is ambiguous (or indeterminate?) out of context, but almost always disambiguated by its collocates in running text. For example, only its "drink" meaning is relevant in frequent phrases, such as *a black coffee, coffee bar,* or *coffee cup.* Often, the "drink" meaning is signalled by the (+ liquid) feature of collocating verbs, as in *poured coffee* or *sipped my coffee*. In some occurrences, even more help with identifying the relevant meaning is available, since *coffee* occurs in parallel constructions, as in these attested corpus examples with *and* or *or:*

- a coffee and tea shop
- she had ordered coffee and I had ordered beer
- come up to my house to have a coffee or a sherry
- if the coffee is too weak or the tea is too strong.



- come up to my house to have a coffee or a sherry
- if the coffee is too weak or the tea is too strong.

Parallel syntactic frames are another feature of chunking, discussed by Miller (1993). (Incidentally, the corpus reveals a fourth meaning, as in *a light coffee colour.* Often, a corpus will reveal a use of a word which is obvious once it has been seen, but which did not occur to one's intuition.)

Gee also mentioned in passing the example of *take a bath,* which shows that words must be learned in phrases. The 2.3 million word corpus provided over 400 examples of the lemma pair *take a.* Only about 10% of these use *take* in the sense of "reach out and hold," "transport," and so forth. The most common use by far is in combinations, such as *take a look at, took an interest in, taken a breath, takes a photograph, take a decision,* where *take* is delexicalized and where almost all the meaning is provided by the noun. That is, *take a look* means little more than "look," *take a decision* means "decide," and so on.

In addition, collocations of three or more words are often involved. For example, *take a breath* usually occurs in the combination *take a deep breath,* and *take a look* often occurs in phrases such as *take a close look at.* Learning a language involves learning such clusters of frequently co-occurring words. Such clusters are not consistently dealt with in dictionaries, whether for native speakers or nonnative learners, and reveal that we do not understand much about what units of meaning are involved in learning a language or in understanding a text.