

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВАСИЛЯ СТУСА
ФАКУЛЬТЕТ ІНОЗЕМНИХ МОВ
КАФЕДРА ІНОЗЕМНИХ МОВ ПРОФЕСІЙНОГО СПРЯМУВАННЯ

В. І. Калініченко, Н. В. Стрюк

**МЕТОДИЧНІ ВКАЗІВКИ
ДЛЯ САМОСТІЙНОЇ РОБОТИ З ДИСЦИПЛІНИ
«ПРАКТИЧНИЙ КУРС ІНОЗЕМНОЇ МОВИ (АНГЛІЙСЬКОЇ)»**

*для студентів 3 курсу спеціальності «Прикладна лінгвістика»
філологічного факультету*

Методичний посібник

Вінниця
ДонНУ імені Василя Стуса
2018

УДК 811.111'42(076.5)

М 545

Укладачі:

В. І. Калініченко, кандидат філологічних наук, доцент кафедри іноземних мов професійного спрямування факультету іноземних мов Донецького національного університету імені Василя Стуса.

Н. В. Стрюк, асистент кафедри іноземних мов професійного спрямування факультету іноземних мов Донецького національного університету імені Василя Стуса.

Рецензенти:

Я. В. Григошкіна, кандидат філологічних наук, доцент кафедри теорії та практики перекладу факультету іноземних мов Донецького національного університету імені Василя Стуса.

М. М. Юрковська, кандидат філологічних наук, доцент кафедри англійської філології факультету іноземних мов Донецького національного університету імені Василя Стуса.

*Рекомендовано до друку Вченою радою факультету іноземних мов
Донецького національного університету імені Василя Стуса
(протокол № 9 від 25.05.2018 р.)*

М 545 **Калініченко В. І., Стрюк Н. В.** Методичні вказівки для самостійної роботи з дисципліни «Практичний курс іноземної мови (англійської)» для студентів 3 курсу спеціальності «Прикладна лінгвістика» філологічного факультету. Вінниця : ДонНУ імені Василя Стуса, 2018. 84 с.

Методичний посібник призначений для самостійної роботи студентів 3 курсу філологічного факультету спеціальності «Прикладна лінгвістика» під час вивчення дисципліни «Практичний курс іноземної мови (англійської)». Високий рівень володіння англійською мовою є визначальною складовою ефективною навчальною та науковою діяльністю прикладних лінгвістів. Рекомендовані завдання розроблені на матеріалі статей довідника «The Handbook of Applied Linguistics» (2004).

УДК 811.111'42(076.5)

© Калініченко В. І., 2018

© Стрюк Н. В., 2018

© ДонНУ імені Василя Стуса, 2018

ЗМІСТ

ВСТУП.....	4
UNIT 1. WHAT IS LEXICOGRAPHY?.....	5
UNIT 2. WHAT IS A DICTIONARY? DICTIONARIES AND ENCYCLOPEDIAS. TYPES OF DICTIONARY.....	14
UNIT 3. DICTIONARIES IN APPLIED LINGUISTICS.....	24
UNIT 4. LANGUAGE CORPORA. EMPIRICAL LINGUISTICS.....	33
UNIT 5. MODERN CORPORA AND SOFTWARE.....	41
UNIT 6. WORDS AND PHRASES. RECURRENT PHRASES, COLLOCATIONS AND PHRASAL SCHEMAS.....	50
UNIT 7. DISCOURSE ANALYSIS.....	61
UNIT 8. RULES AND PRINCIPLES, CONTEXTS AND CULTURES OF LANGUAGE IN USE. INTERACTION.....	70
СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ.....	80

ВСТУП

Самостійна робота студентів є необхідною умовою підготовки висококваліфікованих фахівців, ключовим елементом у процесі самовдосконалення та невід'ємною частиною безперервного навчання (Lifelong learning). Мета самостійної роботи полягає в формуванні вміння систематизувати, планувати, контролювати й регулювати свою діяльність без допомоги й контролю викладача.

Методичний посібник призначений для самостійної роботи студентів 3 курсу філологічного факультету спеціальності «Прикладна лінгвістика» під час вивчення дисципліни «Практичний курс іноземної мови (англійської)». Високий рівень володіння англійською мовою є визначальною складовою ефективною навчальною та науковою діяльністю прикладних лінгвістів. Рекомендовані завдання розроблені на матеріалі статей довідника «The Handbook of Applied Linguistics» (2004).

Посібник складається з восьми розділів і охоплює такі актуальні теми прикладної лінгвістики, як: «What is Lexicography?», «What is a Dictionary? Types of Dictionary», «Dictionaries in Applied Linguistics», «Language Corpora», «Modern Corpora and Software», «Words and Phrases», «Discourse Analysis», «Rules and Principles, Contexts and Cultures of Language in Use».

Кожен розділ містить словник базової лексики та завдання для контролю та самоконтролю різних типів, що є, головню, пов'язаними з питаннями лексикології, граматики та перекладу: проблемні запитання, вправи на словотворення, заповнення пропусків, вправи на підстановку, корегування типових помилок у поданих текстових фрагментах, завдання на підбір еквівалентних лексичних одиниць в англійській та українській мовах, завдання на виконання перекладу речень з використанням активної лексики конкретного розділу та ін.

Посібник також містить список джерел для подальшого поглибленого читання з актуальних проблем прикладної лінгвістики.

UNIT 1

WHAT IS LEXICOGRAPHY?

It is difficult to arrive at a succinct and satisfying working definition of lexicography. Even a cursory glance in dictionaries and other reference works and in the secondary literature reveals many variations on a theme, reflecting a variety of standpoints. In a narrow sense lexicography may be described as the art and craft of writing a dictionary. Certainly, a lexicographer is essentially someone who writes or contributes to a dictionary or dictionaries, be it as an individual or a member of a team, as a freelancer or an in-house employee, as a full-time professional or part-time alongside other activities such as university lecturing. *Lexicographer* is also used more generally to refer to writers of other reference works, including encyclopedias. Like other definitions, however, and indeed like much dictionary writing itself, this definition of lexicography is derivative (Landau, 2001), and it is a compromise for the sake of brevity. It raises many questions: why dictionary, why not e.g., thesaurus, lexicon, or encyclopedia and other reference works? Why write, why not, for example, plan, edit, publish or make, produce, compile, let alone study, review, or use? Why art and craft, why not, for example, activity, process, technique, science, job, profession or practice, let alone history, study, use, or theory?

There are justifiable answers to such questions. The dictionary is widely regarded as the prototypical work of lexical reference, but this claim requires much further explication. Writing is the essential lexicographic activity, especially writing and rewriting semantic, pragmatic, or etymological descriptions; planning and data collection precede and accompany the writing, editing and publishing follow it. Good lexicography is more than compilation. Extracting meanings and uses from authentic texts and explaining them clearly and fully in a minimum of words is an art, as is the selection of appropriate illustrative examples. Writing with dictionary users uppermost in mind in an attempt to meet their needs is a practical and useful activity, a craft. Defining lexicography in this narrow sense as the art and craft of writing a dictionary is meant to locate it explicitly at the center of the applied linguistic endeavor and to emphasize the high degree of human knowledge, insight, judgment and skill required to produce the text of a successful reference work designed to be of practical use and benefit in real-life situations. Certainly, a dictionary that does not prove useful is unlikely to prove successful.

The advent of electronic corpora and media can make the lexicographers' work better, but not necessarily easier. Computers can store and process quantities of textual data quite unmanageable by humans. Where several million manually and painstakingly excerpted citation slips were once considered a sufficient basis for a multi-volume scholarly dictionary, now even one-volume trade dictionaries rest on hundreds of millions of rapidly and automatically entered running words. The differences are not only in quantity, but more importantly in quality. Lexicographers now have at their disposal vastly superior language data. Neutral frequency counts of masses of words can act as a counterbalance to intuition, memory and possible bias in many of the decisions they must make in accordance with the specifications of the particular

dictionary project. They help determine which usages are central and which are peripheral, which new items should be included and which items should be excluded as obsolescent or archaic, which combining forms and multi-word items warrant status as main lemmas or headwords rather than as run-ons and sub-lemmas, or how homographs and senses can be ordered, to mention but a few possibilities. Lexicographers have been at the forefront in utilizing language corpora and applying the findings of corpus linguistics to good effect in their analysis and description of lexis and hence to the benefit of their users. The corpus revolution is very real; computerphoria would be misplaced, however. There may be huge savings in storage space and processing time, but it is humans who continue to choose the texts and analyze the vastly increased data, which can now in fact require more time, experience, and skill to process than before. Humans discern and describe sense distinctions in polysemous words and between sets of synonyms, antonyms, and hyponyms. They select appropriate illustrative examples or establish usage and usage restrictions in tune with changing sociocultural conventions. And specialist material from a directed reading program still has a place alongside the mass data entered by means of optical scanners, magnetic tapes, and the like.

Similarly, electronic media open up quite new possibilities for the presentation and use of lexicographical material. They can, for instance, help overcome the constraints of space that have long plagued lexicographers and their editors and limited the coverage, description, and illustration of lexical items even in comprehensive or unabridged dictionaries. The size of the computer screen and of the «search word» box remain limitations, however, and favor directed searches for specific items over the incidental consulting of neighboring entries and the general, even random browsing so dear to word and dictionary buffs brought up on printed books. They can help overcome the tyranny of the printed alphabet that has severely limited accessibility and fostered the modern dominance of the alphabetic mode of presentation over the older thematic or systematic mode. Access through the alphabet has become a practical necessity for most users, however, and modern thesauruses are either arranged alphabetically or have an alphabetical index. Online e-dictionaries and e-cyclopedias available free or by subscription on the Internet and CD-ROM are already vying with and in some cases supplanting conventional printed books. Large and expensive multi-volume reference works seem to be leading the paradigm shift from book to bank and byte. Academic researchers working on and with scholarly historical dictionaries are among the major beneficiaries. At the click of a mouse they can conveniently search from their desks the full resources of the *Oxford English Dictionary Online* in ways simply not possible on visits to the library to consult the 20 large and alphabetically ordered volumes of the *Second Edition*. Now that wordbanks and wordnets, such as the *British National Corpus* or the *Bank of English*, the *Princeton WordNet*, and the multilingual *EuroWordNet*, can be accessed in full or in part on the Internet, users can effectively become their own lexicographers. The future of lexicography is undoubtedly electronic. Nonetheless, however much the computer can aid lexicographers as dictionary writers, it will not replace them.

The questions raised above also point to a need to understand lexicography in a wider sense as used in the rapidly increasing number of university courses, conferences and workshops, books, journals and articles on the subject. These concern not only lexicography as practice, namely the planning, writing, editing, and publishing of dictionaries and other lexicographical reference works, but also lexicography as theory, notably the study of dictionary history, criticism, typology, structure, and use (Wiegand, 1998). Some scholars distinguish theory, also known as metalexicography or dictionary research, from practice as lexicography proper. Others include all aspects of both theory and practice in their definition of lexicography. Be that as it may, many different sub-branches of lexicography can be distinguished, ranging from computational to pedagogical and terminographical. Postgraduate degree or diploma courses on lexicography aim to provide academic qualifications and professional training for future dictionary writers. However, most lexicographers still train as before in-house or on the job. The literature on lexicography involves university and other scholars as well as lexicographers and ex-lexicographers. The former mediate the findings of research in (theoretical) linguistics and other academic disciplines, which most practicing lexicographers cannot possibly keep abreast of. They also describe and re-edit or reprint historical dictionaries, and make suggestions for the improvement of all aspects of lexicographical description. In line with the lexicographers' constant emphasis on utility, the literature on lexicography now devotes much attention to dictionary uses in academic research, educational practice, and leisure activity. It focuses particularly on dictionary users and seeks to ascertain who uses which dictionary when and where, for what purpose and with what result. This focus on the user perspective (Hartmann, 2001, pp. 80–95, pp. 115–20) and the need for empirical studies of what dictionary users do in real look-up situations (Atkins, 1998; Nesi, 2000; Tono, 2001) are important concerns of applied linguistics. Among the scientific commissions of the International Association of Applied Linguistics (AILA) is one devoted to Lexicology and Lexicography as research areas which can contribute to a better understanding and facilitation of language learning and language use and are studied from several perspectives. However, important as it is, the user's perspective is not the only one: lexicographers as dictionary writers, scholars as dictionary researchers, (language) teachers as mediators also offer essential perspectives on the complex and multi-faceted activity that is lexicography, quite apart from publishers, consultants, and others. At the center of this activity is the dictionary itself as text (Hartmann, 2001, pp. 24–5), and the dictionary is thus the focus of the discussion that follows.

Lexicography is in essence an art and a craft. It is also a profession and a hobby, a scholarly and commercial enterprise, and an academic discipline. It is, further, a longstanding cultural practice and an integral part of the intellectual tradition in literate societies.

TASKS

I. Study the topical vocabulary below. Find the contexts where the units under consideration are used in the text.

1	succinct (adj.)	[sək'sɪŋ(k)t]	стислий, короткий
2	standpoint (n.)	[ˈstɑnd(p)ˌpɔɪnt]	точка зору
3	thesaurus (n.)	[θɪ'sɔːrəs]	вичерпний словник; тезаурус
4	lexicon (n.)	[ˈleksɪk(ə)n]	лексикон, словник; лексика
5	compile (v.)	[kəm'paɪl]	укладати (довідники); компілювати
6	language corpora (n.)	[ˈkɔːp(ə)rə]	множ. від language corpus – мовний корпус
7	textual data (n.)	[ˈtɛkstʃʊəl ˈdeɪtə]	текстові дані
8	multi-volume dictionary (n.)	[wʌn-'vɒljʊm ˈdɪkʃ(ə)n(ə)rɪ]	багатотомний словник
9	one-volume dictionary (n.)	[mʌlti-'vɒljʊm ˈdɪkʃ(ə)n(ə)rɪ]	однотомний словник
10	bias (n.)	[ˈbaɪəs]	упередженість; необ'єктивність; (for) схильність, ухил
11	obsolescent (adj.)	[əbsə'les(ə)nt]	який застаріває, виходить із вжитку, відживає
12	lemma (n.)	[ˈlemə]	лема; провідне слово, гніздове слово (в словнику або глосарії з алфавітно-гніздовою системою)
13	run-on (n.)	[rʌn-ɒn]	текст, набраний у підбір
14	discern (v.)	[dɪ'sɜːn]	відрізняти, відділяти, виділяти; розрізняти
15	sense distinctions (n.)	[sens dɪs'tɪŋkʃənz]	сміслові відмінності
16	usage restrictions (n.)	[usage ˈrestɹɪkʃnz]	обмеження у використанні
17	constraint (n.)	[kən'streɪnt]	обмеження
18	coverage (n.)	[ˈkʌvərɪdʒ]	охоплення
19	comprehensive (adj.)	[kəmprɪ'hensɪv]	повний, докладний, детальний
20	unabridged (adj.)	[ʌnə'brɪdʒd]	повний, нескорочений
21	foster (v.)	[ˈfɒstə]	сприяти розвитку; заохочувати
22	supplant (v.)	[sə'plɑnt]	витіснити (що-небудь)
23	conventional (adj.)	[kən'venʃənəl]	звичний, традиційний, загальноприйнятий;
24	scholarly (adj.)	[ˈskɒləli]	учений; властивий ученим, науковий
25	computational (adj.)	[ˌkɒmpju(:)'teɪʃən(ə)l]	обчислювальний, комп'ютерний

26	keep abreast of (v.)	[ki:p ə'brɛst ɒv]	не відставати від, іти в ногу з
27	utility (n.)	[ju(:)'tɪlɪti]	корисність, практичність, вигідність
28	facilitation (n.)	[fə'sɪlɪ'teɪʃən]	полегшення, допомога
29	multifaceted (adj.)	[mʌltɪ'fæsɪtɪd]	багатогранний, багатосторонній
30	endeavor (n.)	[en'devər]	спроба, намагання

II. Answer the following questions to the text.

- 1) How may lexicography be described in a narrow sense?
- 2) Who is a lexicographer? In what meaning is this term used more generally?
- 3) What are the essential lexicographic activities?
- 4) How can the advent of electronic corpora and media make the lexicographers' work better? How many words do one-volume trade dictionaries rest on now?
- 5) What do humans continue to do in lexicography despite the corpus revolution and computerphoria?
- 6) What new possibilities for the presentation and use of lexicographical material do electronic media open up?
- 7) Why are academic researchers working on and with scholarly historical dictionaries among the major beneficiaries of shifting large and expensive multi-volume reference works from book to bank and byte?
- 8) What is the aim of postgraduate degree or diploma courses on lexicography?
- 9) What does the literature on lexicography devote much attention to now? What does it focus on particularly?
- 10) What are important concerns of applied linguistics in lexicography?

III. Match the words to their definitions.

1	language corpus	a	having a variety of different and important features or elements
2	compile	b	a word of more specific meaning than a general or superordinate term applicable to it
3	constraint	c	the state of being useful
4	bias	d	a concentration on or interest in one particular area or subject
5	utility	e	encourage the development of something
6	foster	f	the state of being checked, restricted, or compelled to avoid or perform some action
7	lemma	g	a collection of recorded utterances used as a basis for the descriptive analysis of a language
8	obsolescent	h	it is no longer needed because something better has been invented

9	hyponym	i	a word or phrase defined in a dictionary or entered in a word list
10	multi-faceted	j	produce by assembling information collected from other sources

IV. Match the parts below to complete a single syntactic unit from the text.

1	Even a cursory glance in dictionaries and other reference works	a	applying the findings of corpus linguistics to good effect in their analysis and description of lexis.
2	The dictionary is widely regarded as the prototypical work of lexical reference,	b	and between sets of synonyms, antonyms, and hyponyms.
3	Computers can store and process quantities of	c	reveals many variations on a theme, reflecting a variety of standpoints.
4	The differences are not only in quantity,	d	or dictionary research, from practice as lexicography proper.
5	Lexicographers have been at the forefront in utilizing language corpora and	e	but this claim requires much further explication.
6	They can help overcome the tyranny of the printed alphabet that has severely limited accessibility and	f	as research areas which can contribute to a better understanding and facilitation of language learning and language use.
7	Humans discern and describe sense distinctions in polysemous words	g	however, and modern thesauruses are either arranged alphabetically or have an alphabetical index.
8	Access through the alphabet has become a practical necessity for most users,	h	but more importantly in quality.
9	Some scholars distinguish theory, also known as metalexicography	i	textual data quite unmanageable by humans.
10	Among the scientific commissions of the International Association of Applied Linguistics is one devoted to Lexicology and Lexicography	j	fostered the modern dominance of the alphabetic mode of presentation over the older thematic or systematic mode.

V. Fill in the gaps using the appropriate words from the box. Underline the words inserted.

reference work	run-ons	unabridged	sub-branches	electronic corpora
keep abreast of	multi-faceted	obsolescent	coverage	lecturing
supplanting	endeavor	usage restrictions	contributes	

1) A lexicographer is essentially someone who writes or to a dictionary or dictionaries, be it as an individual or a member of a team, as a freelancer or an in-house employee, as a full-time professional or part-time alongside other activities such as university

2) Lexicography in this narrow sense as the art and craft of writing a dictionary is meant to locate it explicitly at the center of the applied linguistic and to emphasize the high degree of human knowledge, insight, judgment and skill required to produce the text of a successful designed to be of practical use and benefit in real-life situations.

3) The advent of and media can make the lexicographers' work better, but not necessarily easier.

4) They help determine which new items should be included and which items should be excluded as or archaic, which combining forms and multi-word items warrant status as main lemmas or headwords rather than as and sub-lemmas.

5) Humans select appropriate illustrative examples or establish usage and in tune with changing sociocultural conventions.

6) Electronic media can help overcome the constraints of space that have long plagued lexicographers and their editors and limited the, description, and illustration of lexical items even in comprehensive or dictionaries.

7) Online e-dictionaries and e-cyclopedias are already vying with and in some cases conventional printed books.

8) Many different of lexicography can be distinguished, ranging from computational to pedagogical and terminographical.

9) Ex-lexicographers mediate the findings of research in (theoretical) linguistics and other academic disciplines, which most practicing lexicographers cannot possibly

10) Lexicographers, scholars and language teachers offer essential perspectives on the complex and activity that is lexicography, quite apart from publishers, consultants, and others.

VI. Find the English equivalents from the text to the following words and word-combinations.

1	застарілий	
2	довідник	
3	заново відредагувати	
4	гніздове слово	
5	багатослівний	
6	викладання в університеті	
7	полегшення вивчення мови	
8	витіснення звичайних друкованих книг	
9	повні або нескорочені словники	

10	можлива упередженість	
11	стисле і задовільне робоче визначення	
12	різноманітні точки зору	
13	обсяги обробки текстових даних	
14	розпізнавання та опис смислових відмінностей	
15	обмежити охоплення	

VII. Find appropriate synonyms (if any) and antonyms (if any) to the words below.

№	Word	Synonym	Antonym
1	succinct (adj.)		
2	essential (adj.)		
3	compilation (n.)		
4	attempt (n.)		
5	misplaced (adj.)		
6	limitation (n.)		
7	aim (v.)		
8	involve (v.)		
9	utility (n.)		
10	focus (v.)		

VIII. Provide all the possible derivatives to the word forms below. Make up sentences with the words derived.

compile, explicitly, restriction, coverage, facilitation, contribute, produce, derivative, convention, edit
--

IX. Correct the possible mistakes in terms of grammar, spelling and inappropriate use of vocabulary in the sentences below.

1) *Lexicografer* is also using more generally to refere to writers of other reference works, including enciclopedias.

2) Writing is a essential lexicographic activity, especially writing and rewriting semantic, pragmatic, or etymological descriptions; planning and data collection precede and accompany the writing, editing and publishing follow it.

3) Certainly, a dictionary that do not proved useful is unlikely to prove succssful.

4) Lexicografers now has at their disposal vastly super languages data.

5) The corpus revolution are very realistic; computerphoria would be misspelled, however.

6) Specialist material from a directed reading program steell have a place alongside the mass data entered by meaning of optical scanners, magnetic tapes, and the like.

7) Electronic media open up quiet new possibilities for the presentation and use of lexicographical material.

8) It focus particulaly on dictionary users and seak to ascertain who use whose dictionary when and where, for what purpose and with that result.

9) At the center of this activity is the dictionary itself as text, and the dictionary is thus the focus of the discussion that follows.

10) Lexicograpy is an longstanding cultural practice and a integral part of the intelectual tradition in literature societies.

X. Translate the following sentences into English, using the topical vocabulary under consideration.

1) Лексикографія – це більше, ніж просто укладання; визначення значень та сфери використання на основі автентичних текстів і чітке та повне пояснення їх за допомогою мінімальної кількості слів є мистецтвом.

2) Лексикографи найактивніше використовують мовні корпуси і застосовують здобутки корпусної лінгвістики для високо ефективного аналізу та опису лексики.

3) Вони можуть допомогти подолати тиранію друкованого алфавіту, що суттєво обмежує доступність, і сприяє сучасному домінуванню алфавітної моделі презентації над старішими тематичною або систематичною моделями.

4) Увага на перспективі користувача та потреба в емпіричних дослідженнях того, що роблять користувачі словників в реальних ситуаціях пошуку, є важливими проблемами прикладної лінгвістики.

5) Існує величезна економія місця зберігання та часу обробки, але саме люди продовжують вибирати тексти та аналізувати інформацію, яка значно збільшується, що наразі може вимагати більше часу, досвіду та майстерності, ніж раніше.

UNIT 2

WHAT IS A DICTIONARY? DICTIONARIES AND ENCYCLOPEDIAS. TYPES OF DICTIONARY

What is a dictionary? Dictionaries and encyclopedias.

As already mentioned, the dictionary is widely regarded as the prototypical work of lexical reference. It classifies and stores information in print or, increasingly, electronic form and has an access system or systems designed to allow users to retrieve the information in full or in part as readily as possible. The information is essentially linguistic and may include material on the form, meaning, use, origin, and history of words, phrases, and other lexical items. In a dictionary phonetic and grammatical information is word-related and thus essentially lexical. Put very simply, a dictionary is a book or bank about words.

In theory linguistic or lexical information may be distinguished from extralinguistic or encyclopedic information. Certainly, there are classes of words which lend themselves to either linguistic or encyclopedic treatment. The former include function words such as prepositions, determiners, or conjunctions and discourse-marking chunks such as *you know*, *I mean*, and many others. They derive their meaning from their function within a linguistic text rather than from any reference to extralinguistic reality and are properly treated in a dictionary.

The latter may include proper names of people and places, biographical data, and descriptions of historical events, political, social, and cultural institutions, geographical and geopolitical entities, works of art, literature and music, myths and mythological figures, beliefs and religions, academic disciplines, and the like. A reference work that stores and classifies such factual information on all or some branches of knowledge or a single subject area is generally known as an encyclopedia. Put simply, an encyclopedia is a book or bank about facts. It is notable in this connection that multilingual and especially bilingual dictionaries have long been and continue to be very common, but this is not true of encyclopedias.

Conversely, the latter can be and have been translated, but this does not seem to be the case with dictionaries, except perhaps for the fast-developing genre of bilingualized, semi-bilingual, or bridge dictionaries in the area of pedagogical lexicography. In practice, however, a hard and fast distinction between lexical and encyclopedic information is not possible. Humans use language to communicate about facts, things, and people; words and the world are inextricably linked. A linguistic description of nouns as names for plants, animals, or insects and of adjectives as names for colors, for instance, necessarily involves encyclopedic information.

Such items are entered in both dictionaries and encyclopedias. Their semantic explanation will differ in degree rather than kind, namely in the amount of factual information required or provided to identify and characterize the object referred to according to the intended purpose of the particular reference work. Lexicographers must be concerned with words in their own right as linguistic items and with what words refer to in the world of extralinguistic reality or with their referents as such. Dictionaries and encyclopedias are best seen as two types of reference work, among

others, which stand at opposite ends of a continuum, one concerned with words as linguistic or lexical items, the other with facts as such. There are many mixed or blended forms in between (McArthur, 1986, pp. 102–4).

In the titles and/or subtitles of subject-area and biographical reference works, which are most commonly published in one volume, *dictionary* can be used alternatively and synonymously with *encyclopedia*. In this same sense *companion* and *handbook* are also found. In the titles of dictionaries-cumencyclopedias, which combine lexical and encyclopedic information, the attribute *encyclopedic* sometimes explicitly qualifies the head noun *dictionary*, sometimes not. A successful example of a fully integrated encyclopedic dictionary is the *Reader's Digest Great Illustrated Dictionary*, 1984, which features small color photographs and drawings at the appropriate alphabetical place in the outside columns of virtually every page, color maps, and part- and fullpage panels and tables, most also in color. Clearly, the genre of encyclopedic dictionary is established as a blend between the dictionary as a word book/bank on the one hand and the encyclopedia as a fact book/bank on the other.

This is certainly true of the American and French traditions, less so in the British and German ones. Equally clearly, the genre is regarded in English as a type of dictionary, and thus belongs to the province of lexicography. The question whether encyclopedias as such also belong has been variously answered. My own view is that it is justifiable to regard encyclopedias as falling within the scope of lexicography in the wider sense discussed above, and it would definitely enhance and advance metalexicography if encyclopedias were given fuller attention. If the present chapter nonetheless restricts itself largely to dictionaries as word books, it is for practical reasons of space, especially as there are so many different types of dictionary.

Types of dictionary

Given that dictionaries belong to the oldest, most widespread, and best-selling books in literate societies, it is hardly surprising that their number is legion. Different societies have different lexicographical traditions, and ideas on what might constitute the prototypical dictionary vary accordingly. The range of languages, varieties, and vocabularies, of sizes, formats, and prices, or intended purposes, uses, and users seems inexhaustible. Most dictionaries codify natural languages, but there are also dictionaries of international auxiliary languages, sign languages, shorthands, and braille. The time interval between new impressions and even new editions of popular trade dictionaries grows ever shorter, and their covers and dust jackets resemble ever more strongly billboards advertising the virtues and unique features of their product in a highly competitive market. This is perhaps particularly true of English dictionaries, not least for second/foreign language learners as a reflection of the current worldwide dominance of English as an additional language.

What impact electronic publishing will have on this situation is not yet clear. Currently, prospective dictionary users and buyers are faced with a bewildering *embarras de richesses*. Language teachers and librarians are faced with the problem of continuously updating their resources. Dictionary scholars are faced with a rich, diverse, and ever-changing field of study. It is small wonder that dictionary typology

has become an integral component of metalexicography, that different criteria, including size, scope of linguistic and subject-area coverage, number of languages, period covered, target groups and intended uses among others, have been advanced as the basis of different typologies, and that no agreed taxonomy has emerged to classify the variety of dictionary types. In the practical typology that underlies the organization of much of their international encyclopedia of lexicography, Hausmann et al. Distinguish first between monolingual and multilingual dictionaries. Of the latter, the vast majority are bilingual and cover two national standard languages.

Bilingual dictionaries continue to be the most-used reference book in second/foreign language learning at all levels. There are specialized bilingual dictionaries, such as dictionaries of deceptive cognates or false friends, subject-specific technical dictionaries, and pictorial dictionaries that feature line drawings largely of thematically grouped concrete objects with their designations in two languages. The prototypical bilingual dictionary, however, is the general translation dictionary. Headwords or lemmas in one (source) language, usually presumed to be the user's first language, are supplied at least with translation equivalents in the other (target) language. Full equivalents may need mere listing, while partial and surrogate equivalents require further explanation or exemplification to ensure sense identification and discrimination.

Passive or receptive dictionaries help in decoding or translating from the target/foreign to the source/native language, active or productive dictionaries help in encoding or translating from the source to the target language. For each language pair there are in theory four directions to consider, for example, German-French for French users and French-German for German users (passive), German-French for German users and French-German for French users (active). In practice most bilingual dictionaries are bidirectional: French-German and German-French. Monolingual dictionaries are divided into general and specialized works. The former are found in two major types, the encyclopedic dictionary and above all the semasiological defining dictionary. Aimed at adult native speakers and usually published in a single volume – although the volume may range from compact and portable to very bulky and unwieldy – this latter is the prototypical dictionary of dictionaries in most European lexicographies. Alphabetically ordered lemmas, representing in the main unmarked contemporary standard vocabulary, are supplied with semantic explanations or descriptions of various kinds. Often there is much other information as well.

The more than 70 types of specialized dictionaries derive mainly from different types of marked lemmas in the macro-structure or from different types of lexicographic information other than the definitions in the micro-structure. Marked lemmas include archaisms, neologisms, regionalisms, and internationalisms. There are dictionaries devoted to all these and many other lemma types. Syntagmatic information underlies dictionaries of syntactic patterns or valency, collocations, fixed phrases and idioms, proverbs and quotations. Paradigmatic information underlies onomasiological dictionaries, which move from concepts or word meanings to word forms as the expression of these concepts.

They include dictionaries which classify and list synonyms with or without sense discrimination and meaning description – the former are discriminating, the

latter cumulative synonymies – reverse and word-family dictionaries, and the thesaurus. From other categories of lexicographic information derive dictionaries of spelling, pronunciation, inflections, frequency and etymology, and chronological dictionaries.

There are dictionaries dealing *inter alia* with specific text types, texts by individual authors, and concordances. This essentially phenomenological typology is complemented by a functional one based on the intended use and target group. Included here are children's and learners' dictionaries, both for native and non-native speakers, as well as dictionaries of core vocabulary, all of which are pedagogic in orientation.

This typology is neither exhaustive nor uncontested. It does not seek explicitly to account for all of the many mixed or hybrid types of lexicographic reference works. Nor can it reflect the fact that different traditions can favor different dictionary types. It also needs to be said that the typology classifies printed dictionaries and that it remains to be seen what impact the electronic presentation of lexicographic information with its different possibilities will have on dictionary typology. The many types of reference works classified in this typology are all dictionaries or word books. The overwhelming majority contain the term *dictionary* (*dictionnaire*, *Wörterbuch*) in the title, and it is this term that is firmly entrenched as the overall designation of works of lexical or word-centered reference. Few others have survived.

Glossary is used of an alphabetical list of selected items with definitions and/or translation equivalents as found commonly at the back of subject-area textbooks or language course books. *Vocabulary* can be used similarly, but most commonly refers to the lexical items of a given language, also of a language variety, speaker, or text, taken collectively and studied in lexicology but not necessarily codified and described in lexicography. Part synonyms are *lexis* and *lexicon*, both of which are also used as antonyms of *grammar*. *Lexicon* is used further, often in the collocation or compound *mental lexicon*, for words and vocabulary stored and processed in the speaker's mind. As a label for a lexicographic reference work it is now generally applied in English to specialized or technical works or to dictionaries of classical languages such as Greek or Arabic. It is thus more restricted than its one-time synonym *dictionary*. McArthur's *Longman Lexicon of Contemporary English*, 1981, however, is a type of thesaurus. In modern lexicographic use (Hellen, 1999), *thesaurus* refers to a word book that classifies and groups lexical items of a language, variety, or subject area according to sense relations, especially synonymy, in semantic sets and arranges and presents them alphabetically and/or thematically or conceptually. All thematic and some alphabetical thesauruses now have alphabetical indexes to ensure easy access, especially when the items are grouped according to a philosophical world view such as those which determined the organization of older thematic encyclopedias. At one level *thesaurus* is used as a hyponym and at another level as an antonym of *dictionary*: the thesaurus is both a type of dictionary and it also contrasts with the dictionary proper, as reflected in the titles of combined dictionaries-cum-thesauruses such as *Collins (Concise, Compact) Dictionary and Thesaurus*. The dictionary proper here is the alphabetical semasiological defining dictionary, and this type represents the stock answer to the question, what is a dictionary?

TASKS

I. Study the topical vocabulary below. Find the contexts where the units under consideration are used in the text.

1	exemplification (n.)	[ɪg,zɛmplɪfɪ'keɪʃən]	ілюстрація, пояснення прикладом
2	emerge (v.)	[ɪ'mɜ:dʒ]	з'являтися, виникати; з'ясовуватися
3	widespread (adj.)	['waɪdspred]	розповсюджений, поширений
4	bewildering (adj.)	[bɪ'wɪldərɪŋ]	той, що бентежить, дивує; що заганяє в глухий кут
5	inextricably (adv.)	[ɪn'ɛkstri:kəbli]	нерозривно
6	notable (adj.)	['nəʊtəbl]	помітний, відчутний, очевидний; вартий уваги, значущий
7	nonetheless (adv.)	[,nʌnðə'les]	тим не менш, все таки, однак, проте
8	justifiable (adj.)	['dʒʌstɪfaɪəbl]	виправдано; що може бути виправданим
9	bulky (adj.)	['bʌlki]	громіздкий, великий
10	pictorial (adj.)	[pɪk'tɔ:riəl]	ілюстрований (про періодичне видання); яскравий, рисунковий, образотворчий
11	unwieldy (adj.)	[ʌn'wi:ldi]	Громіздкий
12	proper (adj.)	['prɒpə]	відповідний, правильний; властивий
13	process (v.)	['prəʊses]	обробляти (дані, інформацію тощо)
14	ensure (v.)	[ɪn'ʃʊə]	забезпечувати, гарантувати
15	restrict (v.)	[rɪs'trɪkt]	обмежувати
16	contrast (v.)	['kɒntrɑ:st]	порівнювати, протиставляти
17	arrange (v.)	[ə'reɪndʒ]	організовувати, налагоджувати, пристосовувати, налаштовувати
18	refer (v.)	[rɪ'fɜ:]	відносити до чогось, посилатися, звертатися
19	access (n.)	['ækses]	доступ
20	enhance (v.)	[ɪn'hɑ:ns]	покрещувати; збільшувати; підвищувати
21	determine (v.)	[dɪ'tɜ:mɪn]	визначати, зумовлювати; (<i>іноді контекст.</i> обчислювати)
22	store (v.)	[stɔ:]	зберігати; запасати; постачати
23	item (n.)	['aɪtəm]	одиниця; пункт; питання; деталь; повідомлення

24	overwhelming (adj.)	[,əʊvə'welmɪŋ]	переважний; нездоланний; незліченний
25	portable (adj.)	['pɔ:təbl]	портативний, переносний
26	designation (n.)	[,deziɡ'neɪʃən]	вказівка, призначення; визначення; назва
27	uncontested (adj.)	[,ʌnkən'testɪd]	безперечний, очевидний
28	exhaustive (adj.)	[ɪg'zɔ:stɪv]	вичерпний; виснажливий
29	majority (n.)	[mə'dʒɔ:prɪti]	Більшість
30	surrogate (adj.)	['sʌrəɡɪt]	несправжній, недійсний; той, що може виступати в якості замітника

II. Answer the following questions to the text.

- 1) How can you comment upon the following statement from the text: «Dictionary is a book or bank about words.»?
- 2) Is a hard and fast distinction between lexical and encyclopedic information possible? Explain your point of view.
- 3) Provide the definitions to the following notions: *glossary*, *vocabulary*, *dictionary*, *thesaurus*, *lexicon*, *encyclopedia*. How can they be properly differentiated?
- 4) What is a chronological dictionary? What are the particular ways it can be used in Applied Linguistics?
- 5) How many types of specialized dictionaries derive from different types of marked lemmas in the macrostructure?
- 6) Why do all thematic thesauruses now have alphabetical indexes?
- 7) What is the meaning of *inter alia* word combination used in the text?
- 8) What is the function of passive/receptive dictionaries?
- 9) What type of dictionary can be regarded as prototypical of dictionaries in most European lexicographies?
- 10) What types of bilingual dictionaries do they single out in Applied Linguistics? What can be viewed as a prototypical bilingual dictionary?

III. Match the words to their definitions.

1	idiom	a	a book that contains lists of words that have similar meanings
2	archaism	b	an expression whose meaning is different from the meaning of the individual words
3	internationalism	c	the words used for talking about a particular subject;
4	quotation	d	a list of the difficult words used in a piece of writing or subject, with explanations of their meaning

5	dictionary	e	a reference resource which provides information about many different subjects or about one particular subject; may be published as a single book, a series of books, or as a digital product such as a website or an app.
6	glossary	f	words from a book, play, film etc that you mention when you are speaking or writing
7	encyclopedia	g	a reference resource which provides information about words and their meanings, uses, and pronunciations; may be published as a printed book, or as a digital product such as a website or app, and it may be monolingual, bilingual, or multilingual.
8	vocabulary	h	a new word or expression, or an existing word used with a new meaning
9	thesaurus	i	a loanword that occurs in several languages (that is, translingually) with the same or at least similar meaning and etymology.
10	neologism	g	an old word or phrase that is no longer used

IV. Match the parts below to complete a single syntactic unit from the text.

1	<i>Lexicon</i> is used further, often in the collocation or compound <i>mental lexicon</i> ,	a	reference book in second/foreign language learning at all levels.
2	Alphabetically ordered lemmas, representing in the main unmarked contemporary standard vocabulary,	b	further explanation or exemplification to ensure sense identification and discrimination.
3	Included here are children's and learners' dictionaries, both for native and non-native speakers,	c	for words and vocabulary stored and processed in the speaker's mind.
4	As a label for a lexicographic reference work it is now generally applied in English to specialized or technical works	d	different traditions can favor different dictionary types.
5	Bilingual dictionaries continue to be the most-used	e	and with what words refer to in the world of extralinguistic reality or with their referents as such.
6	Full equivalents may need mere listing, while partial and surrogate equivalents require	f	active or productive dictionaries help in encoding or translating from the source to the target language.
7	Passive or receptive dictionaries help in decoding or translating from the target/foreign to the source/native language,	g	are supplied with semantic explanations or descriptions of various kinds.

8	Lexicographers must be concerned with words in their own right as linguistic items	h	as well as dictionaries of core vocabulary, all of which are pedagogic in orientation.
9	Paradigmatic information underlies onomasiological dictionaries,	i	or to dictionaries of classical languages such as Greek or Arabic.
10	Nor can it reflect the fact that	j	which move from concepts or word meanings to word forms as the expression of these concepts.

V. Fill in the gaps using the appropriate words from the box. Underline the words inserted.

proverbs inter alia lexicographic standard surrogate inextricably constitute frequency philosophical equivalents typology alphabetical exemplification
--

- 1) Syntagmatic information underlies dictionaries of syntactic patterns or valency, collocations, fixed phrases and idioms,and quotations.
- 2) All thematic and somethesauruses now have alphabetical indexes to ensure easy access, especially when the items are grouped according to aworld view such as those which determined the organization of older thematic encyclopedias.
- 3) It also needs to be said that theclassifies printed dictionaries and that it remains to be seen what impact the electronic presentation of information with its different possibilities will have on dictionary typology.
- 4) From other categories of lexicographic information derive dictionaries of spelling, pronunciation, inflections,and etymology, and chronological dictionaries.
- 5) Alphabetically ordered lemmas, representing in the main unmarked contemporaryvocabulary, are supplied with semantic explanations or descriptions of various kinds.
- 6) Headwords or lemmas in one (source) language, usually presumed to be the user's first language, are supplied at least with translationin the other (target) language.
- 7) Different societies have different lexicographical traditions, and ideas on what mightthe prototypical dictionary vary accordingly.
- 8) There are dictionaries dealingwith specific text types, texts by individual authors, and concordances.
- 9) Full equivalents may need mere listing, while partial and equivalents require further explanation orto ensure sense identification and discrimination.
- 10) Humans use language to communicate about facts, things, and people; words and the world arelinked.

VI. Find the English equivalents from the text to the following words and word-combinations.

1	алфавітний покажчик	
2	забезпечувати вільний доступ до	
3	лексична одиниця	
4	носій мови	
5	двомовний/тлумачний словник	
6	бути зафіксованим у лексикографічному джерелі	
7	класична мова	
8	постійно оновлювати довідкові джерела	
9	переважна більшість	
10	сміслові розрізнення	
11	потребувати подальшого роз'яснення	
12	простіше кажучи	
13	макро- і мікроструктура	
14	власне мовна та позамовна (екстралінгвістична) інформація	
15	галузь, область лексикографії	

VII. Find appropriate synonyms (if any) and antonyms (if any) to the words below.

№	Word	Synonym	Antonym
1	reverse (adj.)		
2	native (adj.)		
3	exhaustive (adj.)		
4	explicitly (adv.)		
5	commonly (adv.)		
6	passive (adj.)		
7	simplification (n.)		
8	productive (adj.)		
9	bulky (adj.)		
10	justifiable (adj.)		

VIII. Provide all the possible derivatives to the word forms below. Make up sentences with the words derived.

differ, include, contrast, refer, codify, reflect, combine, favor, compete, continue

IX. Correct the possible mistakes in terms of grammar, spelling and inappropriate use of vocabulary in the sentences below.

1) The *dictionary* proper here is the alphabetical semasiological defining *dictionary*, and *these type* represents the stock answer *on the question*, what is a *dictionary*?

2) The many types of reference works classified in this typology are all dictionaries or word books.

3) Marked lemma include archaisms, neologisms, regionalisms, and internationalisms.

4) Alphabetical ordered lemmas, representing in the main unmarked contemporary standard vocabulary, are supplied by semantic explanations or descriptions with various kinds.

5) Lexicon uses further, often in the collocation or compound mental lexicon, for words and vocabulary stored and processing in the speaker's mind.

6) Language teachers and librarians are faced under the problem of continuous updating their resource.

7) Bilingual dictionaries continues to be the most-used reference book in second/foreign language learning in all levels.

8) This essentially phenomenological typology is complemented by a functional one based for the intended use and target group.

9) Most dictionary codifying natural language, but there are also dictionaries of international auxiliary language, sign language, shorthands, and brailless.

10) It is notable in this connection that multilingual and especially bilingual dictionaries had long been and continue to be very common, but this is not false of encyclopedias.

X. Translate the following sentences into English, using the topical vocabulary under consideration.

1) У цьому зв'язку очевидним є те, що багатомовні і особливо двомовні словники вже протягом доволі тривалого часу залишаються достатньо поширеними і популярними, що не можна зазначити про енциклопедії.

2) Це, безумовно, стосується більшою мірою американської та французької лінгвістичної традиції, меншою мірою – британської та німецької.

3) Словники та енциклопедії загалом розглядаються як два окремих типи довідкового лексикографічного джерела, і головною відмінністю між ними є два протилежні полюси у так званому інформаційному континуумі – слова як лексичні одиниці та одиниці мови загалом, й з іншого боку – факти як такі.

4) Ця інформація, по суті, має лінгвістичну природу та може містити дані про форму, значення, вживання, походження та історію слів, фраз та інших лексичних одиниць.

5) Слова набувають свого значення залежно від своєї функції в межах лінгвістичного тексту, важливим при цьому є урахування екстралінгвістичних чинників, що не завжди є фіксованими у словнику при поданні лексико-семантичних варіантів конкретного слова.

UNIT 3

DICTIONARIES IN APPLIED LINGUISTICS

Wherever languages are used and wherever languages are taught and learned, especially in educational settings, dictionaries play a central role. As already discussed, lexicography is thus not only a field of professional, commercial, and academic activity in its own right, but also very much an integral part of applied linguistics and its constituent subject areas. The most obvious area is first and second/foreign language teaching and learning at all ages and levels of education, an area that some virtually equate with applied linguistics and that is by common consent certainly one of the core activities of applied linguistics. A few examples of other areas of professional applied linguistics must suffice here.

One such area is translation. Professional translators need and use dictionaries of different types according to the nature of the translation, general or specialized, literary or scientific. The dictionaries range from general-purpose dictionaries of the second language and thesauruses and synonym dictionaries of the first language to mono- and bilingual subject-specific technical dictionaries and glossaries. Not for nothing is the general bilingual dictionary known as a translation dictionary, although in this context translation must be seen as a traditional exercise in second/foreign language teaching and learning as well as a professional activity. The work of lexicographers and translators has much in common, and the latter can be expert informants for practicing lexicographers, more so perhaps than linguists. Technical translators must have the combination of linguistic and encyclopedic or content knowledge and an ability at written expression needed by specialist lexicographers. Literary translators must have an ability to extract meaning from text in one language and to arrive at an equivalent formulation in another that could only benefit bilingual lexicographers. They also have a highly developed feeling for sense discrimination and explanation that would make them ideal consultants on or compilers of thesauruses.

Other areas of applied linguistics are communication in the professions and languages for special purposes, both of which have at their disposal a vast range of specialized, subject-specific reference works, be it in law, medicine and engineering, or in the sciences and technologies (Bergenholtz & Tarp, 1995). Both areas draw *inter alia* on terminological lexicography or terminography and use as editors and/or consultants experts in the relevant subject area or areas being treated. Linguistic knowledge as such may or may not play a role. Dictionaries and glossaries of technical terms may be mono- and, increasingly frequently, multilingual, with international standards organizations seeking to establish equivalence of standardized terms and concepts across languages. They tend to be thematic rather than alphabetical in organization and presentation in accordance with their concentration on word meanings rather than word forms and on concepts within a given taxonomy. To handle the problem of the sheer number of terms in some areas they make full use of the possibilities now offered by electronic storage and presentation.

A further area is language planning, both corpus planning and status planning, in which the role of lexicography has been and is as central as it is complex. In the

modern period of western European lexicography, mainstream dictionaries have been absolutely instrumental in the establishment of standard varieties of the different vernaculars, especially in written use, and in their gradual emancipation from Latin. Regardless of whether they have been avowedly descriptive or explicitly prescriptive and normative in intention and approach, they have codified and helped standardize spelling, pronunciation, meaning, and usage and they have acquired the status of linguistic authorities in the eyes of many, if not most users. The authoritarian tradition is firmly established, and publishers still often appeal to it in their advertising. Indeed, the history of mainstream dictionaries can be seen *inter alia* as a history of the longstanding and ongoing conflict between the descriptive and the prescriptive, one notable chapter of which was the controversy over *Webster's Third New International Dictionary* in the 1960s (Sledd & Ebbitt, 1962; Morton, 1994). The dictionary editors favored a strongly descriptive policy aiming to record and describe authoritatively contemporary English usage as documented in extensive citation files. Where appropriate, they included clear pragmatic information on debated usage, but did not set out to be an authoritarian *arbiter usus*, being concerned to avoid prescribing or proscribing usage. A case in point is the entry on *ain't* reproduced slightly enlarged in Figure 2.1.

ain't \ 'ānt\ *also* **an't** \ " *also* 'ant *or like* AREN'T\ [prob. contr. of *are not, is not, am not, & have not*] **1 a** : are not <you ~ going> <they ~ here> <things ~ what they used to be> **b** : is not <it ~ raining> <he's here, ~ he> **c** : am not <I ~ ready> — though disapproved by many and more common in less educated speech, used orally in most parts of the U. S. by many cultivated speakers esp. in the phrase *ain't I* **2** *sub-***stand** **a** : have not <I ~ seen him> <you ~ told us> **b** : has not <he ~ got the time> <~ the doctor come yet>

Figure 2.1 Definition of *ain't* from Webster's Third New International Dictionary By permission. From *Webster's Third New International® Dictionary, Unabridged*, © 1993 by Merriam-Webster, Incorporated.

In some quarters this policy was viewed as a permissive abdication of the alleged responsibility of lexicographers not only to describe what is used and how but also to prescribe what should or should not be used. While attempts to buy out the publishers and remove the dictionary from circulation failed, the controversy produced avowedly rival works such as *The American Heritage Dictionary*, 1969, which featured usage notes informed by a panel of more than 100 representatives of the literary establishment. Its echoes can still be clearly heard in later dictionaries, where a separate usage note on *ain't*, for instance, is often longer than the actual lexicographic description itself. One example is *The Reader's Digest Great Illustrated Dictionary*, 1984 (see Figure 2.2).

ain't (aynt). *Nonstandard*. Contraction of *am not*. Also extended in use to mean *are not*, *is not*, *has not*, and *have not*.

Usage: Although widely used in colloquial speech, *ain't* is considered nonstandard by educated speakers. It should always be avoided in writing or formal speech, unless you are deliberately trying to create a humorous effect, or using a fixed phrase like *Things ain't what they used to be*. *Aren't I* (as in *aren't I coming too?*) has sometimes also been attacked on the grounds that it misleadingly suggests a corresponding form *I are*. But the full form, *am I not*, is so formal that in many contexts it may be considered ridiculously stilted, and *aren't I* is therefore a quite acceptable usage in educated British English. The form *amn't I* has some currency in regional English, especially in Scotland and Ireland, but is considered nonstandard.

Figure 2.2 Definition of *ain't* from *The Reader's Digest Great Illustrated Dictionary* By permission of *The Readers's Digest Association Limited*, *Reader's Digest Great Illustrated Dictionary* (1984).

The same European dictionaries played as much a role in status planning as in corpus planning, certainly in terms of nation building. The multi-volume scholarly and historical dictionaries inaugurated in nineteenth-century Europe, for example, were seen as national dictionaries, and the lexicography of Noah Webster was consciously and patriotically American. Nation building is not just a historical issue, but is equally important in contemporary lexicography. It underlies and supports, for instance, efforts to establish a standardized variety of «lesser-used» European languages such as Luxembourgish or Rhaeto-Romance. It is an important motivation in the lexicographical recording and describing of endangered and indigenous languages by anthropological linguists and also in the planning of comprehensive monolingual dictionaries for languages such as Samoan and Tongan which have previously relied on bilingual dictionaries with English. It is also an integral component of the codification of the different standard varieties of both contiguous and dispersed pluricentric languages. An example of the former is German, where *Österreichisches Wörterbuch*, 1951, 39th edn. 2001, a government sponsored endonormative dictionary used officially in schools, codifies Austrian Standard German as a standard variety distinct from German Standard German and Swiss Standard German. An example of the latter is English, where different native speaker standard varieties are now covered in national dictionaries, for example, *The Australian National Dictionary. A Dictionary of Australianisms on Historical Principles*, 1988; and *The Macquarie Dictionary*, 1981, 3rd edn. 1997, which advertises itself as «the arbiter of Australian English» and as «Australia's National Dictionary.»

TASKS

I. Study the topical vocabulary below. Find the contexts where the units under consideration are used in the text.

1	constituent (adj.)	[kən'stɪtjʊənt]	складовий
2	extract (v.)	[ɪks'trækt]	вибирати; діставати
3	arrive (at)	[ə'raɪv]	досягати (чого-небудь)
4	disposal (n.)	[dɪs'pəʊzəl]	право розпоряджатися; управління; використання
5	editor (n.)	[ˈɛdɪtə]	редактор
6	relevant (adj.)	[ˈrɛlɪvənt]	відповідний
7	frequently (adv.)	[ˈfri:kwəntli]	часто
8	equate	[ɪ'kweɪt]	вважати рівним; рівняти, прирівнювати; ототожнювати
9	core (adj.)	[kɔː]	основний; центральний
10	suffice	[sə'faɪs]	бути достатнім; вистачати; задовольняти
11	accordance	[ə'kɔːdəns]	відповідність, злагода
12	sheer (adj.)	[ʃɪə]	єдиний; тільки один
13	vernacular (n.)	[və'nækjʊlə]	рідна мова; національна мова
14	avowedly (adv.)	[ə'vɔɪdli]	відкрито, прямо, явно, гласно; за загальним визнанням
15	arbiter usus (n.)	[ˈɑːbɪtə usʊs]	авторитетний суддя
16	controversy (n.)	[ˈkɒntɹəvɜːsi]	суперечка, полеміка, розбіжність у думках
17	proscribe (v.)	[prəʊs'kraɪb]	виганяти, висилати засуджувати, забороняти
18	contemporary (adj.)	[kən'tempərəri]	сучасний
19	permissive (adj.)	[pə'mɪsɪv]	дозвільний, який дозволяє, який допускає
20	abdication (n.)	[ˌæbdɪ'keɪʃ(ə)n]	складання повноважень, відмова
21	alleged (adj.)	[ə'ledʒd]	нібито наявний; такий, наявність якого стверджується; заявлений
22	rival (adj.)	[ˈraɪvəl]	який суперничає, конкуруючий
23	in terms of	[ɪn tɜːmz ɒv]	з точки зору
24	indigenous (adj.)	[ɪn'dɪdʒɪnəs]	місцевий, корінний
25	taxonomy (n.)	[tæk'sɒnəmi]	систематика, таксономія
26	dispersed (adj.)	[dɪs'pɜːst]	розсіяний; розосереджений
27	contiguous (adj.)	[kən'tɪgjuəs]	суміжний, дотичний, прилягаючий; сусідній; близький
28	endonormative (adj.)	[ˌendəʊ'nɔːmətɪv]	місцево визначений
29	distinct (adj.)	[dɪs'tɪŋkt]	відмінний, який відрізняється; несхожий

30	pluricentric language (n.)	[,pluəri'sentrik 'læŋgwɪdʒ]	плюрицентрична мова (з декількома стандартними варіантами, наприклад, англійська з британським і американським варіантами)
----	----------------------------	-----------------------------	--

II. Answer the following questions to the text.

- 1) What subject area is one of the core activities of applied linguistics?
- 2) How must translation be seen in the context of the general bilingual dictionary?
- 3) What knowledge must technical translators have?
- 4) Why are literary translators ideal consultants on or compilers of thesauruses? What specific abilities do they have?
- 5) Who are used as editors and/or consultants in communication in the professions and languages for special purposes areas of applied linguistics?
- 6) How do dictionaries and glossaries of technical terms handle the problem of the sheer number of terms in some areas?
- 7) Why have mainstream dictionaries in western European lexicography acquired the status of linguistic authorities in the eyes of many users?
- 8) How can the history of mainstream dictionaries be seen? What was its most notable chapter?
- 9) Why is nation building equally important in contemporary lexicography?
- 10) Provide examples of the codification of the different standard varieties of both contiguous and dispersed pluricentric languages.

III. Match the words to their definitions.

1	discrimination	a	systematic organization of methods, rules, etc
2	tend	b	failure to fulfil a responsibility or duty
3	corpus	c	a language with several interacting codified standard versions, often corresponding to different countries
4	indigenous	d	regularly or frequently behave in a particular way or have a certain characteristic
5	contiguous	e	recognition and understanding of the difference between one thing and another
6	constituent	f	vary or extend between specified limits
7	abdication	g	originating or occurring naturally in a particular place
8	range	h	next or together in sequence
9	codification	i	servicing to compose or make up a thing
10	pluricentric language	j	a collection of written or spoken material in machine-readable form, assembled for the purpose of linguistic research

IV. Match the parts below to complete a single syntactic unit from the text.

1	Lexicography is thus not only a field of professional, commercial, and academic activity in its own right,	a	«lesser-used» European languages such as Luxembourgish or Rhaeto-Romance.
2	Professional translators need and use dictionaries of different types	b	that would make them ideal consultants on or compilers of thesauruses.
3	The work of lexicographers and translators has much in common,	c	possibilities now offered by electronic storage and presentation.
4	Literary translators also have a highly developed feeling for sense discrimination and explanation	d	but did not set out to be an authoritarian <i>arbiter usus</i> , being concerned to avoid prescribing or proscribing usage.
5	Both areas draw <i>inter alia</i> on terminological lexicography or terminography and	e	they have codified and helped standardize spelling, pronunciation, meaning, and usage and they have acquired the status of linguistic authorities in the eyes of many users.
6	Dictionaries and glossaries of technical terms tend to be thematic rather than alphabetical in organization and presentation in accordance	f	and the latter can be expert informants for practicing lexicographers, more so perhaps than linguists.
7	To handle the problem of the sheer number of terms in some areas they make full use of the	g	use as editors and/or consultants experts in the relevant subject area or areas being treated.
8	Regardless of whether they have been avowedly descriptive or explicitly prescriptive and normative in intention and approach,	h	according to the nature of the translation, general or specialized, literary or scientific.
9	Where appropriate, they included clear pragmatic information on debated usage,	i	with their concentration on word meanings rather than word forms and on concepts within a given taxonomy.
10	Contemporary lexicography underlies and supports efforts to establish a standardized variety of	j	but also very much an integral part of applied linguistics and its constituent subject areas.

V. Fill in the gaps using the appropriate words from the box. Underline the words inserted.

glossaries	<i>inter alia</i>	endonormative	in terms of	indigenous
comprehensive	distinct	editors	avowedly rival	mainstream
prescriptive	contemporary	controversy	translators	core ability
	equate	vernaculars	equivalence	

1) First and second language teaching and learning is an area that some virtually with applied linguistics and that is by common consent certainly one of the activities of applied linguistics.

2) Technical must have the combination of linguistic and encyclopedic or content knowledge and an at written expression needed by specialist lexicographers.

3) Dictionaries and of technical terms may be mono- and, increasingly frequently, multilingual, with international standards organizations seeking to establish of standardized terms and concepts across languages.

4) Western European lexicography dictionaries have been absolutely instrumental in the establishment of standard varieties of the different, especially in written use, and in their gradual emancipation from Latin.

5) The history of mainstream dictionaries can be seen as a history of the longstanding and ongoing conflict between the descriptive and the

6) The dictionary favored a strongly descriptive policy aiming to record and describe authoritatively English usage as documented in extensive citation files.

7) The over *Webster's Third New International Dictionary* produced works such as *The American Heritage Dictionary*, which featured usage notes informed by a panel of more than 100 representatives of the literary establishment.

8) European dictionaries played as much a role in status planning as in corpus planning, certainly nation building.

9) Nation building is an important motivation in the lexicographical recording and describing of endangered and languages by anthropological linguists and also in the planning of monolingual dictionaries.

10) *Österreichisches Wörterbuch*, a government sponsored dictionary used officially in schools, codifies Austrian Standard German as a standard variety from German Standard German and Swiss Standard German.

VI. Find the English equivalents from the text to the following words and word-combinations.

1	складові предметні області	
2	тематичні технічні словники	
3	практикуючі лексикографи	
4	корпусне планування	
5	відкрито описовий	
6	корінні мови	
7	редактори словників	
8	відповідна предметна сфера	

5) In some quarters this policy was been viewed as a permissive abdication of the alleged responsibility of lexicographers not only to describe that is used and how but also to prescribe that should or should not be used.

6) The multi-volumes scholarly and historical dictionaris inaugurated in nineteenth-century Europe were saw as national dictionaris.

7) In this context translation must to be seen as a traditional exercise in first/foreign language teaching and learning as well as a profeccional activity.

8) A further area be language planning, both corpus planning and status planning, in whose the role of lexicografy have been and is as central as it is complex.

9) Nation building were the integral component of the codification of the different standard varieties of both contiguous and dispersed pluricentric languages.

10) A example of the latter is English, were different native speaker standard varieties are now covered in national dictionaris.

X. Translate the following sentences into English, using the topical vocabulary under consideration.

1) Літературні перекладачі повинні уміти виділити значення з тексту однією мовою та отримати еквівалентне формулювання іншою, що може стати в нагоді двомовним лексикографам.

2) Такі галузі прикладної лінгвістики як професійне спілкування та мови для спеціальних цілей мають у своєму розпорядженні широкий спектр спеціалізованих довідників.

3) Словники та глосарії технічних термінів, як правило, є тематичними, а не алфавітними за організацією та викладом відповідно до їх концентрації на значеннях, а не на формах слів та поняттях в межах певної систематики.

4) Строго описова політика розглядалася як дозвільна відмова від передбачуваної відповідальності лексикографів не лише, щоб описувати, що і як використовується, але також і для того, щоб визначити, що слід або не слід використовувати.

5) Збереження нації є важливою мотивацією для лексикографічного запису та опису антропологічними лінгвістами мов, що знаходяться під загрозою зникнення та мов корінних народів, а також для планування комплексних одномовних словників.

UNIT 4

LANGUAGE CORPORA. EMPIRICAL LINGUISTICS

<...> Since the 1990s, a «language corpus» usually means a text collection which is:

- large: millions, or even hundreds of millions, of running words, usually sampled from hundreds or thousands of individual texts;
- computer-readable: accessible with software such as concordancers, which can find, list and sort linguistic patterns;
- designed for linguistic analysis: selected according to a sociolinguistic theory of language variation, to provide a sample of specific text-types or a broad and balanced sample of a language.

Much «corpus linguistics» is driven purely by curiosity. It aims to improve language description and theory, and the task for applied linguistics is to assess the relevance of this work to practical applications. Corpus data are essential for accurately describing language use, and have shown how lexis, grammar, and semantics interact. This in turn has applications in language teaching, translation, forensic linguistics, and broader cultural analysis. In limited cases, applications can be direct. For example, if advanced language learners have access to a corpus, they can study for themselves how a word or grammatical construction is typically used in authentic data. Hunston (2002, pp. 170–84) discusses data-driven discovery learning and gives further references.

However, applications are usually indirect. Corpora provide observable evidence about language use, which leads to new descriptions, which in turn are embodied in dictionaries, grammars, and teaching materials. Since the late 1980s, the influence of this work is most evident in new monolingual English dictionaries (CIDE, 1995; COBUILD, 1995a; LDOCE, 1995; OALD, 1995) and grammars (e.g., COBUILD, 1990), aimed at advanced learners, and based on authentic examples of current usage from large corpora. Other corpus-based reference grammars (e.g., G. Francis, Hunston, & Manning, 1996, 1998; Biber et al., 1999) are invaluable resources for materials producers and teachers.

Corpora are just sources of evidence, available to all linguists, theoretical or applied. A sociolinguist might use a corpus of audio-recorded conversations to study relations between social class and accent; a psycholinguist might use the same corpus to study slips of the tongue; and a lexicographer might be interested in the frequency of different phrases. The study might be purely descriptive: a grammarian might want to know which constructions are frequent in casual spoken language but rare in formal written language. Or it might have practical aims: someone writing teaching materials might use a specialized corpus to discover which grammatical constructions occur in academic research articles; and a forensic linguist might want to study norms of language use, in order to estimate the likelihood that linguistic patterns in an anonymous letter are evidence of authorship.

So, if corpus linguistics is not (necessarily) applied linguistics, and is not a branch of linguistics, then what is it? It is an empirical approach to studying language, which uses observations of attested data in order to make generalizations about lexis, grammar, and semantics. Corpora solve the problem of observing patterns of language use. It is these patterns which are the real object of study, and it is findings about

recurrent lexico-grammatical units of meaning which have implications for both theoretical and applied linguistics.

Large corpora have provided many new facts about words, phrases, grammar, and meaning, even for English, which many teachers and linguists assumed was fairly well understood. Valid applications of corpus studies depend on the design of corpora, the observational methods of analysis, and the interpretation of the findings.

Applied linguists must assess this progression from evidence to interpretation to applications, and this chapter therefore has sections on empirical linguistics (pre- and post-computers), corpus design and software, findings and descriptions, and implications and applications.

I use these presentation conventions. LEMMAS (LEXEMES) are in upper case. *Word-forms* are lower case italics. ‘Meanings’ are in single quotes. Collocates of a node are in angle brackets: UNDERGO <surgery>.

Since corpus study gives priority to observing millions of running words, computer technology is essential. This makes linguistics analogous to the natural sciences, where it is observational and measuring instruments (such as microscopes, radio telescopes, and x-ray machines) which extended our grasp of reality far beyond «the tiny sphere attainable by unaided common sense» (Wilson, 1998, p. 49).

Observation is not restricted to any single method, but concordances are essential for studying lexical, grammatical, and semantic patterns. Printed concordance lines (see Appendix) are limited in being static, but a computer accessible concordance is both an observational and experimental tool, since ordering it alphabetically to left and right brings together repeated lexico-grammatical patterns. A single concordance line, on the horizontal axis, is a fragment of language use (*parole*). The vertical axis of a concordance shows repeated co-occurrences, which are evidence of units of meaning in the language system (*langue*).

The tiny sample of concordance lines in the Appendix is not representative. In a real study one might have hundreds or thousands of concordance lines, but I can use this sample for illustration. Concordance data are often especially good at distinguishing words with related propositional meanings, but different connotations and patterns of usage. The Appendix therefore gives examples of *endure*, *persevere*, *persist*, and *undergo*, which are all used to talk about unpleasant things which last a long time, but which differ in their surrounding lexis and grammar. For example, we can observe how the word-form *persist* occurs in distinct constructions. When its subject is an abstract noun, it often denotes unpleasant things (*fears*, *problems*), often medical (*symptoms*, *headaches*), and often has a time reference (*for over a year*, *for up to six weeks*). Alternatively, when the subject of *persist in* is animate, it is often used of someone who persists, often unreasonably or *in the face of* opposition, in doing something which is difficult or disapproved of. Such recurrent co-occurrence patterns provide evidence of typical meaning and use.

It is sometimes objected that concordances place words in small, arbitrary contexts, defined by the width of a computer screen, and ignore contexts of communication. However, it is an empirical finding that evidence for the meaning of a node word often occurs within a short span of co-text. In addition, corpora allow

individual utterances to be interpreted against the usage of many speakers and the intertextual norms of general language use.

The observation of large publicly available data sets implies (a weak sense of) inductive methods, that is, gathering many observations and identifying patterns in them. This does not imply mechanical methods of generalizing from observations, but (as Fillmore, 1992, pp. 38, 58 puts it) a combination of corpus linguistics (getting the facts right) and armchair linguistics (thinking through the hypotheses that corpus data suggest). It does mean, however, that corpus study belongs to a philosophical tradition of empiricism. Contrary to a loss of confidence, from Saussure to Chomsky, in the ability to observe real language events, corpora show that language use is highly patterned.

Although there are limitations on corpus design (see below), and although we can never entirely escape subjective interpretations, corpora allow «a degree of objectivity» about some central questions, «where before we could only speculate» (Kilgarriff, 1997, p. 137). There are no automatic discovery procedures, but inductive generalizations can be tested against observations in independent corpora.

Corpus methods therefore differ sharply from the view, widely held since the 1960s, that native speaker introspection gives special access to linguistic competence. Although linguists' careful analyses of their own idiolects have revealed much about language and cognition, there are several problems with intuitive data and misunderstandings about the relation between observation and intuition in corpus work. Intuitive data can be circular: data and theory have the same source in the linguist who both proposes a hypothesis and invents examples to support or refute it. They can be unreliable or absent: many facts about frequency, grammar, and meaning are systematic and evident in corpora, but unrecorded in pre-corpus dictionaries. They are narrow: introspection about small sets of invented sentences cannot be the sole and privileged source of data.

There is no point in being purist about data, and it is always advisable to compare data from different sources, both independent corpora, and also introspection and experiments. Corpus study does not reject intuition, but gives it a different role. Concordances focus intuition, and this «confirms rather than produces the data» (de Beaugrande, 1999, pp. 247–8). Without this retrospective competence, native speakers could not recognize untypical collocations in literature, advertising, or jokes. We cannot know in advance what kinds of evidence might bear on a theory of linguistic competence (as even Chomsky, 2000, pp. 139–40 admits). Nevertheless, with some striking exceptions (Fillmore, 1992), cognitive approaches have neglected corpus data on recurrent semantic patterns as evidence of cognitive structures. <...>

TASKS

I. Study the topical vocabulary below. Find the contexts where the units under consideration are used in the text.

1	corpus (<i>pl.</i> corpora) (n.)	[ˈkɔ:pəs]	корпус; збірник
2	sample (n.)	[ˈsɑ:mpəl]	зразок, приклад

3	concordance (n.)	[kən'kɔ:dəns]	узгодження, відповідність
4	pattern (n.)	['pætən]	модель, зразок
5	assess (v.)	[ə'ses]	оцінювати
6	interact (v.)	[,ɪntər'ækt]	взаємодіяти
7	forensic (adj.)	[fə'rensɪk]	судовий, судочинний
8	authentic (adj.)	[ɔ:'θentɪk]	автентичний, оригінальний
9	reference (n.)	['refrəns]	посилання; довідка
10	embody (v.)	[ɪm'bɒdi]	включати; об'єднувати
11	observable (adj.)	[əb'zɜ:vəbl]	видимий, помітний, доступний для огляду
12	monolingual (adj.)	[,mɒnəʊ'lɪŋgwəl]	одномовний
13	invaluable (adj.)	[ɪn'veljʊəbl]	безцінний
14	evidence (n.)	['eɪdəns]	свідчення
15	descriptive (adj.)	[dɪs'krɪptɪv]	описовий
16	authorship (n.)	['ɔ:θəʃɪp]	авторство
17	refute (v.)	[rɪ'fju:t]	спростовувати
18	denote (v.)	[dɪ'nəʊt]	позначати
19	persist (v.)	[pə'sɪst]	наполягати
20	disapprove (v.)	[,dɪsə'pru:v]	не схвалювати, спростовувати
21	recurrent (adj.)	[rɪ'kɹərənt]	періодичний, повторюваний
22	likelihood (n.)	['laɪklɪhʊd]	подібність, схожість
23	hypothesis (n.)	[haɪ'pɒθɪsɪs]	гіпотеза
24	circular (adj.)	['sɜ:kjʊlə]	круговий, що рухається по колу
25	generalization (n.)	[,dʒenərəlaɪ'zeɪʃən]	узагальнення
26	valid (adj.)	['vælɪd]	дійсний; чинний
27	unreliable (adj.)	[,ʌnrɪ'laɪəbl]	ненадійний
28	attainable (adj.)	[ə'teɪnəbl]	досяжний
29	unaided (adj.)	[ʌn'eɪdɪd]	без сторонньої допомоги
30	axis (n.)	['æksɪs]	вісь

II. Answer the following questions to the text.

- 1) Why is computer technology essential in corpus study? What makes linguistics analogous to the natural sciences?
- 2) Are printed or computer accessible concordances more efficient? Give your reasons.
- 3) What problems do corpora solve? Is corpus linguistics equal or similar to applied linguistics?
- 4) What are the reasons for corpus study to belong to a philosophical tradition of empiricism?
- 5) Why do you think language use is highly patterned?
- 6) Do concordances really ignore contexts of communication? Give your reasons.
- 7) What are the benefits of using language corpora?

- 8) Can a sample of concordance lines in the Appendix be representative?
- 9) What are concordance data often good at?
- 10) What is the vertical axis of a concordance? How is it related to the horizontal axis?

III. Match the words to their definitions.

1	inductive	a	a word or phrase that someone speaks;
2	introspection	b	the fact that two things have similar features or qualities; a list produced by a computer that shows every example of a particular word that is used in the books, newspapers etc stored on the computer;
3	idiolect	c	something that you suggest is true, although you do not say it directly;
4	cognitive	d	the process of carefully examining your own feelings, thoughts, and ideas;
5	corpus	e	a piece of computer software that is designed to do a particular job;
6	application	f	one person's individual way of speaking or writing a language;
7	implication	g	something that is connected with recognizing and understanding things;
8	utterance	h	not based on any particular plan, or not done for any particular reason; used about actions that are considered to be unfair;
9	arbitrary	i	a collection of written and spoken language stored on computer and used for language research and writing dictionaries;
10	concordance	g	reasoning from particular facts or ideas to a general rule or law;

IV. Match the parts below to complete a single syntactic unit from the text.

1	There are no automatic discovery procedures,	a	cannot be the sole and privileged source of data.
2	There is no point in being purist about data, and it is always advisable to compare data from different sources,	b	but gives it a different role.
3	A single concordance line, on the horizontal axis,	c	in the ability to observe real language events, corpora show that language use is highly patterned.
4	Contrary to a loss of confidence, from Saussure to Chomsky,	d	which are evidence of units of meaning in the language system (<i>langue</i>).

5	They are narrow: introspection about small sets of invented sentences	e	but different connotations and patterns of usage.
6	It is an empirical approach to studying language,	f	the observational methods of analysis, and the interpretation of the findings.
7	Corpus study does not reject intuition,	g	both independent corpora, and also introspection and experiments.
8	Concordance data are often especially good at distinguishing words with related propositional meanings,	h	is a fragment of language use (<i>parole</i>).
9	Valid applications of corpus studies depend on the design of corpora,	i	which uses observations of attested data in order to make generalizations about lexis, grammar, and semantics.
10	The vertical axis of a concordance shows repeated co-occurrences,	j	but inductive generalizations can be tested against observations in independent corpora.

V. Fill in the gaps using the appropriate words from the box. Underline the words inserted:

semantic hypothesis collocations formal descriptive evidence intertextual empirical concordance introspection grammatical
--

- 1) Without this retrospective competence, native speakers could not recognize untypicalin literature, advertising, or jokes.
- 2) In a real study one might have hundreds or thousands oflines, but I can use this sample for illustration.
- 3) In addition, corpora allow individual utterances to be interpreted against the usage of many speakers and thenorms of general language use.
- 4) Nevertheless, with some striking exceptions, cognitive approaches have neglected corpus data on recurrentpatterns as evidence of cognitive structures.
- 5) Such recurrent co-occurrence patterns provideof typical meaning and use.
- 6) Intuitive data can be circular: data and theory have the same source in the linguist who both proposes aand invents examples to support or refute it.
- 7) Corpus methods therefore differ sharply from the view, widely held since the 1960s, that native speakergives special access to linguistic competence.
- 8) Observation is not restricted to any single method, but concordances are essential for studying lexical,, and semantic patterns.
- 9) However, it is anfinding that evidence for the meaning of a node word often occurs within a short span of co-text.
- 10) The study might be purely.....: a grammarian might want to know which constructions are frequent in casual spoken language but rare inwritten language.

VI. Find the English equivalents from the text to the following words and word-combinations:

1	знати заздалегідь	
2	спростувати гіпотезу	
3	індуктивне узагальнення	
4	одномовні словники	
5	обмеження щодо побудови корпусу	
6	горизонтальна вісь	
7	вертикальна вісь	
8	мовна система	
9	ступінь об'єктивності	
10	індивідуальне висловлення	
11	ігнорувати комунікативний контекст	
12	розрізняти слова	
13	когнітивний підхід	
14	суб'єктивне тлумачення	
15	ненадійні дані	

VII. Find appropriate synonyms (if any) and antonyms (if any) to the words below:

№	Word	Synonym	Antonym
1	observable (adj.)		
2	valuable (adj.)		
3	attainable (adj.)		
4	generalization (n.)		
5	approve (v.)		
6	concordance (n.)		
7	valid (adj.)		
8	implication (n.)		
9	aided (adj.)		
10	competence (n.)		

VIII. Provide all the possible derivatives to the word forms below. Make up sentences with the words derived.

induce, deduce, introspect, apply, pure, construct, occur, valid, aid, recognize, public

IX. Correct the possible mistakes in terms of grammar, spelling and inappropriate use of vocabulary in the sentences below:

- 1) The tiny sample of concordance lines at the Appendix is not representative.

2) It is sometimes been objected that concordances place words in small, arbitrary contexts, defined by width of computer screen, and ignore contexts about communication.

3) We must not know for advance what kinds of evidence might bear on a theory of linguistic competence.

4) Corpora provides observable evidence about language use, which leads to new descriptions, which on turn are embodied in dictionaries, grammars, and teaching materials.

5) Applied linguists must assess these progression from evidence to interpretation to applications, and this chapter therefore have sections on impirical linguistics (pre- and post-computers), corpus deizgn and software, findings and descriptions, and implications and fapplications.

6) For example, we can obzerve how the word-form *persist* occur in distinct construction.

7) However, it is empirical finding that evidence by the meaning of node word often occur within a short span of co-text.

8) Corpus data is essential for accurately describing language use, and has shown how lexis, grammar, and semantics interact.

9) So, if corpus linguistics are not (necessarily) applied linguistics, and are not branch of linguistics, then what it is?

10) It aims for improve language description and theory, and the task for applied linguistics to assess the relevance of this work to practical applications.

X. Translate the following sentences into English, using the topical vocabulary under consideration.

1) Активний розвиток корпусної лінгвістики останніми роками засвідчує, що цей напрям стає одним з провідних серед прикладних мовознавчих досліджень.

2) Перспективність, необхідність та актуальність подібних повнотекстових корпусів є очевидною – філологи (чи представники інших наукових дисциплін) отримують увесь масив потрібних для досліджень художніх творів (або інших текстів) з повним інструментарієм у вигляді систем пошуку і цитування, бібліографії, тезаурусів і глосаріїв, а у випадку поєднання з лінгвістичним корпусом – ще й повну лінгвістичну інформацію про тексти та їхні одиниці.

3) Такий корпус відрізняється від електронної бібліотеки повнотою охоплення матеріалу та підходами до його параметризації й опису, сферами застосування і цільовою аудиторією.

4) З урахуванням переваг і недоліків існуючих корпусів ми спробуємо описати структуру і представлення даних корпусу художніх текстів.

5) Цей корпус поєднується з лінгвістичним, що створює широкі можливості для лінгвістичних або міждисциплінарних досліджень. Певним недоліком обох корпусів є недостатня зручність у представленні метатекстових даних, у класифікації та пошуку за авторами тощо.

UNIT 5

MODERN CORPORA AND SOFTWARE

Modern computer-assisted corpus study is based on two principles.

1 *The observer must not influence what is observed.* What is selected for observation depends on convenience, interests and hypotheses, but corpus data are part of natural language use, and not produced for purposes of linguistic analysis.

2 *Repeated events are significant.* Quantitative work with large corpora reveals what is central and typical, normal and expected. It follows (Teubert, 1999) that corpus study is inherently sociolinguistic, since the data are authentic acts of communication; inherently diachronic, since the data are what has frequently occurred in the past; and inherently quantitative. This disposes of the frequent confusion that corpus study is concerned with «mere» performance, in Chomsky's (1965, p. 3) pejorative sense of being characterized by «memory limitations, distractions, shifts of attention and interest, and errors.» The aim is not to study idiosyncratic details of performance which are, by chance, recorded in a corpus. On the contrary, a corpus reveals what frequently recurs, sometimes hundreds or thousands of times, and cannot possibly be due to chance.

Available corpora

Any list of extant corpora would be quickly out of date, but there are two sets of important distinctions between

- small first generation corpora from the 1960s onward and much larger corpora from the 1990s, and
- carefully designed reference corpora, small and large, and other specialized corpora, opportunistic text collections, archives and the like.

The first computer-readable corpora, compiled in the 1960s, are very small by contemporary standards, but still useful because of their careful design. The Brown corpus (from Brown University in the USA) is one million words of written American English, sampled from texts published in 1961: both informative prose, from different text-types (e.g., press and academic writing), and different topics (e.g., religion and hobbies); and imaginative prose (e.g., detective fiction and romance). Parallel corpora were designed to enable comparative research: the LOB corpus (from the universities of Lancaster, Oslo, & Bergen) contains British data from 1961; Frown and FLOB (from Freiburg University, Germany) contain American and British data from 1991; and ICE (International Corpora of English) contains regional varieties of English, such as Indian and Australian. Similar design principles underlie the Lund corpus of spoken British English (from University College London and Lund University), which contains around half a million words, divided into samples of the usage of adult, educated, professional people, including face-to-face and telephone conversations, lectures and discussions.

By the late 1990s, some corpora consisted of hundreds of millions of words. The Bank of English (at COBUILD in Birmingham, UK) and the British National Corpus (BNC) had commercial backing from publishers, who have used the corpora to produce dictionaries and grammars. The 100-million-word BNC is also carefully designed to include demographically and stylistically defined samples of written and spoken language. The Bank of English arguably over-emphasizes mass media texts, but these

are very influential, and it still has a range of text-types and advantages of size: over 400 million words by 2001. Because constructing large reference corpora is so expensive, it may be that huge new corpora cannot again be created in the near future. These corpora will remain standard reference points, which can be supplemented by small specialized corpora, designed by individual researchers, and by large opportunistic collections.

Many other corpora for English, and increasingly for other languages, are available (see Michael Barlow's website: address in the further reading section below).

Corpus design

Some basic principles of corpus design (Kennedy, 1998, pp. 13–87; Hunston, 2002, pp. 25–37) are simple enough. A corpus which claims to be a balanced sample of language use must represent variables of demography, style, and topic, and must include texts which are spoken and written, casual and formal, fiction and non-fiction, which vary in level (e.g., popular and technical), age of audience (e.g., children or adults), and sex and geographical origin of author, and which illustrate a wide range of subject fields (e.g., natural and social sciences, commerce, and leisure). However, no corpus can truly represent a whole language, since no one quite knows what should be represented. It is not even obvious what are appropriate proportions of mainstream text-types such as quality newspapers, literary classics, and everyday conversation, much less text-types such as newspaper ads, business correspondence, and church sermons. (Even carefully designed corpora have odd gaps: despite their influence as a text-type, textbooks are not represented in Brown and LOB.) A realistic aim is a corpus which samples widely, is not biased toward data which are easy to collect (e.g., mass media texts), does not under-represent data which are difficult to collect (e.g., casual conversation), and is not unbalanced by text-types which have over-specialized lexis and grammar (e.g., academic research articles).

Since large quantities of data are necessary in order to study what is typical and probable, an important criterion is size, which is usually measured in important: How large is the corpus measured as word-types (i.e., different words), or as the number of different texts or text-types it contains? A corpus might be very large, but consist entirely of American newswire texts, with a correspondingly narrow vocabulary. One can also attempt to measure linguistic influence: How large is the audience for the texts in the corpus? Casual conversation is a linguistic universal, but a typical conversation is private, whereas the language of the mass media is public, and therefore much more influential. And whereas some texts are heard once by millions of people (sports commentaries), others (literary classics) are constantly re-read over generations. A reception index, which weights texts by their audience size, can be constructed at least in a rough way.

In summary, any corpus is a compromise between the desirable and the feasible, and although design criteria cannot be operationalized, large balanced corpora reveal major regularities in language use. In any case, there is no reason to rely on any single corpus, and it is often advisable to combine large general corpora designed according to principles of sociolinguistic variation, small corpora from specific knowledge domains (since much lexis is determined by topic), and opportunistic text collections.

Huge text collections (such as the world-wide-web) can be used to study patterns which do not occur even in large reference corpora. For example, concordance lines in the Appendix show that *undergo* is typically used of someone who is forced to undergo something unpleasant, often a medical procedure or a test of some kind, or of a situation which undergoes some profound and often unwelcome change. Typical examples are:

had to *undergo* a stringent medical examination
is about to *undergo* dramatic changes

However generalizations must be checked against potential counter-examples. First, comparison of different text-types shows that, in scientific and technical English, *undergo* usually has no unpleasant connotations. An example from the BNC (which still involves ‘change’) is:

the larvae *undergo* a complex cycle of 12 stages

Second, people ‘unwillingly’ undergo unpleasant experiences. But does the collocation *willingly UNDERGO* occur and does it provide a counter-example? Now we have a problem: the lemma UNDERGO is fairly frequent (around 25 occurrences per million words in the BNC), and even *willingly* is not infrequent (around 5 per million), but the combination *willingly UNDERGO* does not occur at all in the 100-million-word BNC. However, a search of the world-wide-web quickly provided 200 examples, which revealed another pattern: people *willingly undergo* a sacrifice for the sake of others or for the sake of religious beliefs. Characteristic examples are:

one can *willingly undergo* some painful experience for one who is dearly loved
sufferings and dangers the early Christians *willingly underwent* for the sake of . . .

A corpus is specifically designed for language study, but other text collections (such as newspapers on CD-ROM) can be useful for some types of study. Again, I see no point in being purist about data, as long as their source is stated in a way which allows findings to be assessed. The world-wide-web has the advantage of enormous size, but it is impossible to characterize its overall range of texts. Words and phrases in the world-wide-web can be searched for directly with search engines, or with a concordancer which uses these engines, such as one developed at the University of Liverpool (<http://www.webcorp.org.uk/>).

Raw, lemmatized, and annotated corpora

A corpus may consist of raw text (strings of orthographic word-forms), or it can be lemmatized, and annotated or tagged, for intonation (for spoken corpora), grammatical or semantic categories. Part-of-speech tagging allows a corpus to be searched for grammatical constructions, such as adjective-noun combinations (*persistent rain*), and make it possible to study the frequency of grammatical categories in different text-types (e.g., see Biber, Conrad, & Reppen, 1998, pp. 59–65 on nominalizations; and Carter & McCarthy, 1999, on passives). Information on the frequencies of lexical and grammatical features can indicate to language teachers where it is worthwhile devoting pedagogical effort (Kennedy, 1998, pp. 88–203).

Nevertheless, a simple example illustrates the value of working with raw text. Many occurrences of the lemmas of the verbs PERSIST and ENDURE share the semantic and pragmatic features that something ‘unpleasant’ is lasting ‘for a long

time'. However, although the adjectives *persistent* and *enduring* also share the feature «for a long time», their typical collocates show their very different connotations:

persistent <ambiguity, bleeding, confusion, headaches>

enduring <appeal, legacies, peace, significance, values>

Traditionally, lemmas comprise words within a single part of speech. *Persistent* is an adjective, and shares the connotations of the verb PERSIST. *Enduring* might be considered an adjective, or the *-ing* form of the verb ENDURE, but has very different connotations from the verb.

In addition, the grammatical categories needed for unrestricted naturally occurring text can be very different from those required for the invented data described in abstract syntax. This draws into question centuries-old assumptions about the part-of-speech system (Sinclair, 1991, pp. 81–98; Sampson, 1995; Hallan, 2001). So, tagging may make unwarranted assumptions about appropriate grammatical categories. Again, the principle is that observer and data should be kept independent. The facts never «speak for themselves,» but inductive methods aim for the minimum of preconceptions. How to lemmatize words is by no means always obvious, and there are no standardized systems for part-of-speech tagging (Atwell et al., 2000) or full parsing (Sampson, 1995).

TASKS

I. Study the topical vocabulary below. Find the contexts where the units under consideration are used in the text.

1	computer-assisted (adj.)	[kəm'pjʊ:tər-ə'sɪstɪd]	(проведений) за допомогою еом, машинний
2	significant (adj.)	[sɪg'nɪfɪkənt]	значимий, несучий значення (про афікс); значущий
3	inherently (adv.)	[ɪn'hɪərəntli]	за своєю суттю, у своїй основі; від природи
4	pejorative (adj.)	['pi:dʒərətɪv]	зневажливий, пейоративний
5	shift (n.)	[ʃɪft]	переміщення, перестановка, перенесення; переміна; зміна
6	idiosyncratic (adj.)	[,ɪdɪəsɪŋ'krætɪk]	своєрідний, унікальний характерний, особливий
7	recur (v.)	[rɪ'kɜː]	відбуватися знову
8	extant (adj.)	[eks'tænt]	існуючий; збережений; який дійшов до нас
9	distinction (n.)	[dɪs'tɪŋkʃən]	розрізнення, розпізнавання; відмінність, розходження
10	onward (adv.)	['ɒnwəd]	уперед, далі
11	comparative (adj.)	[kəm'pærətɪv]	порівняльний
12	underlie (v.)	[,ʌndə'laɪ]	лежати в основі (чого-н.)
13	sermon(n.)	['sɜːmən]	проповідь; повчання, нотація
14	gap (n.)	[gæp]	огріх, прогалина (у знаннях)

15	criteria (n.)	[kraɪ'tɪəriə]	множ. від criterion – критерій, мірило, ознака
16	feasible (adj.)	['fi:zəbl]	можливий, імовірний
17	pattern (n.)	['pætən]	зразок, приклад
18	concordance (n.)	[kən'kɔ:dəns]	відповідність, конкорданція, алфавітний покажчик слів
19	sacrifice (n.)	['sækrɪfaɪs]	жертва
20	assess (v.)	[ə'ses]	оцінювати, давати оцінку
21	raw (unannotated) corpus (n.)	[rɔ: (,ʌn'ænəuteɪtɪd) 'kɔ:pəs]	не анований корпус
22	lemmatized corpus (n.)	[lemmə'taɪzd 'kɔ:pəs]	лематизований корпус англ. lemmatize – формувати гнізда; визначати головні слова і сортувати словосполучення
23	annotated corpus (n.)	['ænəuteɪtɪd 'kɔ:pəs]	анотований корпус
24	raw text (n.)	[rɔ: tɛkst]	необроблений текст
25	annotated(adj.)	['ænəuteɪtɪd]	з коментарями, примітками, анований
26	occurrence (n.)	[ə'klʌrəns]	випадок; частотність; поширеність
27	comprise (v.)	[kəm'praɪz]	включати, містити в собі, становити; охоплювати; складатися з; входити до складу
28	assumption (n.)	[ə'sʌmpʃ(ə)n]	припущення, допущення; вихідне положення
29	preconception (n.)	[,pri:kən'sepʃən]	заздалегідь складена думка; упереджена думка, упередження
30	parsing (n.)	['pɑ:zɪŋ]	синтаксичний аналіз; аналіз, розбір

II. Answer the following questions to the text.

- 1) What are two principles of modern computer-assisted corpus study?
- 2) What do the following abbreviations stand for: LOB, FLOB, ICE? What kind of data do they contain?
- 3) Why will the late 1990s corpora remain standard reference points in the near future?
- 4) What must a corpus which claims to be a balanced sample of language use represent and include?
- 5) How can you comment upon the following statement from the text: «Any corpus is a compromise between the desirable and the feasible.»
- 6) What is the difference between a corpus and other text collections (such as newspapers on CD-ROM)?

- 7) Can huge text collections be used to study patterns which do not occur even in large reference corpora? Explain your point of view. Provide examples.
- 8) Why is it impossible to characterize the world-wide-web overall range of texts?
- 9) What does part-of-speech tagging allow a corpus?
- 10) How do lemmas traditionally comprise words? Provide examples.

III. Match the words to their definitions.

1	sermon	a	a principle or standard by which something may be judged or decided
2	gap	b	be the cause or basis of (something)
3	criterion	c	a small part or quantity intended to show what the whole is like
4	parsing	d	a talk on a religious or moral subject, especially one given during a church service
5	underlie	e	shared by most people and regarded as normal or conventional
6	occur	f	add an extra element or amount to
7	mainstream	g	resolving (a sentence) into its component parts and describing their syntactic roles
8	sample	h	the rise and fall of the voice in speaking
9	supplement	i	exist or be found to be present in a place or under a particular set of conditions
10	intonation	g	a difference, especially an undesirable one, between two views or situations

IV. Match the parts below to complete a single syntactic unit from the text.

1	Quantitative work with large corpora reveals	a	who have used the corpora to produce dictionaries and grammars.
2	The Brown corpus is one million words	b	that something 'unpleasant' is lasting 'for a long time'.
3	The Bank of English and the British National Corpus had commercial backing from publishers,	c	whereas the language of the mass media is public, and therefore much more influential.
4	However, no corpus can truly represent a whole language,	d	<i>undergo</i> usually has no unpleasant connotations.
5	Casual conversation is a linguistic universal, but a typical conversation is private,	e	what is central and typical, normal and expected.
6	Comparison of different text-types shows that, in scientific and technical English,	f	grammatical constructions, such as adjective-noun combinations.
7	The world-wide-web has the advantage of enormous size,	g	since no one quite knows what should be represented.

8	Part-of-speech tagging allows a corpus to be searched for	h	and there are no standardized systems for part-of-speech tagging or full parsing.
9	Many occurrences of the lemmas of the verbs PERSIST and ENDURE share the semantic and pragmatic features	i	of written American English, sampled from texts published in 1961.
10	How to lemmatize words is by no means always obvious,	j	but it is impossible to characterize its overall range of texts.

V. Fill in the gaps using the appropriate words from the box. Underline the words inserted:

due to chance	syntax	connotations	raw text	collocates
intonation	concordance lines	demographically	unrestricted	
computer-readable	assessed	sacrifice	recurs	pattern
samples	reference points	opportunistic		

1) A corpus reveals what frequently, sometimes hundreds or thousands of times, and cannot possibly be

2) The first..... corpora, compiled in the 1960s, are very small by contemporary standards, but still useful because of their careful design.

3) The 100-million-word BNC is also carefully designed to include and stylistically defined of written and spoken language.

4) These corpora will remain standard, which can be supplemented by small specialized corpora, designed by individual researchers, and by large collections.

5) For example, show that *undergo* is typically used of someone who is forced to undergo something unpleasant, often a medical procedure or a test of some kind, or of a situation which undergoes some profound and often unwelcome change.

6) However, a search of the world-wide-web quickly provided 200 examples, which revealed another: people *willingly undergo* a for the sake of others or for the sake of religious beliefs.

7) There is no point in being purist about data, as long as their source is stated in a way which allows findings to be

8) A corpus may consist of, or it can be lemmatized, and annotated or tagged, for, grammatical or semantic categories.

9) Although the adjectives *persistent* and *enduring* also share the feature «for a long time», their typical show their very different

10) The grammatical categories needed for..... naturally occurring text can be very different from those required for the invented data described in abstract

VI. Find the English equivalents from the text to the following words and word-combinations.

1	випадково	
2	значущі події	
3	компроміс між бажаним і можливим	
4	сміслові відмінності	
5	кількісний за своїм характером	
6	церковні проповіді	
7	критерії проектування	
8	мінімум упереджень	
9	висока частотність лем	
10	містити регіональні різновиди англійської мови	
11	складатись з необроблених текстів	
12	спеціально розроблений для вивчення мови	
13	об'єднати великі загальні корпуси	
14	необґрунтовані припущення щодо відповідних граматичних категорій	
15	лежати в основі корпусу розмовної британської англійської мови	

VII. Find appropriate synonyms (if any) and antonyms (if any) to the words below.

№	Word	Synonym	Antonym
1	assumption (n.)		
2	pejorative (adj.)		
3	frequent (adj.)		
4	recur (v.)		
5	effort (n.)		
6	reveal (v.)		
7	authentic (adj.)		
8	represent (v.)		
9	influence (n.)		
10	biased(adj.)		

VIII. Provide all the possible derivatives to the word forms below. Make up sentences with the words derived.

infrequent, comparative, produce, contain, educate, influential, generation, occur, combination, observer

IX. Correct the possible mistakes in terms of grammar, spelling and inappropriate use of vocabulary in the sentences below.

1) What is selected for observation depend from convenience, interests and hypotheses, but corpus data are a part of natural language use, and not produced for purposes of linguistic analysis.

2) The aim is not to study idiosyncratic details of performance who are, by chance, recorded in a corpora.

3) The LOB corpus contain British data from 1961; Frown and FLOB contain American and British data from 1991; and ICE contain regional varieties of English, such as India and Australia.

4) Many other corpuses for English, and increasingly for others languages, is available.

5) A corpora which claims to be a balancing sample of language use must to include texts which are spoken and written, casual and formal, fiktion and non-fiktion, which vary in level, age of, and sex and geografical origin of author.

6) How large the corpus is measured as word-types, or as the number of diferent texts or text-types it contain?

7) Any corpora is a compromise between the desirable and the feacible, and although design criterias cannot be operationalized, large balanced corpora reveals major regularities in language use.

8) Huge text collections can used to study patterns who do not occur even in large reference corporas.

9) Words and phrases in the world-wide-web can be searched for directly with seerch engines, or with a concordancer who uses this engines.

10) Tagging may making unwarranted assumptions about an appropriate grammatical categories.

X. Translate the following sentences into English, using the topical vocabulary under consideration.

1) Подібні принципи створення лежать в основі корпусу розмовної британської англійської мови, який містить близько півмільйона слів, поділених на зразки за вживанням дорослих, освічених, професійних людей, включаючи особисті та телефонні розмови, лекції та дискусії.

2) Немає підстав покладатися на жодний єдиний корпус, і часто доцільно об'єднати великі загальні корпуси.

3) Корпус спеціально створений для вивчення мови, але інші зібрання текстів також можуть бути корисними для деяких типів дослідження.

4) Оскільки велика кількість даних необхідна для вивчення того, що являється типовим та ймовірним, важливим критерієм є розмір.

5) Не зовсім очевидно, які доречні пропорції основних типів текстів, таких як якісні газети, літературна класика та повсякденні розмови, значно менших типів текстів, таких як оголошення в газеті, ділове листування та церковні проповіді.

UNIT 6

WORDS AND PHRASES. RECURRENT PHRASES, COLLOCATIONS AND PHRASAL SCHEMAS

<...> Many corpus studies reject individual words as units of meaning, and propose a theory of abstract phrasal units. Nevertheless, words are a good place to start, since, «a central fact about a word is how frequent it is» (Kilgarriff, 1997, p. 135), and other things being equal, the more frequent a word is, the more important it is to know it, and to teach it early to learners: hence the interest, since the 1890s, in reliable word-frequency lists for many applications.

Frequency shows that system and use are inseparable (Halliday, 1991). More frequent words tend to be shorter, irregular in morphology and spelling, and more ambiguous out of context: a glance at a dictionary shows that short frequent words require many column inches. A few, mainly grammatical, words are very frequent, but most words are very rare, and in an individual text or smallish corpus, around half the words typically occur only once each. In addition, a word with different senses usually has one meaning which is much more frequent. These relations imply a balance between economy of effort for the speaker and clarity for the hearer, and in the 1930s and 1940s Zipf (1945) tried to formulate statistical relations between word frequency, word length, and number of senses. (These regularities apply to many other aspects of human behavior. In a library, a few books are frequently borrowed, but most books collect dust.)

The simplest frequency lists contain unlemmatized word-forms from a general corpus, in alphabetical or frequency order, but there are considerable differences between even the top ten words from an unlemmatized written corpus (in 1), a spoken corpus (in 2), and a lemmatized mixed written and spoken corpus (in 3):

(1) *the, of, and, a, in, to [infinitive marker], is, to [preposition], was, it*

(2) *I, you, it, the, 's, and, n't, a, that, yeah*

(3) *the, BE, of, and, a, in, to [infinitive marker], HAVE, it*

These examples are from frequency lists for the 100-million-word BNC, made available by Kilgarriff (<ftp://ftp.itri.bton.ac.uk/bnc/>).

Unlemmatized lists show that different forms of a lemma differ greatly in frequency, and may have very different collocational behavior: see above on *endure* and *enduring*. However, raw frequency lists cannot distinguish words in different grammatical classes (e.g., *firm* as adjective or noun) and the different meanings of a word (e.g., *cold* as 'low temperature' versus 'lacking in feeling'). This requires a grammatically tagged corpus and a method of automatic sense disambiguation, and makes an apparently trivial counting task into a considerable theoretical problem.

Frequency lists require careful interpretation to provide what is really wanted, which is a measure of the relative importance of words, and more important than raw frequency may be even distribution across many text-types. Conversely, we want to know not only what is frequent in general, but what distinguishes a text-type. For example, words may be frequent in academic texts but unlikely in fiction, or vice-versa:

constants, measured, thermal, theoretically sofa, kissed, damned, impatiently.

These examples are from Johansson (1981; discussed also by Kennedy, 1998, p. 106). For important reference data on word frequency and distribution, see W. Francis and Kucera (1982), Johansson and Hofland (1988–9), and Leech, Rayson, and Wilson (2001; and <http://www.comp.lancs.ac.uk/ucrel/bncfreq/flists.html>).

We come back to the distinction between evidence and interpretation. Frequency and distribution (which are all we have) are indirect objective measures of the subjective concept of salience (which is what we really want). The objective measures have limitations, but allow analysis to be based on public and replicable data. The only alternative is intuition, which may be absent, speculative, or wrong. A very useful applied frequency study is reported by Coxhead (2000), who used a corpus of 3.5 million words to set up the Academic Word List (AWL).

This contains words which have both high frequency and wide distribution in academic texts, irrespective of subject area (but excluding approximately the 2,000 most frequent words in English, from West, 1953). AWL comprises 570 word families: not just word-forms, but head-words plus their inflected and derived forms, and therefore around 3,100 word-forms altogether, e.g.:

concept: conception, concepts, conceptual, conceptualization, conceptualize, conceptualized, conceptualizes, conceptualizing, conceptually.

Coxhead's corpus comprised texts from academic journals and university textbooks from arts, commerce, law, and natural science. To be included in AWL, a word had to occur at least 100 times altogether in the whole academic corpus, at least ten times in each of the four sub-corpora, and in at least half of 28 more finely defined subject areas, such as biology, economics, history, and linguistics.

AWL gives very good coverage of academic texts, irrespective of subject area. Here it must be remembered that words are *very* uneven in their frequency. In a typical academic text, the single word *the* covers around 6 or 7 per cent of running text, the top ten words cover over 20 per cent, and the 2,000 most frequent words cover around 75 per cent. The words in AWL typically cover a further 10 per cent. The remaining 15 per cent will be specialized words which are specific to a given topic, plus proper names, etc. AWL is further divided into ten sub-groups, from most to least frequent. Group 1 covers 3.6 per cent of the corpus, which means that a student reading academic prose could expect to come across *each word* in group 1, on average, once every four pages or so.

A list is, of course, just a list, not teaching materials, and requires interpretation by materials designers and teachers. However, even as a bare list, AWL can provide a check, for teachers or students themselves, on what words students should know. Word frequency lists are limited, especially for very common words, since these are common, not in their own right, but because they occur in common phrases. For example, *back* is usually in the top 100 in lemmatized frequency lists, and (including compounds such as *backward* and *backwater*) gets nearly five full pages in the COBUILD (1995a) dictionary. This is not because speakers frequently use *back* to mean a part of the body, but because it occurs in many phrases with only residual relations to this denotation. It has many meanings, but vanishingly few uses with the part-of-body meaning. The following examples are from Cobuild (1995a), and Sinclair (1991, p. 116) gives a detailed analysis of its nominal, prepositional and idiomatic uses.

lying on his back; the back of the chair; on the back of a postcard; at the back of the house; round the back; do something behind her back; get off my back; you scratch my back . . . ; see the back of someone; turn your back on

In summary: Frequent words are frequent because they occur in frequent phrases. In these phrases, frequent words are often delexicalized, because meaning is dispersed across the whole phrase. Since frequent content words are rarely used with their full lexical meaning, the boundary between content and function words is fuzzy. It is for these reasons that the co-occurrence of words and grammatical constructions has been studied so intensively: the central principle is that it is not words, but phrase-like units, which are the basic units of meaning.

The simplest definition of a phrase is a string of two or more uninterrupted word-forms which occur more than once in a text or corpus: see Altenberg (1998) on «recurrent word-combinations» and Biber et al. (1999) on «lexical bundles.» I used a program to identify strings in this sense, in a written corpus of four million words. (Since 2002, when I did this work with a locally written program, excellent n-gram software has been made available by William Fletcher at <http://kwicfinder.com/kfNgram/>.) The most frequent five-word string, over twice as frequent as any other, was *at the end of the*. And almost 30 out of the top 100 five-word strings had the pattern *PREP + the + NOUN + of + the*.

Examples included:

at the end of the; in the middle of the; at the beginning of the; at the bottom of the

The program operationalizes, in a very simple way, the concept of repeated units. It cannot automatically identify linguistic units, but presents data in a way which helps the analyst to see patterns. These findings are not an artifact of my small corpus. I looked at the same strings in the 100-million-word BNC, and found that, normalized to estimated occurrences per million words, the frequencies in the two corpora were remarkably similar. These examples represent only one pattern, of course. Other frequent five-word strings have discourse functions:

as a matter of fact; it seems to me that; it may well be that; but on the other hand

Altenberg (1998) identifies other recurrent multi-word strings, and some of their typical pragmatic functions. These multi-word strings are already evidence that recurrent lexicogrammatical units are not fixed phrases, but abstract semantic units. For example, the program above counts separately the strings *on the top of the*, *on the very top of the*, or *on top of the*, although, to the human analyst, they are semantically related.

A few dozen concordance lines can be manually inspected for patterns, but if we have thousands of lines, then we require a method of summarizing concordances and showing patterns. We can write a program which finds the most frequent collocates of a node, one, two, and three words to the left and right, and lists them in descending frequency.

The positional frequency table for *undergo* shows that it often occurs in a passive construction (*was forced to, is required to*), is often followed by an adjective signaling the seriousness of the event (*extensive, major*), and is often used of medical events (*surgery, operation*).

Raw frequency of co-occurrence is important, but we need to check the frequency of collocation relative to the frequency of the individual words. If two words

are themselves very frequent, they may co-occur frequently just by chance. Conversely, a word might be infrequent, but when it does occur, it usually occurs with a small set of words. For example, the word *vegetative* is not frequent, but when it occurs, especially in journalism, it often co-occurs with *persistent*, in the phrase *persistent vegetative state*, with reference to patients in a coma.

The variability of phrasal units makes it doubtful whether there could be a useful «phrase frequency list,» but corpus studies show that all words occur in habitual patterns which are often much stronger than is evident to intuition. For example, in a 200-million-word corpus, the word-form *persistent* occurred over 2,300 times, with clear semantic preferences, shown by the top 20 collocates, ordered by frequency:

persistent <*offenders, reports, most, rumours, state, vegetative, despite, young, juvenile, problem, injury, problems, rain, allegations, critic, offender, rumors, speculation, amid, cough*>

The most frequent single collocate (in 5 percent of cases) was *offenders*; and the most frequent set of collocates were words for *reports, rumors, and speculations*.

Table 4.1 Positional frequency table for NODE *undergo* in a span of 3 words to left and right (only collocates occurring five or more times are shown, in descending frequency, independently for each position).

N-3	N-2	N-1	NODE	N+1	N+2	N+3
was	Forced	to	*	a	medical	And
is	Required	will	*	an	surgery	Tests
be	Have	and	*	further	testing	examination
are	Had	would	*	extensive	tests	Of
and	Is	must	*	the	treatment	surgery
that	They	he'll	*	major	Change	operation
been	About	should	*	surgery	changes	transformation
were	And	who	*	treatment	For	Before
where	Patients	women	*	medical	Heart	Test
children	That	often	*	heart	And	medical
he	He		*	his	Major	For
in	Will		*	testing	operation	In
the	Women		*		examination	On
women	Due		*		extensive	training
will	Ordered		*		transformation	To
for			*		Radical	testing
last			*		Test	The
not			*		training	A
of			*		The	As
						By
						changes

Persistent is used of bad situations (collocates include *problem* and *problems*), which include medical conditions (*cough, injury, vegetative*) and criminal activities

(*juvenile, offenders*). Some collocates frequently occur in longer phrases (*persistent juvenile offenders, persistent vegetative state*), and most examples involving «crime» and «allegations» are from journalism. With comparable data on a broad sample of words, we can then ask whether *persistent* exerts a stronger than average collocational attraction on its surrounding collocates. The brief answer is that *persistent* is typical of many words in this respect.

The top collocates of a word provide evidence of its characteristic semantic preferences and syntactic frames. Figures for a broad sample of words show how pervasive collocational attraction is, and allow generalizations about its strength and variability. The example of *persistent* is taken from a data-base (COBUILD, 1995b), which provides a suitable sample of node-words and their collocates for quantitative statements about phraseology. For the 10,000 most frequent content words (word-forms) in the 200-million-word corpus, the database gives the 20 most frequent collocates in a span of four words to left and right. For each node-collocate pair, it gives 20 randomly selected concordance lines, each with a rough description of its source (e.g., British fiction, American journalism). For individual words, this provides figures on the strength of attraction between node and top collocate:

undergoing <*surgery* 11 %>, *undergo* <*surgery* 9 %>, *endured* <*years* 6 %>, *persistent* <*offenders* 5 %>

(That is, in 11 percent of occurrences, *undergoing* co-occurs with *surgery*, etc.)

The data-base shows that around 75 percent of content words in the central vocabulary of English have a strength of attraction of between 2 and 9 percent. And over 20 percent co-occur with one specific collocate in over 10 percent of occurrences. Conversely, few words have less than one chance in 50 of co-occurring with one specific collocate.

These are figures for the attraction between two single unlemmatized wordforms. Collocational attraction is much stronger if it is calculated between a node and a set of approximate synonyms. For example:

achieving <*goal(s)* 7 %, *success, aim, results, objectives*> 15 %
ambitious <*plan(s)* 7 %, *project, program(me), scheme*> 16 %

The strength of attraction between all common content words is surprisingly high, yet not taken into account in most language description. Corpus study shows kinds of linguistic organization which are not predictable by rule, but are recurrent and observable.

TASKS

I. Study the topical vocabulary below. Find the contexts where the units under consideration are used in the text.

1	salience (n.)	['seɪljəns]	характерна особливість
2	figure (n.)	['fɪgə]	цифра, число
3	pervasive (adj.)	[pɜː'veɪsɪv]	проникаючий, всепроникний
4	rough (adj.)	[rʌf]	приблизний, нерівний, грубий

5	predictable (adj.)	[prɪ'dɪktəbl]	передбачуваний
6	descending (adj.)	[dɪ'sendɪŋ]	спадний
7	node (n.)	[nəʊd]	вузол (<i>іноді контекст. вершина</i>)
8	ascending (adj.)	[ə'sendɪŋ]	висхідний
9	remarkably (adv.)	[rɪ'mɑ:kəbli]	надзвичайно, дивовижно
10	respective (adj.)	[rɪs'pektɪv]	відповідний
11	bare (adj.)	[beə]	порожній, незаповнений
12	lexicalize (v.)	['leksɪkəlaɪz]	лексикалізувати
13	disperse (v.)	[dɪs'pɜ:s]	розпорошувати, розсіювати; розходитися
14	positional (adj.)	[pə'zɪʃənl]	позиційний
15	persistent (adj.)	[pə'sɪstənt]	наполегливий
16	residual (adj.)	[rɪ'zɪdʒʊəl]	залишковий
17	denotation (n.)	[,di:nəʊ'teɪʃən]	позначення, денотація
18	average (adj.)	['ævərɪdʒ]	середньостатистичний
19	span (n.)	[spæn]	інтервал, коротка відстань між елементами; (<i>іноді контекст. група, комплект</i>)
20	randomly (adv.)	['rændəmli]	випадковим чином
21	allegation (n.)	[,æli'geɪʃ(ə)n]	твердження, заява (<i>іноді контекст. голослівне твердження, не підтвержене доказами</i>)
22	approximate (adj.)	[ə'prɒksɪmɪt]	приблизний
23	word-form (n.)	[wɜ:d-fɔ:m]	словоформа
24	set (n.)	[set]	набір, комплекс
25	speculation (n.)	[,spekjʊ'leɪʃən]	міркування
26	conversely (adv.)	['kɒnvɜ:sli]	навпаки, протилежно, всупереч
27	boundary (n.)	['baʊndəri]	межа, розподільна лінія, кордон
28	distribution (n.)	[,dɪstrɪ'bju:ʃən]	розподіл, розповсюдження, дистрибуція
29	replicable (adj.)	['replɪkəbl]	повторюваний, відтворюваний
30	disambiguation (n.)	[,dɪsæm'bigjuetʃən]	усунення неоднозначності

II. Answer the following questions to the text.

- 1) What is the central fact about a word?
- 2) What proves system and use to be inseparable?
- 3) What kind of evidence do the top collocates of the word provide?
- 4) Can a useful «phrase frequency list» be elaborated? Give your reasons.
- 5) When can concordance lines be inspected for patterns manually? When is a method of summarizing concordances required?
- 6) What does an abbreviation of AWL stand for? What does AWL provide?
- 7) Why are word frequency lists limited?

- 8) Provide examples of linguistic functioning in their nominal, prepositional and idiomatic uses.
- 9) What does word frequency mean? When do we observe words to be delexicalized?
- 10) Can the program discussed in the text under consideration identify linguistic units automatically? In what way is the linguistic data presented?

III. Match the words to their definitions.

1	variability	a	usually or often done by someone;
2	pervasive	b	something observed in a scientific investigation or experiment that is not naturally present but occurs as a result of the experiment itself;
3	manually	c	the fact that something changes often or is not always the same;
4	analyst	d	written or spoken language, especially when it is studied in order to understand how people use language;
5	statement	e	relating to the meaning of words;
6	artifact	f	something that you say or write that states a fact or gives information in a formal way; a written or spoken announcement on an important subject that someone makes in public;
7	habitual	g	words that are often used together;
8	semantic	h	operated by a person instead of automatically or using a computer;
9	discourse	i	spreading through the whole of something and becoming a very obvious feature of it;
10	collocate	g	someone whose job is to carefully examine a situation, event etc in order to provide other people with information about it;

IV. Match the parts below to complete a single syntactic unit from the text.

1	The top collocates of a word provide evidence of its	a	but we need to check the frequency of collocation relative to the frequency of the individual words.
2	Raw frequency of co-occurrence is important,	b	we can then ask whether <i>persistent</i> exerts a stronger `than average collocational attraction on its surrounding collocates.
3	For each node-collocate pair, it gives 20 randomly selected concordance lines,	c	but are recurrent and observable.
4	With comparable data on a broad sample of words,	d	and allow generalizations about its strength and variability.
5	Corpus study shows kinds of linguistic organization which are not predictable by rule,	e	characteristic semantic preferences and syntactic frames.

6	Figures for a broad sample of words show how pervasive collocational attraction is,	f	each with a rough description of its source (e.g., British fiction, American journalism).
7	It cannot automatically identify linguistic units,	g	since these are common, not in their own right, but because they occur in common phrases.
8	These multi-word strings are already evidence that	h	but if we have thousands of lines, then we require a method of summarizing concordances and showing patterns.
9	Word frequency lists are limited, especially for very common words,	i	recurrent lexicogrammatical units are not fixed phrases, but abstract semantic units.
10	A few dozen concordance lines can be manually inspected for patterns,	j	but presents data in a way which helps the analyst to see patterns.

V. Fill in the gaps using the appropriate words from the box. Underline the words inserted:

disambiguation construction meanings morphology collocates distribution pervasive lexical dictionary denotation
--

1) Since frequent content words are rarely used with their full meaning, the boundary between content and function words is fuzzy.

2) Figures for a broad sample of words show how collocational attraction is, and allow generalizations about its strength and variability.

3) Some frequently occur in longer phrases (*persistent juvenile offenders*, *persistent vegetative state*), and most examples involving «crime» and «allegations» are from journalism.

4) This requires a grammatically tagged corpus and a method of automatic sense, and makes an apparently trivial counting task into a considerable theoretical problem.

5) More frequent words tend to be shorter, irregular in and spelling, and more ambiguous out of context: a glance at a dictionary shows that short frequent words require many column inches.

6) The positional frequency table for *undergo* shows that it often occurs in a passive (*was forced to*, *is required to*), is often followed by an adjective signaling the seriousness of the event (*extensive*, *major*), and is often used of medical events (*surgery*, *operation*).

7) This is not because speakers frequently use *back* to mean a part of the body, but because it occurs in many phrases with only residual relations to this..... It has many....., but vanishingly few uses with the part-of-body meaning.

8) More frequent words tend to be shorter, irregular in morphology and spelling, and more ambiguous out of context: a glance at ashows that short frequent words require many column inches.

9) A few, mainly grammatical, words are very frequent, but most words are very rare, and in an individual text or smallish....., around half the words typically occur only once each.

10) Frequency and(which are all we have) are indirect objective measures of the subjective concept of salience (which is what we really want).

VI. Find the English equivalents from the text to the following words and word-combinations:

1	підсумувати кількість співпадінь	
2	важлива теоретична проблема	
3	перелік високочастотних слів	
4	узагальнення щодо потужності та варіативності	
5	часто повторювані лексико-граматичні одиниці	
6	досліджувати в ручному режимі	
7	розмежувальна лінія між самостійними та допоміжними частинами мовами	
8	нерегулярний характер у площині морфології та орфографії	
9	вибрані випадковим чином рядки узгодження (відповідності)	
10	автоматично визначати (ідентифікувати) мовні одиниці	
11	низька частотність сполучуваності одиниць	
12	організація мовних одиниць, не передбачена конкретним правилом	
13	порівняльні дані	
14	варіативність фразових одиниць	
15	приблизне (нечітке, неповне) описання джерела одиниць дослідження	

VII. Find appropriate synonyms (if any) and antonyms (if any) to the words below:

№	Word	Synonym	Antonym
1	rough (adj.)		
2	abstract (adj.)		
3	salience (n.)		
4	objective (adj.)		
5	summarize (v.)		
6	comparable (adj.)		
7	respective (adj.)		

8	bare (adj.)		
9	lexicalize (v.)		
10	disperse (v.)		

VIII. Provide all the possible derivatives to the word forms below. Make up sentences with the words derived.

mean, regular, ambiguous, subject, object, vanish, part, allow, very, attract

IX. Correct the possible mistakes in terms of grammar, spelling and inappropriate use of vocabulary in the sentences below.

1) The strengths of attraction between all common content words is surprising highly, yet not taken into account in most language description.

2) If two words themselves very frequently, they may co-occur frequently just by chance.

3) Coxhead's corpus comprise texts from academic journals and university textbooks in arts, commerce, law, and natural science.

4) I looked to the same strings in the 100-million-word BNC, and found that, normalized to estimated occurrences per million words, the frequencies in the two corpuses were remarkably similarly.

5) This is not because speakers frequently uses *back* to mean part of the body, but because it occur in many phrases with only residual relations to this denotation.

6) Conversely, few words have lesser than one chance in 50 of co-occurring with one specific collocates.

7) The brief answer is that *persistent* is typical for many words in this disrespect.

8) It is for this reasons that the co-occurrence of words and grammatical constructions has being studied so intensively: the central principle is that it is not words, but phrase-like units, which are the basic units of meaning.

9) We can write program which finds the most frequently collocates of a node, one, two, and three words on the left and right, and lists them in descending frequency.

10) More frequently words are tending to be shorter, irregular in morphology and spelling, and more ambiguously out of context: a glance at a dictionary shows that short frequent words require many column inches.

X. Translate the following sentences into English, using the topical vocabulary under consideration.

1) Застосування комп'ютерів дозволяє миттєво обробити величезний обсяг мовного матеріалу і відібрати всі можливі у конкретному корпусі приклади вживання необхідних для аналізу одиниць.

2) У цьому сенсі важливою є увага до квантитативного компонента мови, тобто врахування в першу чергу більш частотних елементів порівняно з менш

частотними, визнання квантитативних відношень суттєвим фактором у мовній еволюції і структурі мовних правил.

3) Фахівці у галузі прикладної лінгвістики повинні звернути увагу на синхронічну варіативність мови, тобто визнати той факт, що не існує єдиної жорсткої системи засобів вираження змісту, а існують її різні реалізації, в тому числі залежні від психологічних, біологічних і соціальних факторів.

4) У традиційному підручнику англійської мови буде сказано, що конструкція *I'm not* можлива в літературному англійському, а от конструкція *I ain't* не існує. Корпусна лінгвістика додатково вивчає й імовірність (*probability*) лінгвістичних явищ. Тобто, з погляду корпусної лінгвістики, ми не можемо сказати, що вживання *I ain't* у літературній мові зовсім неможливо. Воно всього лише малоімовірне.

5) Процедура корпусного аналізу включає три кроки: 1) ідентифікація мовних даних за допомогою категоріального аналізу; 2) співвідношення мовних даних за допомогою статистичних методів; 3) інтелектуальна інтерпретація результатів.

UNIT 7 DISCOURSE ANALYSIS

Who does Discourse Analysis, and Why?

Discourse analysts do what people in their everyday experience of language do instinctively and largely unconsciously: notice patternings of language in use and the circumstances (participants, situations, purposes, outcomes) with which these are typically associated. The discourse analyst's particular contribution to this otherwise mundane activity is to do the noticing consciously, deliberately, systematically, and, as far as possible, objectively, and to produce accounts (descriptions, interpretations, explanations) of what their investigations have revealed.

Since the study of language *in use*, as a goal of education, a means of education, and an instrument of social control and social change, is the principal concern of applied linguistics, indeed its *raison d'être*, it is easy to see why discourse analysis has such a vital part to play in the work that applied linguistics does, and why so much of the work that has been done over the last few decades on developing the theory and practice of discourse analysis been done by applied linguists (Widdowson, Candlin, Swales, for example) or by linguists (notably Halliday and his followers) for whom the integration of theory and practice is a defining feature of the kind of linguistics that they do.

Much of the work, but not by any means all. A great deal of discourse analysis is done by linguists who would not call themselves applied and much by scholars in other disciplines – sociology, psychology, psychotherapy, for example – who would not call themselves linguists. Discourse analysis is part of applied linguistics but does not belong exclusively to it; it is a *multi-disciplinary* field, and hugely diverse in the range of its interests.

For many the interest in discourse is *beyond* language in use (Jaworski & Coupland, 1999, p. 3) to «language use relative to social, political and cultural formations . . . , language reflecting social order but also language shaping social order, and shaping individuals' interaction with society.»

That this is no overstatement may quickly be demonstrated by indicating something of the range of discourse-related books published in recent years: discourse and politics (Schäffner & Kelly-Holmes, 1996; Howarth et al., 2000); ideologies (Schäffner, 1997), and national identity (Wodak et al., 1999); environmental discourse (Hajer, 1997; Harre, Brockmeier, & Muhlhausler, 1999); discourse and gender (Walsh, 2001; Wodak, 1997; Romaine, 1998); discourse of disability (Corker & French, 1999) and the construction of old age (Green, 1993); applied discursive psychology (Willig, 1999); professional discourse (Gunnarson, Linell, & Nordberg, 1997) and professional communication across cultural boundaries (Scollon, Scollon, & Yuling, 2001); the discourse of interrogation and confession (Shuy, 1998); academic discourse (Swales, 1998); discourse in cross-cultural communication (Hatim, 2000) and translation (Schäffner, 2002); discourse in everyday life (Locke, 1998; Cameron, 2000; Delin, 2000) and, at some remove from the everyday, divine discourse (Wolterstorff, 1995).

Jaworski and Coupland (1999, pp. 3–6) explain why so many areas of academic study have become so gripped by enthusiasm for discourse analysis in terms, firstly,

of a shift in epistemology, «a falling off of intellectual security in what we know and what it means to know . . . The question of *how* we build knowledge has come to the fore, and this is where issues to do with language and linguistic representation come into focus.» They point, secondly, to a broadening of perspective in linguistics, with a growth of linguistic interest in analysis of conversation, stories, and written text, in «the subtleties of implied meaning» and in the interaction of spoken language with nonlinguistic communication. And, thirdly, they note how, in the changed political, social and technological environment in which we now live – the postmodern world of service industry, advertising, and communications media – discourse «ceases to be ‘merely’ a function of work; it becomes work [and the] analysis of discourse becomes correspondingly more important.»

Defining Discourse

Discourse analysis may, broadly speaking, be defined as the study of language viewed communicatively and/or of communication viewed linguistically. Any more detailed spelling out of such a definition typically involves reference to concepts of language *in use*, *language above or beyond the sentence*, language as meaning *in interaction*, and language in *situational and cultural context*. Depending on their particular convictions and affiliations – functionalism, structuralism, social interactionism, etc. – linguists will tend to emphasize one, or some, rather than others in this list. (On the origins and implications of the language in use vs. language above the sentence distinction see for example Schiffrin, 1994, pp. 20–39; Pennycook, 1994a, p. 116; Widdowson, 1995, p. 160; Cameron, 2001, pp. 10–13.)

To illustrate this point, let us imagine four linguists preparing to work with the following small sample:

A: You THREW it so you GET it

B: MOI↓ra + I'll call my MUM

Linguist 1 sees a *text* – the verbal record of a speech event, something visible, palpable and portable, consisting of various bits of linguistic meaning (words, clauses, prosodic features, etc.). This linguist is mainly interested in the way the parts of the text relate to each other to constitute a unit of meaning.

Linguist 2 sees beyond the text to the *event* of which it is the verbal record. Linguist 2 is most likely the person who collected the data; and who made the following note describing some features of the situation in which the exchange took place:

[sunny Sunday afternoon, Edinburgh Botanic Garden, two girls, both aged 7 or 8, on a path; one of them has kicked the ball they are playing with into the bushes]

This linguist is mainly interested in the relationships between the various factors in the event: the participants, their cultural backgrounds, their relationship to each other, the setting, what is going on, the various linguistic choices made, etc.

Linguist 3 sees the text and the event but then beyond both to the *performance* being enacted, the *drama* being played out between the two girls: what has happened, who is responsible, how the girls evaluate these facts (relate them to some existing framework of beliefs and attitudes about how the world – their world – works), how they respond to them, what each is trying to achieve, their strategies for attempting to

achieve these objectives, etc. This linguist is mainly interested in the dynamics of the process that makes the event happen.

Linguist 4 sees the text, the event, and the drama; but beyond these, and focally, the *framework of knowledge and power* which, if properly understood, will explain how it is possible for the two children, individually and jointly, to enact and interpret their drama in the way they do.

We may, not unreasonably, imagine that our four linguists are colleagues in the same university department. Each recognizes the validity of the perspective of each of the others, and the fact that, far from there being any necessary conflict or «incommensurability» between them (but cf. Pennycook, 1994a), the perspectives are complementary: all are needed for a full understanding of what discourse is and how it works.

As implied by the above, I do not think there is much to be gained from attempts to achieve a single definition of discourse that is both comprehensive and succinct. (For a list and discussion of such definitions, see for example Jaworski & Coupland 1999: 1–7.) Here instead is a set of definitions in the style of a dictionary entry for «discourse»:

discourse

1 the linguistic, cognitive and social processes whereby meanings are expressed and intentions interpreted in human interaction (linguist 3);

2 the historically and culturally embedded sets of conventions which constitute and regulate such processes (linguist 4);

3 a particular event in which such processes are instantiated (linguist 2);

4 the product of such an event, especially in the form of visible text, whether originally spoken and subsequently transcribed or originally written (linguist 1).

TASKS

I. Study the topical vocabulary below. Find the contexts where the units under consideration are used in the text.

1	discourse (n.)	[dɪs'kɔ:s]	дискурс; висловлення
2	unconsciously (adv.)	[ʌn'kɒŋʃəsli]	несвідомо
3	patterning (n.)	['pætənɪŋ]	структурування, компонування
4	circumstance (n.)	['sɜ:kəmstəns]	обставина; умова; подробиця, деталь
5	mundane (adj.)	['mʌndeɪn]	світський
6	deliberately (adv.)	[dɪ'libərətli]	обдуманно; зважено
7	raison d'être (french).	[,reizɔ̃ 'dɛtrə]	підстава; сенс існування
8	multi-disciplinary	[,mʌltɪ'dɪsɪplɪnəri]	багатогалузевий
9	boundary (n.)	['baʊndəri]	кордон; межа
10	divine (adj.)	[dɪ'vaɪn]	релігійний

11	gripped (adj.)	[grɪpt]	захоплений
12	epistemology (n.)	[ɪˌpɪstəˈmɒlədʒi]	теорія пізнання, епістемологія
13	subtlety (n.)	[ˈsʌtlti]	тонке розходження
14	implied (adj.)	[ɪmˈplaɪd]	передбачуваний, очікуваний
15	correspondingly (adv.)	[ˌkɒrɪsˈpɒndɪŋli]	відповідно
16	interaction (n.)	[ˌɪntərˈæksjən]	взаємодія; вплив один на одного
17	conviction (n.)	[kənˈvɪkʃən]	переконання, погляди
18	affiliation (n.)	[əˌfɪliˈeɪʃ(ə)n]	простежування джерел, походження; встановлення зв'язку (з чим-небудь)
19	palpable (adj.)	[ˈpælpəbl]	очевидний, явний
20	clause (n.)	[klaʊz]	речення (частина складного речення)
21	prosodic (adj.)	[prəˈsɒdɪk]	просодичний, інтонаційний
22	evaluate (v.)	[ɪˈvæljuːeɪt]	оцінювати, давати оцінку
23	framework (n.)	[ˈfreɪmwɜːk]	загальна схема, структура
24	focally (adv.)	[ˈfəʊkəli]	фокусно
25	validity (n.)	[vəˈlɪdɪti]	вагомість, обґрунтованість; вірність, правильність
26	Incommensurability (n.)	[ˌɪnkəˌmenʃərəˈbɪlɪti]	несумірність; непорівнянність
27	complementary (adj.)	[ˌkɒmplɪˈmentəri]	комплементарний, взаємодоповнюючий
28	embedded (adj.)	[ɪmˈbedɪd]	включений, вкладений; вбудований
29	instantiate (v.)	[ɪnˈstænsɪeɪt]	проілюструвати конкретним прикладом
30	subsequently(adv.)	[ˌsʌbsɪkwəntli]	згодом, потім, пізніше

II. Answer the following questions to the text.

- 1) What is the discourse analyst's particular contribution to the everyday experience of language?
- 2) Why have applied linguists done much work on developing the theory and practice of discourse analysis?
- 3) Who except applied linguists has done a great deal of discourse analysis?
- 4) Why is the interest in discourse beyond language in use?
- 5) What discourse-related books were published in recent years?
- 6) How may discourse analysis be defined?
- 7) What kind of person is Linguist2?
- 8) What is Linguist 3 mainly interested in?

- 9) What concepts of language are involved in the more detailed definition of discourse analysis?
 10) How does Linguist 4 see the text, the even, and the drama?

III. Match the words to their definitions.

1	deliberately	a	combining in such a way as to enhance or emphasize the qualities of each other or another
2	multi-disciplinary	b	not able to be judged by the same standards; having no common standard of measurement
3	divine	c	the aggregate of people living together in a more or less ordered community
4	complementary	d	devoted to God
5	incommensurable	e	a physical or mental condition that limits a person's movements, senses, or activities
6	framework	f	in a careful and unhurried way
7	Society	g	state or describe exactly the nature, scope, or meaning of
8	disability	h	firmly hold the attention or interest of
9	define	i	combining or involving several academic disciplines or professional specializations in an approach to a topic or problem
10	grip	j	a basic structure underlying a system, concept, or text

IV. Match the parts below to complete a single syntactic unit from the text.

1	Discourse analysts do what people in their everyday experience of language do instinctively and largely unconsciously:	a	discourse «ceases to be ‘merely’ a function of work; it becomes work [and the] analysis of discourse becomes correspondingly more important.»
2	A great deal of discourse analysis is done by linguists	b	whether originally spoken and subsequently transcribed or originally written.
3	They note how, in the changed political, social and technological environment in which we now live	c	notice patternings of language in use and the circumstances with which these are typically associated.
4	Discourse analysis may be defined as the study of language viewed communicatively	d	to the <i>performance</i> being enacted, the <i>drama</i> being played out between the two girls.
5	This linguist is mainly interested in the dynamics	e	who would not call themselves applied and much by scholars in other disciplines.

6	Discourse is the product of such an event, especially in the form of visible text,	f	to enact and interpret their drama in the way they do.
7	Linguist 2 sees beyond the text to the <i>event</i>	g	of discourse that is both comprehensive and succinct.
8	Linguist 3 sees the text and the event but then beyond both	h	and/or of communication viewed linguistically.
9	The <i>framework of knowledge and power</i> will explain how it is possible for the two children, individually and jointly,	i	of the process that makes the event happen.
10	There is much to be gained from attempts to achieve a single definition	j	of which it is the verbal record.

V. Fill in the gaps using the appropriate words from the box. Underline the words inserted.

gripped	various	mundane	divine	defining feature	epistemology
investigations	subtleties	features	prosodic features	instantiated	
perspective	speech event	overstatement	boundaries	palpable	
embedded	exchange	cultural	discourse analysis		

1) The discourse analyst's particular contribution to this otherwise activity is to do the noticing consciously, deliberately, systematically, and, as far as possible, objectively, and to produce accounts of what their have revealed.

2) Much of the work that has been done over the last few decades on developing the theory and practice of been done by applied linguists or by linguists for whom the integration of theory and practice is a of the kind of linguistics that they do.

3) That this is no may quickly be demonstrated by indicating something of the range of discourse-related books published in recent years: discourse and gender (Walsh, 2001; Wodak, 1997; Romaine, 1998); professional discourse (Gunnarson, Linell, & Nordberg, 1997) and professional communication across cultural (Scollon, Scollon, & Yuling, 2001); discourse (Wolterstorff, 1995).

4) Many areas of academic study have become so by enthusiasm for discourse analysis in terms, firstly, of a shift in

5) Discourse is a particular event in which social processes are

6) They point to a broadening of in linguistics, with a growth of linguistic interest in analysis of conversation, stories, and written text, in «the of implied meaning».

7) Linguist 2 is most likely the person who made the following note describing some of the situation in which the took place.

8) Linguist 1 sees a *text* – the verbal record of a, something visible, and portable, consisting of various bits of linguistic meaning (words, clauses,, etc.).

9) This linguist is mainly interested in the relationships between the various factors in the event: the participants, their backgrounds, their relationship to each other, the setting, what is going on, the linguistic choices made, etc.

10) Discourse is the historically and culturally sets of conventions which constitute and regulate social processes.

VI. Find the English equivalents from the text to the following words and word-combinations.

1	компонування	
2	допит	
3	багатогалузевий	
4	перебільшення	
5	обґрунтованість перспективи	
6	оцінювати факти	
7	визначальна особливість	
8	прикладна дискурсивна психологія	
9	мовленнєва дія	
10	академічний дискурс	
11	захоплений ентузіазмом	
12	розширення перспектив в мовознавстві	
13	складають одиницю значення	
14	просодичні особливості	
15	єдине визначення дискурсу	

VII. Find appropriate synonyms (if any) and antonyms (if any) to the words below.

№	Word	Synonym	Antonym
1	vital (adj.)		
2	mundane (adj.)		
3	goal (n.)		
4	respond (n.)		
5	various (adj.)		
6	boundary (n.)		
7	relate (v.)		
8	evaluate (v.)		
9	framework (n.)		
10	reveal (v.)		

VIII. Provide all the possible derivatives to the word forms below. Make up sentences with the words derived.

investigation	embedded	unconsciously	disability	diverse	interpret
understand	construction	environmental	systematically		

IX. Correct the possible mistakes in terms of grammar, spelling and inappropriate use of vocabulary in the sentences below.

1) Discourse is the linguistic, cognitive and social processes whereby meanings are expressing and intentions interpreting in humanistic interaction.

2) That these is no overstatement may be demonstrated by indicating something of the range of discourse-related books published in recent years: discourse and politic (Schäffner & Kelly-Holmes, 1996; Howarth et al., 2000); ideologies (Schäffner, 1997), and nation identity (Wodaketal., 1999); proffessional discourse (Gunnarson, Linell, & Nordberg, 1997).

3) Jaworski and Coupland are pointed to a broadening of perspectives in linguistics, with a growth of linguistic interest in analyses of conversation, stories, and writing text, in «the subtleties of implied meaning» and in the interaction of speaking language with nonlinguistic communication.

4) The question of *how* we built knowledge has come to the fore, and this is where issues to do with language and linguistics representation comes into pocus.

5) To ilustrate this point, let us to imagine four linguists preparing to work with the follow smoll sample.

6) Many of the work that has been done over the last few decades on developing a theory and practise of discourse analyses been done by linguists for whose the integration of theory and practise is a defining feature of the kinds of linguistics that they do.

7) Any more detailed spelling out of such a definition typically involve reference to concepts of language *in use*, language *above or below the sentence*, language as meaning *in interactional*, and language in *situation and culture context*.

8) A great deal of discourse analyses are done by scholas of others disciplines – sociology, psychology, psychotherapy, for example – which would not be called themselves linguists.

9) Linguist 2 is most likely the personal which collected the data; and which made the following note describing some features of the situation in which the exchange took place:

10) The question is what is happened, who esponsible and how the girls evaluating this facts.

X. Translate the following sentences into English, using the topical vocabulary under consideration.

1) Оскільки вивчення мови у вживанні є основною проблемою прикладної лінгвістики, то легко зрозуміти, чому аналіз дискурсу має таку визначальну роль у роботі, яку виконує прикладна лінгвістика.

2) Термін «дискурс» є багатозначним, але його численні визначення не суперечать, а доповнюють одне одного.

3) Дискурс це мовні, когнітивні та соціальні процеси, за допомогою яких виражаються значення та наміри, інтерпретовані людською взаємодією.

4) На процес створення дискурсу впливають чинники лінгвістичного та екстралінгвістичного характеру, які утворюють комбінації під час взаємодії, зокрема компонентів власне дискурсу, пов'язаних із партнерами по комунікації, з темою і предметом спілкування, мовленнєвою ситуацією.

5) Аналіз дискурсу є частиною прикладної лінгвістики, але не належить їй винятково; він є багатогалузевим, і дуже різноманітним в межах своїх інтересів.

UNIT 8

RULES AND PRINCIPLES, CONTEXTS AND CULTURES OF LANGUAGE IN USE. INTERACTION

<...> Under this heading are grouped approaches which seek to understand the means by which language users – presumably universally, though this is always open to empirical contradiction – make sense, in the light of various contextual factors, of others’ utterances and contrive to have their own understood more or less as they intend.

Included here is work in pragmatics (Levinson, 1983; Mey, 1993; Thomas, 1995; Yule, 1996; Grundy, 2000) on:

- speech act theory (Austin, 1962; Searle, 1969);
- context; deixis and reference; shared knowledge (presuppositions) and frameworks of interpretation (schemata);
- cooperativeness in interaction: the «cooperative principle» and its «maxims» (Grice, 1975) and procedures for determining relevance (Sperber & Wilson, 1995);
- indirectness, indeterminacy and implicature and how these derive from particular ways of performing speech acts and manipulating the «maxims»;
- politeness or tact (Leech, 1983; Brown & Levinson, 1987; Kasper, 1997).

Politeness theory deals with the concept of face, with acts which are potentially damaging to face, and with the linguistic stratagems used for limiting such damage, when it is unavoidable. It is informed not only by linguistic pragmatics but also by social psychology and linguistic anthropology.

Table 1. Ways and means of discourse analysis:

Rules and principles	<ul style="list-style-type: none"> • pragmatics (including speech act theory and politeness theory); • conversation analysis;
Contexts and cultures	<ul style="list-style-type: none"> • ethnography of communication; • interactional sociolinguistics;
Functions and structures	<ul style="list-style-type: none"> • systemic-functional linguistics (SFL); • Birmingham school discourse analysis; • text-linguistics;
Power and politics	<ul style="list-style-type: none"> • pragmatic and sociolinguistic approaches to • power in language; • critical discourse analysis;

Work in conversation analysis (CA) (see Chapter 10, this volume), notably on rules of turn-taking and topic-management, and the sequencing rules governing relations between acts, is also included here. Note that the «rules» that CA is interested in are understood as members’ (not analysts’) rules: norms of behaviour, discoverable in the recurring patterns of the action itself, to which members orient in order to manage

and make sense of what is going on. In this respect CA differs from pragmatics. It also differs in its insistent empirical concern with the minutiae of the textual data.

Here are grouped approaches which focus on the sensitivity of ways of speaking (and writing) to situational and cultural differences. Ethnography of communication (Gumperz & Hymes, 1986; Duranti, 1997, Saville-Troike, 2003):

- offers a framework for the study of speech events, seeking to describe the ways of speaking associated with particular speech communities and to understand the role of language in the making of societies and cultures;

- involves both insider-like («emic») understanding of culturally specific ways of communicating (both verbal and non-verbal) and of the various beliefs and attitudes which connect with these ways; and outsider objectivity, encapsulated in Hymes' well-known «SPEAKING» acronym – an «etic» framework of speech event components: setting and scene, participants, ends (purposes, outcomes), act sequences, key (attitudinal aspects), instrumentalities (norms and styles of speech), norms of interaction and interpretation, and genre (the discourse type).

The knowledge that members of communities have of ways of speaking includes knowing when, where and how to speak, what to speak about, with whom, and so forth. The idea that we need, in addition to a theory of grammatical competence, a theory of *communicative* competence (Hymes, 1972) arises from this fact. Speakers need knowledge not only of what is grammatically possible but also of what is appropriate and typically done.

Interactional sociolinguistics (Schiffrin, 1994; Gumperz, 2001) aims at «replicable analysis that accounts for our ability to interpret what participants intend to convey in everyday communicative practice» (Gumperz, 2001). It pays particular attention to culturally specific contextual presuppositions, to the signals – «contextualisation cues» such as code- and style-switching, and prosodic and lexical choices – which signal these, and to the potential for misunderstanding which exists in culturally complex situations. It shares with CA a keen attention to detail and a focus on members' procedures, but differs from it in its interest in processes of inferencing and in the consequences of contextual variation and cultural diversity (for example, Tannen, 1984a).

Grouped here are text-friendly models of language and grammar-friendly approaches to text. Systemic-functional linguistics (SFL) (Halliday, 1978; Halliday & Hasan, 1985; Martin, 1992):

- sees language not as an autonomous system but as part of the wider socio-cultural context, as «social semiotic»; the aim is «to look into language from the outside and specifically, to interpret linguistic processes from the standpoint of the social order» (Halliday, 1978, p. 3);

- sees grammar as meaning potential – a «potential» that is functionally determined by the need of speakers and writers to simultaneously represent experience (the ideational function), manage their relationship with their co-participants (the interpersonal function) and produce dialogue or monologue, whether spoken or written, which is cohesive and coherent (the textual function); the realization of these meta-functions can be discerned both at the micro-level of clause structure (e.g., systems of transitivity) and at the macro-level of context (register features of «field,» «tenor,» and «mode»);

- provides a comprehensive theory of text analysis and genre (Martin, 2002).

Sharing much of the theoretical basis of SFL, Birmingham school discourse analysis originated in the analysis of classroom discourse (Sinclair & Coulthard, 1975). This revealed a hierarchical model of discourse structure (lesson, transaction, exchange, move, act), whose most widely exploited insight has been the regular sequence of moves within a teaching exchange: *Initiating move* (from the teacher), *Responding move* (from the pupil), *Feedback move* (from the teacher). This «IRF» pattern can be detected in other domains, including not only other unequal-power institutional domains such as doctor–patient consultations but also casual conversation (Stubbs, 1983; Tsui, 1994; Eggins & Slade, 1997, pp. 45–7). In the latter case, the third move (renamed follow-up) is likely to involve some kind of interpersonally motivated evaluation, for example a positive gloss on a respondent’s declining the initiator’s invitation.

Text-linguistics (de Beaugrande & Dressler, 1981; Levinson, 1983, p. 288 for the distinction between this and «speech act (or interactional)» approaches;) is not so much a single approach to discourse as a somewhat indeterminate set of interests or predispositions. These include:

- focus on *text*, generally defined as language «above,» «beyond» or «longer than» the sentence, and especially on the structure of texts and on their formal (syntactic and lexical), or surface, features;
- achievement – and the role of various kinds of lexis in signalling these (Hoey, 1991); on cohesion generally (e.g., Halliday & Hasan, 1976); on rhetorical patterns of textual meaning such as general-particular and problem-solution (Hoey, 1983, 2001); and on text structure seen in terms of hierarchies of textual relationships (Mann & Thompson, 1987);
- a particular concern with the analysis of *written* texts (see, for example, Connor & Johns, 1990; Mann & Thompson, 1992).

It is with the concept of interaction that discourse (for the analyst) comes to life. Entrances are made, intentions are formed, topics are introduced, turns are taken, actions are performed, reactions are prompted and in turn reacted to; understandings are checked, contributions are acknowledged, breakdowns occur, repairs are contrived; exits are negotiated. *People* are at work, doing things with meanings (producing them, interpreting them, negotiating them), co-creating an event whose trajectory may be clear to none of them until it is complete, and perhaps not even then.

This is discourse seen not as product (a text on a page) but as process, joint action in the making (Clark, 1996), and in consequence most difficult to capture and analyze without losing sight of its essence. The very smallest details – the falling-from-high pitch tone on which B says «Moira» for example – may be the most telling in revealing what is happening and with what intended, or unintended, effect.

The concept of discourse as interaction is present in all current ways and means of doing discourse analysis. In pragmatics, meaning is seen as «a dynamic process, involving the negotiation of meaning between speaker and hearer, the context of utterance (physical, social, and linguistic) and the meaning potential of an utterance» (Thomas, 1995, p. 22). The interactional workings of intention and effect are central

to speech act theory; Grice's maxims «are essentially ground rules for the interactive management of intentions» (Widdowson, 1998, p. 13); and the mutual establishment and maintenance of rapport (the avoidance of threats to face) underpins theories of politeness and tact.

Conversation analysis and interactional sociolinguistics provide somewhat contrasting approaches to the description of the accomplishment of interaction, the former more focused on the internal (to the text) mechanisms of turn-taking and sequencing, the latter highlighting the links between the micro-processes of the text, for example intonational and other «contextualization cues,» and the macro-world of social structures and cultural presuppositions. IRF analysis provides a somewhat static post hoc view of the *accomplished* interaction as a hierarchical patterning of acts, moves, exchanges, and transactions.

The interactionality of discourse is not restricted to the spoken language. «Text is a form of exchange; and the fundamental form of a text is that of dialogue, of interaction between speakers . . . In the last resort, every kind of text in every language is meaningful because it can be related to interaction among speakers, and ultimately to ordinary everyday spontaneous conversation» (Halliday & Hasan, 1985, p. 11). It can be argued that written no less than spoken interaction involves dynamic processes of interaction between readers and writers. Hoey, for example (2001, p. 11) defines text as «the visible evidence of a reasonably self-contained purposeful interaction between one or more writers and one or more readers, in which the writer(s) control the interaction and most of (characteristically all) the language.» <...>

TASKS

I. Study the topical vocabulary below. Find the contexts where the units under consideration are used in the text.

1	hierarchy (n.)	['haɪərə:ki]	ієрархія
2	interactionality (n.)	[,ɪntərəkʃə'nælɪtɪ]	здатність до взаємодії
3	resort (n.)	[rɪ'zɔ:t]	інстанція; звернення; надія
4	ultimately (adv.)	['ʌltɪmɪtli]	в кінцевому рахунку, в кінцевій позиції; зрештою
5	spontaneous (adj.)	[spɒn'teɪniəs]	спонтанний, безпосередній; невимушений
6	decline (v.)	[dɪ'klaɪn]	зменшуватися, знижуватися, занепадати
7	self-contained (adj.)	[self-kən'teɪnd]	окремий, самостійний; недружній, нетовариський
8	sociolinguistics (n.)	[,səʊsiəʊlɪŋ'ɡwɪstɪks]	соціолінгвістика
9	accomplishment (n.)	[ə'kɒmplɪʃmənt]	виконання, досягнення
10	mutual (adj.)	['mju:tʃʊəl]	взаємний, обопільний
11	maintenance (n.)	['meɪntənəns]	підтримання, утримання; захист
12	rapport (n.)	[ræ'pɔ:]	взаєморозуміння, контакт

13	breakdown (n.)	['breɪk ,daʊn]	невдача (в комунікативному процесі); розподіл, розбиття (елементів цілого)
14	trajectory (n.)	['trædʒɪktəri]	траєкторія
15	negotiate (v.)	[nɪ 'gəʊʃieɪt]	вести перемовини
16	contribution (n.)	[,kɒntrɪ 'bjʊ:ʃən]	внесок; співробітництво
17	reveal (v.)	[rɪ 'vi:l]	відкривати, виявляти, з'ясувати
18	consequence (n.)	['kɒnsɪkwəns]	наслідок
19	capture (v.)	['kæptʃə]	захопити, привернути
20	establishment (n.)	[ɪs 'tæblɪʃmənt]	установлення, налагодження
21	underpin (v.)	[,ʌndə 'pɪn]	підкріпляти, бути базисом
22	latter (adj.)	['lætə]	останній
23	contrive (v.)	[kən 'traɪv]	вигадувати, вимислювати; винаходити
24	concern (n.)	[kən 'sɜ:n]	відношення; стурбованість, занепокоєння
25	cohesion (n.)	[kəʊ 'hi:ʒən]	єдність, узгодженість
26	coherence (n.)	[kəʊ 'hiərəns]	послідовність; злагодженість
27	simultaneous (adj.)	[,sɪməl 'teɪniəs]	одночасний, одномоментний
28	transitivity (n.)	[,trænzə 'tɪvəti]	перехідність (<i>грам. категорія</i>)
29	indeterminate (adj.)	[,ɪndɪ 'tɜ:mɪnɪt]	невизначений, невирішений; сумнівний
30	predisposition (n.)	[,prɪ: dɪspə 'zɪʃən]	схильність

II. Answer the following questions to the text.

- 1) Give the definition to the notion of *discourse analysis*.
- 2) What ways and means of discourse analysis do you know?
- 3) What do the abbreviations of CA, SFL, IRF stand for?
- 4) What concept does discourse come to life with for the analyst?
- 5) How is meaning seen in pragmatics?
- 6) What underpins theories of politeness and tact?
- 7) What does interactional sociolinguistics provide discourse analysts with?
- 8) Can the interactionality of discourse be restricted to the spoken language?
- 9) What is the fundamental form of a text?
- 10) What is the purpose of IRF analysis? How do you understand the notion of the *accomplished interaction*?

III. Match the words to their definitions.

1	genre	a	a way of considering or doing something;
2	interpretation	b	the situation when the parts of something fit together in a natural or reasonable way;
3	cohesion	c	the style of language, grammar, and words used for particular situations;
4	syntax	d	someone who responds to a question or offer;
5	interaction	e	an explanation or opinion of what something means; a particular way of performing a piece of music, a part in a play, etc.;
6	register	f	the state of being likely to behave in a particular way or to suffer from a particular disease
7	approach	g	an occasion when two or more people or things communicate with or react to each other;
8	predisposition	h	the grammatical arrangement of words in a sentence;
9	respondent	i	the state of sticking together, or (of people) being in close agreement and working well together;
10	coherence	g	a style, especially in the arts, that involves a particular set of characteristics;

IV. Match the parts below to complete a single syntactic unit from the text.

1	Politeness theory deals with the concept of face, with acts which are potentially damaging to face,	a	where and how to speak, what to speak about, with whom, and so forth.
2	This «IRF» pattern can be detected in other domains,	b	and prosodic and lexical choices – which signal these, and to the potential for misunderstanding which exists in culturally complex situations.
3	In the latter case, the third move (renamed follow-up) is likely to involve some kind	c	but differs from it in its interest in processes of inferencing and in the consequences of contextual variation and cultural diversity.
4	Interactional sociolinguistics aims at «replicable analysis that accounts for	d	has been the regular sequence of moves within a teaching exchange: <i>Initiating move</i> (from the teacher), <i>Responding move</i> (from the pupil), <i>Feedback move</i> (from the teacher).
5	The knowledge that members of communities have of ways of speaking includes knowing when,	e	as a somewhat indeterminate set of interests or predispositions.

6	Text-linguistics is not so much a single approach to discourse	f	and with the linguistic stratagems used for limiting such damage, when it is unavoidable.
7	It shares with CA a keen attention to detail and a focus on members' procedures,	g	including not only other unequal-power institutional domains such as doctor–patient consultations but also casual conversation
8	This revealed a hierarchical model of discourse structure (lesson, transaction, exchange, move, act), whose most widely exploited insight	h	but also of what is appropriate and typically done.
9	Speakers need knowledge not only of what is grammatically possible	i	our ability to interpret what participants intend to convey in everyday communicative practice».
10	It pays particular attention to culturally specific contextual presuppositions, to the signals – «contextualisation cues» such as code- and style-switching,	j	of interpersonally motivated evaluation, for example a positive gloss on a respondent's declining the initiator's invitation.

V. Fill in the gaps using the appropriate words from the box. Underline the words inserted.

context	separately	meaning	complimenting	dictionary
grammar	assumptions	conversation	interpret	analysis

1) Discourseis sometimes defined as the analysis of language 'beyond the sentence'.

2) Some discourse analysts consider the larger discoursein order to understand how it affects the meaning of the sentence.

3) Charles Fillmore points out that two sentences taken together as a single discourse can have meanings different from each one taken.....

4) This contrasts with types of analysis more typical of modern linguistics, which are chiefly concerned with the study of.....: the study of smaller bits of language, such as sounds (phonetics and phonology), parts of words (morphology), meaning (semantics), and the order of words in sentences (syntax).

5) 'Reframing' is a way to talk about going back and re-interpreting theof the first sentence.

6) When you read a newspaper, you need to know whether you are reading a news story, an editorial, or an advertisement in order to properlythe text you are reading.

7) Discourse analysts who studynote that speakers have systems for determining when one person's turn is over and the next person's turn begins.

8) When speakers have differentabout how turn exchanges are signaled, they may inadvertently interrupt or feel interrupted.

9) Realizing that these words can function as discourse markers is important to prevent the frustration that can be experienced if you expect every word to have itsmeaning every time it's used.

10) Studying speech acts such asallows discourse analysts to ask what counts as a compliment, who gives compliments to whom, and what other function they can serve.

VI. Find the English equivalents from the text to the following words and word-combinations.

1	ієрархічна модель структури дискурсу	
2	лінгвістика тексту	
3	контекстуальні передумови	
4	протилежні (суперечливі) підходи	
5	теорія ввічливості	
6	в останньому випадку	
7	риторичне питання	
8	конкретний набір характеристик	
9	чинники контекстуалізації	
10	оцінка, мотивована в міжособистісній площині	
11	здійснення комунікативної взаємодії	
12	регулярна послідовність рухів	
13	базова форма тексту	
14	найдрібніші деталі	
15	смісловий потенціал висловлення	

VII. Find appropriate synonyms (if any) and antonyms (if any) to the words below.

№	Word	Synonym	Antonym
1	rhetorical (adj.)		
2	coherence (n.)		
3	cohesion (n.)		
4	achievement (n.)		
5	predisposition (n.)		
6	natural (adj.)		
7	reasonable (adj.)		
8	agreement (n.)		
9	simultaneous (adj.)		
10	transitivity (n.)		

VIII. Provide all the possible derivatives to the word forms below. Make up sentences with the words derived.

interact, invite, context, decline, interpret, signal, distinct,
gloss, diverse, associate

IX. Correct the possible mistakes in terms of grammar, spelling and inappropriate use of vocabulary in the sentences below:

1) This include focus on *text*, generally defined as language «above,» «beyond» or «longer than» the sentence, and especial on the structure of texts and on their formal (syntactic and lexical), or surface, features.

2) It can be argue that written no less then spoken interaction involves dynamic processes of interaction between reader and writer.

3) It is informs not only by linguistic pragmatic but also by social psychology and linguistic anthropology.

4) In this respect CA differ from pragmatics. It also differ in its insistent empirical concern with the minutiae of the textual data.

5) IRF analyses provides a somewhat static post hoc view of the *accomplished* interaction as a hierarchic patterning of acts, moves, exchanges, and transactions.

6) The idea that we are needing, in addition to a theory of grammatical competence, a theory of *communicative* competence arise from these fact.

7) Systemic-functional linguistics (SFL) see language not as an autonomous system but as part of the wider socio-cultural context, as «social semiotic»; the aim is «to look into language by the outside and specifically, to interpret linguistic processes on the standpoint of the social order».

8) Ethnography of comunication offer a framework for the study of speech events, seeking to discribe the ways of speaking asociated with particular speech comunities and to understand the role of laguage in the making of soccieties and cultures.

9) The concept of discours as interaction is present in all current way and mean of doing discours analysis.

10) This discours seen not how product (a text on a page) but how process, joint action in the making, and in consequeunce most dificult to capture and analize without losing sight by its esence.

X. Translate the following sentences into English, using the topical vocabulary under consideration.

1) Дискурсивний аналіз тексту – це аналіз дискурсу як результату (текст) і процесу мовленнєво-мисленнєвої діяльності з метою виявлення екстралінгвістичних, семантичних, когнітивних і мовленнєвих аспектів його формування, з'ясування функціональної спрямованості і соціальної організації.

2) Дискурсивний аналіз характеризується подвійною спрямованістю: він охоплює аналіз мовного рівня тексту як продукту мовленнєвої діяльності і вод-

ночас досліджує текст у дискурсі, тобто в дієвому просторі з урахуванням зовнішніх і внутрішніх факторів розгортання текстової комунікації.

3) Традиційно дискурс мав значення упорядкованого письмового, але найчастіше мовного, повідомлення окремого суб'єкта. В останні десятиліття термін дістав широке поширення в гуманітаристиці і набув нових відтінків значення.

4) Протягом становлення дискурсного аналізу як спеціальної галузі досліджень, з'ясувалося, що значення дискурсу не обмежується письмовим і усним мовленням, але позначає, крім того, і позамовні семіотичні процеси.

5) Дискурс не є ізольованою текстовою або діалогічною структурою, тому що набагато більше значення в його рамках здобуває паралінгвістичний супровід мови, що виконує ряд функцій (ритмічну, референтну, семантичну, емоційно-оцінну та інші).

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Altenberg, B. (1998) On the phraseology of spoken English. In A. P. Cowie (ed.), *Phraseology: theory, analysis and applications*, (pp. 101–22). Oxford: Oxford University Press.
2. Atkins, B. T. (ed.) (1998) *Using dictionaries. Studies of dictionary use by language learners and translators*. Tübingen: Max Niemeyer.
3. Atwell, E., Demetriou, G., Hughes, J., Schiffrin, A., Souter, C., & Wilcock, S. (2000) A comparative evaluation of modern English corpus grammatical annotation schemes. *ICAME Journal*, 24, 7–24.
4. Austin, J. (1962) *How to do things with words*. Oxford: Clarendon Press.
5. Beaugrande, R. de (1999) Reconnecting real language with real texts: text linguistics and corpus linguistics. *International Journal of Corpus Linguistics*, 4(2), 243–59.
6. Bergenholtz, H. & Tarp, S. (eds.) (1995) *Manual of specialised lexicography. The preparation of specialised dictionaries*. Amsterdam/Philadelphia: John Benjamins.
7. Biber, D., Conrad, S., & Reppen, R. (1998) *Corpus linguistics*. Cambridge: Cambridge University Press.
8. Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999) *Longman grammar of spoken and written English*. London: Longman.
9. Brown, P. & Levinson, S. (1987) *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.
10. Cameron, D. (2001) *Working with spoken discourse*. London: Sage.
11. Carter, R. & McCarthy, M. (1999) The English *get*-passive in spoken discourse. *English Language and Linguistics*, 3(1), 41–58.
12. Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
13. Chomsky, N. (2000) *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
14. Clark, H. (1996) *Using language*. Cambridge: Cambridge University Press.
15. COBUILD (1995a) *Collins COBUILD English dictionary*. London: Harper Collins.
16. Davies, A. & Elder, K. (ed.) (2004) *The handbook of applied linguistics*. Oxford: Blackwell.
17. Duranti, A. (1997) *Linguistic anthropology*. Cambridge: Cambridge University Press.
18. Fillmore, C. J. (1992) Corpus linguistics or computer-aided armchair linguistics. In J. Svartvik (ed.), *Directions in corpus linguistics* (pp. 35–60). Berlin: Mouton.
19. Francis, W. N. & Kucera, H. (1982) *Frequency analysis of English usage: lexicon and grammar*. Boston: Houghton Mifflin.
20. Gove, P. (1966) Etymology in *Webster's Third New International Dictionary*. *Word*, 22, 7–82.
21. Gove, P. (1968) The international scientific vocabulary in *Webster's Third*. *Journal of English Linguistics*, 2, 1–11.
22. Grundy, P. (2000) *Doing pragmatics*. London: Edward Arnold.

23. Grice, H. (1975) Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and semantics 3: speech acts* (pp. 41–58). New York: Academic Press.
24. Gumperz, J. (2001) Interactional sociolinguistics: a personal perspective. In D. Schiffrin, D. Tannen & H. E. Hamilton (eds.), *The handbook of discourse analysis* (pp. 215–28). Oxford: Blackwell.
25. Gumperz, J. & Hymes, D. (eds.) (1986) *Directions in sociolinguistics: the ethnography of speaking*. New York: Basil Blackwell.
26. Hallan, N. (2001) Paths to prepositions? A corpus-based study of the acquisition of a lexico-grammatical category. In J. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure* (pp. 91–120). Amsterdam: Benjamins.
27. Halliday, M. A. K. & Hasan, R. (1985) *Language, context and text: aspects of language in a social-semiotic perspective*. Victoria, Australia: Deakin University Press.
28. Hartmann, R. R. K. (2001) *Teaching and researching lexicography*. Harlow: Pearson Education Ltd.
29. Hartmann, R. R. K. & James, G. (1998) *Dictionary of lexicography*. London/New York: Routledge.
30. Hausmann, F. J., Reichmann, O., Wiegand, H. E., & Zgusta, L. (eds.) (1989–91) *Wörterbücher Dictionaries Dictionnaires: International encyclopedia of lexicography* (3 vols). Berlin/New York: Walter de Gruyter.
31. Hunston, S. (2002) *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
32. Hüllen, W. (1999) *English dictionaries, 800–1700: the topical tradition*. Oxford: Clarendon Press.
33. Hymes, D. (1972) On communicative competence. In J. B. Pride & J. Holmes (eds.), *Sociolinguistics* (pp. 269–93). Harmondsworth: Penguin.
34. Jaworski, A. & Coupland, N. (eds.) (1999) *The discourse reader*. London: Routledge.
35. Johansson, S. & Hofland, K. (1988–9) *Frequency analysis of English vocabulary and grammar* (2 vols). Oxford: Oxford University Press.
36. Kasper, G. (1997) Linguistic etiquette. In F. Coulmas (ed.), *The handbook of sociolinguistics* (pp. 374–85). Oxford: Blackwell.
37. Kennedy, G. (1998) *An introduction to corpus linguistics*. London: Longman.
38. Kilgarriff, A. (1997) Putting frequencies in the dictionary. *International Journal of Lexicography*, 10(2), 135–55.
39. Landau, S. I. (2001) *Dictionaries. The art and craft of lexicography* (2nd edn.). Cambridge: Cambridge University Press.
40. Leech, G. (1983) *Principles of pragmatics*. London: Longman.
41. Leech, G., Rayson, P., & Wilson, A. (2001) *Word frequencies in written and spoken English*. London: Longman.
42. Levinson, S. (1983) *Pragmatics*. Cambridge: Cambridge University Press.
43. McArthur, T. (1986) *Worlds of reference: lexicography, learning and language from the clay tablet to the computer*. Cambridge: Cambridge University Press.
44. Mey, J. (1993) *Pragmatics: an introduction*. Oxford: Blackwell.
45. Morton, H. C. (1994) *The Story of Webster's Third: Philip Gove's controversial dictionary and its critics*. Cambridge: Cambridge University Press.

46. Murray, K. M. E. (1977) *Caught in the web of words: James Murray and the Oxford English Dictionary*. New Haven: Yale University Press.
47. Nesi, H. (2000) *The use and abuse of EFL dictionaries: how learners of English as a foreign language read and interpret dictionary entries*. Tübingen: Max Niemeyer.
48. Pennycook, A. (1994a) Incommensurable discourses? *Applied Linguistics*, 15(2), 115–38.
49. Sampson, G. (1995) *English for the computer*. Oxford: Clarendon.
50. Saville-Troike, M. (2003) *Ethnography of communication: an introduction* (3rd edn.). Oxford: Blackwell.
51. Schiffrin, D. (1994) *Approaches to discourse*. Oxford: Blackwell.
52. Searle, J. R. (1969) *Speech acts: an essay in the philosophy of language*. Cambridge: Cambridge University Press.
53. Sinclair, J. (1991) *Corpus concordance collocation*. Oxford: Oxford University Press.
54. Sledd, J. & Ebbitt, W. R. (1962) *Dictionaries and that dictionary: a casebook on the aims of lexicographers and the targets of reviewers*. Glenview: Scott, Foresman & Co.
55. Sperber, D. & Wilson, D. (1995) *Relevance: communication and cognition* (2nd edn.). Oxford: Blackwell.
56. Teubert, W. (1999) Korpuslinguistik und Lexikographie. *Deutsche Sprache*, 4/1999, 292–313.
57. Thomas, J. (1995) *Meaning in interaction*. London: Longman.
58. Tono, Y. (2001) *Research on dictionary use in the context of foreign language learning: focus on reading comprehension*. Tübingen: Max Niemeyer.
59. Widdowson, H. G. (1995) Discourse analysis: a critical view. *Language and Literature*, 4(3), 157–72.
60. Widdowson, H. G. (1998) The conditions of contextual meaning. In K. Malmkjær & J. Williams (eds.), *Context in language learning and language understanding* (pp. 1–23). Cambridge: Cambridge University Press.
61. Wiegand, H. E. (1998) *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. [Dictionary research: investigations into dictionary use and the theory, history, criticism and computerisation of lexicography]. Berlin/New York: Walter de Gruyter.
62. Wilson, E. O. (1998) *Consilience: the unity of knowledge*. London: Little, Brown & Co.
63. Wright, J. (1998) *Dictionaries*. Oxford: Oxford University Press.
64. Yule, G. (1996) *Pragmatics*. Oxford: Oxford University Press.

Електронні словники

1. Cambridge Dictionary – Cambridge University Press [Електронний ресурс]. – Режим доступу : <https://dictionary.cambridge.org/>
2. Macmillan Dictionary. Free English Dictionary and Thesaurus Online [Електронний ресурс]. – Режим доступу : <https://www.macmillandictionary.com/>
3. Merriam Webster On-Line Dictionary [Електронний ресурс]. – Режим доступу : <http://www.merriam-webster.com>.
4. Oxford English Dictionary [Електронний ресурс]. – Режим доступу : <http://www.oed.com/>

Навчальне видання

Калініченко Віра Ігорівна
Стрюк Наталя Василівна

**МЕТОДИЧНІ ВКАЗІВКИ
ДЛЯ САМОСТІЙНОЇ РОБОТИ З ДИСЦИПЛІНИ
«ПРАКТИЧНИЙ КУРС ІНОЗЕМНОЇ МОВИ (АНГЛІЙСЬКОЇ)»**

*для студентів 3 курсу спеціальності «Прикладна лінгвістика»
філологічного факультету*

Методичний посібник

Технічний редактор

Т. О. Важеніна

Підписано до друку 13.06.2018
Формат 60 x 84/16. Папір офсетний.
Друк – цифровий. Умовн. друк. арк. 4,8
Тираж 50 прим. Зам. 34

Донецький національний університет імені Василя Стуса
21021, м. Вінниця, 600-річчя, 21
Свідоцтво про внесення суб'єкта видавничої справи
до Державного реєстру
серія ДК № 5945 від 15.01.2018