В. І. Калініченко, Н. В. Стрюк

ПРАКТИЧНИЙ КУРС ІНОЗЕМНОЇ МОВИ

НАВЧАЛЬНИЙ ПОСІБНИК З АНГЛІЙСЬКОЇ МОВИ ДЛЯ СТУДЕНТІВ 3 КУРСУ ФІЛОЛОГІЧНОГО ФАКУЛЬТЕТУ СПЕЦІАЛЬНОСТІ 035 «ФІЛОЛОГІЯ», СПЕЦІАЛІЗАЦІЇ 035.10 «ПРИКЛАДНА ЛІНГВІСТИКА»

Рецензенти: *С. О. Шехавцова,* доктор педагогічних наук, професор, декан факультету іноземних мов ДЗ «Луганський національний університет імені Тараса Шевченка».

І. О. Сімкова, доктор педагогічних наук, доцент, завідувач кафедри англійської мови гуманітарного спрямування Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Ю. О. Соболь, кандидат філологічних наук, доцент, завідувач кафедри іноземних мов Запорізького національного технічного університету.

Т. В. Русановська, кандидат філологічних наук, доцент, доцент кафедри іноземних мов Центральноукраїнського національного технічного університету.

Автори: *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 англійської філології факультету іноземних мов Донецького національного університету імені Василя Стуса.

CONTENTS

ПЕРЕДМОВА

Навчальний посібник «Практичний курс іноземної мови» призначений для студентів 3 курсу філологічного факультету спеціальності 035 «Філологія» спеціалізації 035.10 «Прикладна лінгвістика». Високий рівень володіння англійською мовою є визначальною складовою ефективної навчальної та наукової діяльності прикладних лінгвістів. Мета посібника – формування комунікативної, лінгвістичної, соціокультурної та професійної компетенції студентів, що уможливить ефективне застосовування здобутих знань у професійній діяльності.

Посібник складається з чотирьох розділів і охоплює такі актуальні теми прикладної лінгвістики як: "Lexicography", "Dictionaries", "Language Corpora", "Discourse Analysis".

Кожен розділ містить "Reading Section", "Speaking Section". "Picture Description Worksheet", "Listening (Audiovisual Comprehension) Section", "Writing Section" та "Self-Reflection Sheet". Завдання для читання розроблені на матеріалі збірника наукових публікацій з проблем прикладної лінгвістики "The Handbook of Applied Linguistics" (2004).

Наприкінці посібника подано короткий словник термінів прикладної лінгвістики, "Concise Glossary of Applied Linguistics Terms", який може бути використаним як для аудиторної, так і для самостійної роботи студентів, та завдання для самоконтролю з ключами, "Self-Assessment Final Test".

Посібник містить наукові статті та відео з актуальних питань прикладної лінгвістики, словник базової лексики, завдання різних типів, тематичні фото для опису, теми для есе та інструкції для написання, що в комплексі забезпечує формування відповідних програмних компетентностей:

- здатність спілкуватися іноземною мовою у професійній діяльності, опрацьовувати фахову літературу іноземною мовою;

- використовувати іноземну мову для організації ефективної міжкультурної комунікації;

- використовувати іноземну мову в усній та письмовій формі, у різних жанрово-стильових різновидах і регістрах спілкування, для розв'язання комунікативних завдань у різних сферах життя;

- створювати усні й письмові тексти різних жанрів і стилів іноземною мовою.

INTRODUCTION

Undoubtedly, Aplied Linguistics has developed rapidly and is now flourishing, with academic positions, academic departments, international journals, an international association *(Association Internationale de Linguistique Applique (AILA),* to be as well referred to as a profession in regard of all the above-mentioned apparatus. Apparently, the "linguistics applied" view has derived from the two strong traditions [13]: 1) the European philological tradition, exported to the USA through a number of scholars (Roman Jakobson and oth.); 2) the North American tradition of linguistic-anthropological field-work which required the intensive use of non-literate informants and the linguistic description of indigenous languages for the purposes of cultural analysis.

The textbook "Practical Course of Foreign Language" has been designed for students of the 3rd year of the Faculty of Philology majoring in specialty 035 "Philology" and, in particular, specialization 035.10 "Applied Linguistics". A high level of proficiency in English is a key component of effective academic and scientific work of future applied linguists.

The purpose of the student's book under consideration is the formation of the communicative, linguistic, socio-cultural and professional competence of learners, which will make it possible to effectively apply the gained knowledge in professional activities.

The textbook consists of four sections and covers the relevant topics of applied linguistics such as: "Lexicography", "Dictionaries", "Language Corpora", "Discourse Analysis".

At the end of the elaborated textbook, a brief list of applied linguistics terms with their definitions provided, namely, "Concise Glossary of Applied Linguistics Terms" is presented and can be used for both classroom and student self-study work, as well as and the Self-Assessment Final Test for self-control is elaborated.

Each unit of the textbook contains the "Reading Section", "Speaking Section." "Picture Description Worksheet", "Listening (Audiovisual Comprehension) Section", "Writing Section" and "Self-Reflection Sheet". Reading tasks are developed on the basis of the collection of scientific publications on problems of applied linguistics "The Handbook of Applied Linguistics" (2004).

The book contains scientific articles and videos on topical issues of applied linguistics, basic professional vocabulary, various types of tasks, thematic photos for description, topics for essays and writing instructions, which in the complex provides the formation of relevant program competencies:

- ability to communicate in a foreign language in professional activities, to develop professional literature in a foreign language;
- use a foreign language to organize effective intercultural communication;
- use a foreign language in oral and written form, in different genre-style varieties and registers of communication, for solving communicative tasks in different spheres of life;
- create oral and written texts of different genres and styles in a foreign language.

UNIT 1
LEXICOGRAPHY

 READING SECTION

I. Answer the questions below. Give your reasons. Provide examples from your own experience, if possible.

1. What problems do lexicographers encounter in the process of vocabulary compilation?
2. What are the reasons for the occurrence of lexicography?
3. What is the origin of the word "lexicography"?
4. Why is lexicography viewed as a synthetic science?
5. What is the role of the dictionary for any specialist?

II. Comment upon the following quotes related to the topic of the unit under consideration. Do you agree or disagree to the ideas suggested? Give your reasons.

1. "I am not yet so lost in lexicography, as to forget that words are the daughters of earth, and that things are the sons of heaven. Language is only the instrument of science, and words are but the signs of ideas: I wish, however, that the instrument might be less apt to decay, and that signs might be permanent, like the things which they denote." *(Samuel Johnson)*

2. "The makers of dictionaries are dependent upon specialists for their definitions. A specialist's definition may be true or it may be erroneous. But its truth cannot be increased or its error diminished by its acceptance by the lexicographer. Each definition must stand on its own merits." *(Benjamin Tucker)*

3. "Every other author may aspire to praise; the lexicographer can only hope to escape reproach. *(Samuel Johnson)*

4. "There's no such thing as an unabridged dictionary." *(Jack Lynch)*

5. "The writing of a dictionary, therefore, is not a task of setting up authoritative statements about the "true meanings" of words, but a task of recording, to the best of one's ability, what various words have meant to authors in the distant or immediate past. The writer of a dictionary is a historian, not a lawgiver." *(Samuel Hayakawa)*

III. Study the topical vocabulary and read the article below. Find the contexts where the units under consideration are used in the text.

1	succinct (adj.)	[səkˈsɪŋ(k)t]	стислий, короткий
2	standpoint (n.)	[ˈstan(d)ˌpoint]	точка зору
3	thesaurus (n.)	[θɪˈsɔːrəs]	вичерпний словник; тезаурус
4	lexicon (n.)	[ˈlɛksɪk(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.)	[ˈtɛkstjʊəl ˈdeɪtə]	текстові дані
8	multi-volume dictionary (n.)	[wʌn-ˈvɒljʊm ˈdɪkʃ(ə)n(ə)ri]	багатотомний словник
9	one-volume dictionary(n.)	[mʌltɪ-ˈvɒljʊm ˈdɪkʃ(ə)n(ə)ri]	однотомний словник
10	bias (n.)	[ˈbaɪəs]	упередженість; необ'єктивність; (for) схильність, ухил
11	obsolescent (adj.)	[ɔbsəˈles(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.)	[dɪˈsɜːn]	відрізняти, відділяти, виділяти; розрізняти
15	sense distinctions (n.)	[sɛns dɪsˈtɪŋkʃənz]	смислові відмінності
16	usage restrictions(n.)	[usage restrictions]	обмеження у використанні
17	constraint (n.)	[kənˈstreɪnt]	обмеження
18	coverage (n.)	[ˈkʌvərɪdʒ]	охоплення
19	comprehensive (adj.)	[kɔmprɪˈhensɪv]	повний, докладний, детальний
20	unabridged (adj.)	[ʌnəˈbrɪdʒd]	повний, нескорочений
21	foster (v.)	[ˈfɒstə]	сприяти розвитку; заохочувати
22	supplant (v.)	[səˈplant]	витісняти (що-небудь)
23	conventional (adj.)	[kənˈvɛnʃənl]	звичний, традиційний, загальноприйнятий;
24	scholarly (adj.)	[ˈskɒləli]	учений; властивий ученим, науковий
25	computational (adj.)	[ˌkɒmpju(ː)ˈteɪʃən(ə)l]	обчислювальний, комп'ютерний
26	keep abreast of (v.)	[kiːp əˈbrɛst ɒv]	не відставати від, іти в ногу з
27	utility (n.)	[ju(ː)ˈtɪlɪti]	корисність, практичність, вигідність
28	facilitation (n.)	[fəˈsɪlɪˈteɪʃən]	полегшення, допомога
29	multifaceted (adj.)	[mʌltɪˈfæsɪtɪd]	багатогранний, багатосторонній
30	endeavor (n.)	[enˈdevər]	спроба, намагання

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.

Unit 1. Lexicography

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.

IV. 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?

V. 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

VI. 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.

VI. 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.

VIII. 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	

IX. 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.)		

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

comlpile, explicitly, restriction, coverage, facilitation, contribute,
produce, derivative, convention, edit

XI. 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 sucscssful.

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 midia open up quiet new possibilities for the presentation and use of lexicografical 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) Lexicografy is an longstanding cultural practice and a integral part of the intelectual tradition in literature societies.

XII. Open the brackets, putting the infinitive form of the verb given into the necessary tense form (active or passive).

1) From its beginnings several thousand years ago lexicography .... (to serve) primarily the real-life needs of written communication between members of human communities. 2) Those needs .... (to change) just as all living languages constantly .... (to change). 3) In many literate societies lexicography .... (to have) a centuries-old tradition with word lists and word books in scripts. 4) Since print culture .... (to replace) scribal culture some five centuries ago and .... (to usher) in the modern period in European lexicography, the printed book .... (to predominate). 5) Worldwide, no book on a language or on languages .... (to be) and .... (to be) more popular in education systems and in communities at large than the dictionary. 6) Lexicographers can be regarded as descriptive linguists in that they empirically .... (to analyse) and .... (to describe) language with a traditional emphasis on individual items of vocabulary. 7) However, they .... (not

to require) linguistic knowledge alone, but according to the particular dictionary project may draw on other non-linguistic. 8) Lexicography .... (to regard) as quite central to applied linguistics. 9) At the same time, it .... (to see) as a complex activity *sui generis* with its own principles, practices, problems, and traditions. 10) Over the past 20–30 years lexicography .... (to change) fundamentally and irreversibly. 11) The main factor .... (to be) the dramatic impact of the computer. 12) A secondary factor .... (to be) the rapid emergence of metalexicography or dictionary research as an academic discipline with an explosion of writing on and about dictionaries. 13) This article .... (to write) from a western European perspective and .... (to draw) primarily on material related to British and other English language lexicographies.

XIII. 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.

6) В українському мовознавстві прийнято трактувати лексикографію як розділ мовознавства, до компетенції якого належить створення словників і опрацювання їхніх теоретичних засад і тісний зв'язок із лексикологією. Фактично наша мовознавча традиція ніколи експліцитно не ставила питання про статус лексикографії. Тобто не йшлося про те, є лексикографія теоретичною або практичною галуззю знань чи взагалі окремою галуззю.

7) Лексикографічна наука будується на чіткій відповідності лінгвістичної теорії її законам і принципам, лексикографічна практика як мистецтво є окремим літературним жанром, де найважливіша роль відводиться майстерності й таланту автора – творцю словника.

8) Без словників неможливі наукові дослідження в математиці, біології, хімії, медицині, культурології, інформатиці тощо, словники визнаються основними, програмними джерелами багатьох наукових дисциплін. Саме тому сучасними вітчизняними й зарубіжними лінгвістами лексикографія на сьогоднішній день уважається синтетичною наукою.

9) Завдання, що стоять сьогодні перед лексикографією, належать до різноманітних галузей людської практики: переклад, викладання рідної й іноземних мов, комп'ютерні, інформаційно-пошукові системи, культурологія, етнографія, соціологія, проблеми професійного спілкування, психологія тощо.

10) Однак, нині функції лексикографії розширилися за рахунок вирішення не лише прикладних, але й інформаційно-пізнавально-методологічних завдань. Сьогодні наука лексикографія знаходиться на межі інформаційної технології: комп'ютерного мовознавства, практичних наукових експериментів, загальної й педагогічної лінгвістики.

11) Розширенню теоретичної й дослідницької роботи в галузі лексикографії сприяють наявність кількості лексикографічних центрів, розмаїття словників, потужний науковий потенціал.

12) Суб'єктивний підхід у лексикографії стає очевидним при аналізі подачі значень у словниковій статті багатозначних лексем. Саме лексикограф повинен прийняти рішення про виділення певної кількості значень, про їх порядкове розташування у словниковій статті.

13) Основними критеріями при створенні словника є великий обсяг, раціональний, продуманий спосіб подачі словникового матеріалу, швидкий та зручний пошук необхідного слова.

14) Не викликає сумнівів, що лексикографія – це наука про створення словників. Доказом того є теоретичні принципи, закони, постулати, розроблені багатьма теоретиками й практиками словникової справи.

15) Слід відзначити діалектичну двоєдність лексикографії як науки. Виступаючи, з одного боку, окремою науковою дисципліною поряд з іншими лінгвістичними дисциплінами, лексикографія, з іншого боку, представляє собою універсальну методологічну науку, необхідну найрізноманітнішим галузям людської практики.

SPEAKING SECTION.
PICTURE DESCRIPTION WORKSHEET

DISCUSSING FACTUAL INFORMATION

Where? There is a classroom/hall/library room/lecture room at the university/ library/school. It looks out-of-date/modern. The people in the picture belong to different/the same ethnic group/(s)/nationalit(y)ies/origin.

When? The scene takes place during the lesson/match/excursion/staff meeting/ learned society meeting. We can/cannot see in the picture, what is the weather like outside. The scene takes place in the early morning/at midday/in the evening/at midnight.

What else can be mentioned about the setting? What can you see in the background and the foreground?

Who and what? How many? There are _____ people in the _____, a _____ and three young _____. The professor is wearing a _____ shirt and a grey_____. He has his_____ in front of him. The scholars are dressed quite _____. The young woman _____ is wearing a_____ jumper, the woman in the middle has a white_____ on and the young man is wearing a _____. They are _____ to the professor and making _____. The people in the picture are _____ the _____ topical issues of modern _____.

blouse	on the left	grey t-shirt	laptop	lexicography	library room	blue notes
scholars	four	casually	professor	discussing	jumper	grey	listening

DISCUSSING CONCEPTUAL INFORMATION
1. Who could take the picture?
2. What is the possible title of the picture under consideration? How can it be related to Applied Linguistics area? What is the message of the picture?
3. Are the people in the picture posing for camera on purpose?
4. What do you think had happened before the picture was taken?
5. What will they do next after the picture having been taken?

PICTURE DESCRIPTION
Make up a list of 20 key-words covering factual and conceptual information of the picture considered. Describe the picture using the key-words, cover factual and conceptual information of the picture considered.

LISTENING (AUDIOVISUAL COMPREHENSION) SECTION

THE JOY OF LEXICOGRAPHY

I. Watch the talk by the leading lexicographer Erin McKean presented at an official TED conference in March 2007 and answer the following questions. Justify your answer relying upon the facts form the video material. (https://www.youtube.com/watch?v=J4VzuWmN8zY&t=4s)

1) Why does the presenter blame the Queen?
2) Why is lexicography more about material science?
3) How many books in the Library of congress are in English?
4) What is the common lexicographical cope out?
5) Why do people use the dictionary synecdochically?

II. Indicate whether the statements a true or false, correct the false ones.

1) If the presenter had to think of some kind of occupation as a metaphor for her work, she would be a traffic cop. She wants to let the good words make that difficult left-hand turn into the dictionary, and keep the bad words out. ____

2) McKean considers that all the words should be studied, because beautiful expressions can be made from unknown and rarely used parts. ____

3) For the presenter the English language is a mobile which is moved with the help of words used in a new context or with a new connotation. ____

4) Newspaper archive goes back to 1759, 58.1 million newspaper pages of un-dictionaried words. ____

5) The Oxford English Dictionary has around 13 different numbered definitions for set. ____

6) In the dictionary map of American English there is neither Florida nor California. ____

7) The Internet is created of words and enthusiasm and those two are crucial for lexicography. ____

8) A lot of word-collecting sites are not scientific because they do not show the dictionaries which the words were taken from. ____

9) We can compare a word with an archaeological artifact because if its provenance or the source is unknown it does not have any scientific value. ____

10) The presenter doesn't want her son to use only online dictionaries instead of paper ones. ____

III. Fill in the gaps with appropriate words according to Erin McKean's talk.

People think that …………….. job is to let the good words make that difficult left-hand turn into the dictionary, and keep the bad words out. James Murray was the first …………….. of the Oxford English Dictionary. He's really responsible for a lot of what we consider …………….. in dictionaries today. What a modern dictionary is, the Victorian …………….. merged with a little bit of modern propulsion. It's steampunk because we have Victorian design with an …………….. on it. The design has not …………….... And in fact, online dictionaries …………….. almost all the problems of print, except for searchability. And when you improve …………….., you actually take away the one advantage of print, which is serendipity. Serendipity is when you find things you weren't …………….., because finding what you are looking for is so damned difficult.

WRITING SECTION

Choose one of the topics below to write an opinion essay. Before writing study the tips.
1) The future of lexicography depends on the development of the Internet and digital mass media.
2) The 21st century will be the era of lexicography decay.
3) Lexicographers will be replaced by neural networks in the nearest future.

<u>Tips for writing</u>

1. Decide whether you agree or disagree with the title. Try to think of at least two or three good reasons to support your opinion, including examples of why you think the alternative point of view is wrong.
2. Introduce each paragraph with a topic sentence, outlining the main ideas. Do not write about advantages or disadvantages or points for or against.
3. Write in formal style.
4. Don't use colloquial expressions, short forms or emotive vocabulary.
5. Organise your essay into clear <u>paragraphs</u>:
   a) <u>an introductory paragraph</u> in which you state the topic and your opinion;
   b) <u>a main body</u> which consists of several paragraphs, each presenting a separate viewpoint supported by reasons;
   c) <u>a conclusion</u> in which you restate your opinion using different words.

SELF-REFLECTION SHEET ON THE UNIT 1 MATERIALS:

1) THE FOLLOWING 10 KEY-CONCEPTS OF THE UNIT CAN BE SINGLED OUT:

_____
_____

2) UNIT REVIEW: THE UNIT GENERALLY RUNS ABOUT:

_____
_____
_____
_____

3) WHAT ARE THE ADVANTAGES OF THE UNIT UNDER CONSIDERATION?

_____
_____
_____

4) WHAT ARE THE DISADVANTAGES OF THE UNIT UNDER CONSIDERATION?

_____
_____
_____

5) WHICH TASKS DO YOU CONSIDER TO BE MOST USEFUL IN THE UNIT?

_____
_____
_____

6) TAKE A LOOK AT THE TITLE OF THE NEXT UNIT. WHAT DO YOU EXPECT TO LEARN WHEN YOU START COVERING THE MATERIALS OF THE UNIT?

_____
_____
_____

UNIT 2
DICTIONARIES

PART 1

READING SECTION

I. Answer the questions below. Give your reasons. Provide examples from your own experience, if possible.

1. For what purposes are dictionaries typically used?
2. What types of dictionaries do you know?
3. What is the origin of the word "dictionary"? Why do we call word collections a "dictionary"?
4. What is the purpose of using encyclopedias?
5. What free online encyclopedias have you ever heard about?

II. Comment upon the following quotes related to the topic of the unit under consideration. Do you agree or disagree to the ideas suggested? Give your reasons.

1. Dictionaries should be designed with a special set of users in mind and for their specific needs. *(William Householder)*
2. "A lexicographical definition, we shall argue, does not in most cases identify a meaning independently existing in actual usage and discovered there by the lexicographer: it is deliberately constructed and allocated by the lexicographer on the basis of materials selected for study, and its allocation will depend on the viewpoint the lexicographer has chosen to adopt." *(Roy Harris)*
3. "The notion of a definition adequate to all occasions and all demands is a semantic *ignis fatuus*." *(Roy Harris)*.
4. "Dictionaries are like watches; the worst is better than none, and the best cannot be expected to be quite true." *(Samuel Johnson)*
5. "Spellings are made by people. Dictionaries – eventually – reflect popular choices." *(David Crystal)*

III. Study the topical vocabulary and read the article below. Find the contexts where the units under consideration are used in the text.

1	exemplification (n.)	[ɪgˌzɛmplɪfɪˈkeɪʃən]	ілюстрація, пояснення прикладом
2	emerge (v.)	[ɪˈmɜːʤ]	з'являтися, виникати; з'ясовуватися
3	widespread (adj.)	[ˈwaɪdsprɛd]	розповсюджений, поширений
4	inextricably (adv.)	[ɪnˈɛkstrɪkəbli]	нерозривно

Unit 2. Dictionaries

5	bewildering (adj.)	[bɪˈwɪldərɪŋ]	той, що бентежить, дивує; що заганяє в глухий кутаганяє в
6	notable (adj.)	[ˈnəʊtəbl]	помітний, відчутний, очевидний; вартий уваги, значущий
7	nonetheless (adv.)	[ˌnʌnðəˈlɛs]	тим не менш, все таки, однак, проте
8	justifiable (adj.)	[ˈdʒʌstɪfaɪəbl]	виправдано; що може бути виправданим
9	bulky (adj.)	[ˈbʌlki]	громіздкий, великий
10	pictorial (adj.)	[pɪkˈtɔːrɪəl]	ілюстрований (про періодичне видання); яскравий, рисунковий, образотворчий
11	unwieldy (adj.)	[ʌnˈwiːldi]	громіздкий
12	proper (adj.)	[ˈprɒpə]	відповідний, правильний; властивий
13	process (v.)	[ˈprəʊsɛs]	обробляти (дані, інформацію тощо)
14	ensure (v.)	[ɪnˈʃʊə]	забезпечувати, гарантувати
15	restrict (v.)	[rɪsˈtrɪkt]	обмежувати
16	contrast (v.)	[ˈkɒntrɑːst]	порівнювати, протиставляти
17	arrange (v.)	[əˈreɪndʒ]	організовувати, налагоджувати, пристосовувати, налаштовувати
18	refer (v.)	[rɪˈfɜː]	відносити до чогось, посилатися, звертатися
19	access (n.)	[ˈæksɛs]	доступ
20	enhance (v.)	[ɪnˈhɑːns]	покращувати; збільшувати; підви-щувати
21	determine (v.)	[dɪˈtɜːmɪn]	визначати, зумовлювати; (іноді контекст. обчислювати)
22	store (v.)	[stɔː]	зберігати; запасати; постачати
23	item (n.)	[ˈaɪtəm]	одиниця; пункт; питання; деталь; повідомлення
24	overwhelming (adj.)	[ˌəʊvəˈwɛlmɪŋ]	переважний; нездоланний; незліченний
25	portable (adj.)	[ˈpɔːtəbl]	портативний, переносний
26	designation (n.)	[ˌdɛzɪgˈneɪʃən]	вказівка, призначення; визначення; назва
27	uncontested (adj.)	[ˌʌnkənˈtɛstɪd]	безперечний, очевидний
28	exhaustive (adj.)	[ɪgˈzɔːstɪv]	вичерпний; виснажливий
29	majority (n.)	[məˈdʒɒrɪti]	більшість
30	surrogate (adj.)	[ˈsʌrəgɪt]	несправжній, недійсний; той, що може виступати в якості замінника

Unit 2. Dictionaries

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.

Unit 2. Dictionaries

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.

Unit 2. Dictionaries

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, Wųrterbuch*) in the title, and it is this term that is firmly entrenched as the coverall 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?

Unit 2. Dictionaries

VI. 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?

V. 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

Unit 2. Dictionaries

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

1	*Lexicon* is used further, often in the collocation or compound *mental lexicon*,	**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.

VII. 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 some …………….thesauruses now have alphabetical indexes to ensure easy access, especially when the items are grouped according to a …………….world view such as those which determined the organization of older thematic encyclopedias.

3) It also needs to be said that the …………..classifies 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 contemporary ………….vocabulary, 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 translation …………in the other (target) language.

7) Different societies have different lexicographical traditions, and ideas on what might ………….the prototypical dictionary vary accordingly.

8) There are dictionaries dealing …………..with specific text types, texts by individual authors, and concordances.

9) Full equivalents may need mere listing, while partial and ………equivalents require further explanation or ………..to ensure sense identification and discrimination.

10) Humans use language to communicate about facts, things, and people; words and the world are ………..linked.

VIII. 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	смислове розрізнення	
11	потребувати подальшого роз'яснення	
12	простіше кажучи	
13	макро – і мікроструктура	
14	власне мовна та позамовна (екстралінгвістична) інформація	
15	галузь, область лексикографії	

Unit 2. Dictionaries

IV. 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.)		

X. 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

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

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

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

3) Marked lemma include archayisms, neologysms, regionalysms, and internationalysms.

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 tipology is complemented by a functionel 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 bylingual dictionaries had long been and continueing to be very common, but this is not false of encyclopedias.

XII. Fill the gaps with a necessary preposition from the box below.

of (3)	to (4)	by	with	out	in	from

1) The components additional .... the central word list or the dictionary entries from A to Z consist .... front, middle, and back matter, often including the inside covers and, increasingly, the outside covers and dust jacket. 2) The front matter contains most importantly a user's guide or key .... the dictionary. 3) The key is now considered essential, but often seems to be ignored .... users and reviewers alike. 4) It explains style, structure and content .... the dictionary: the metalanguage, symbols and codes used, the punctuation and the complex typography, and the layout of the entries. 5) It often takes the form of reproductions of sample entries .... each component of the macro- and microstructure highlighted and commented on in turn. 6) It sometimes stands alone and sometimes accompanies a longer introduction .... the dictionary outlining the editorial principles underlying the work. 7) The middle matter might consist of small, half- or full-page panels devoted .... grammar and/or usage notes, frequency charts, word-formation items and patterns, lexical sets or pragmatic conventions; or it might feature inserted study pages, illustrations to make the inserts stand .... 8) In many cases the material .... such inserts is reserved for appendices in the back matter. 9) These might contain both linguistic and encyclopedic information of all kinds ranging .... style guides, prefixes and suffixes, and different alphabets to weights and measures, chemical elements, and countries of the world. 10) Some dictionaries have no back matter, others have as many as 100 pages .... appendices.

XIII. 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.

6) Довідковим виданням називають видання, що містить короткі відомості наукового чи прикладного характеру, розташовані в порядку, зручному для їх швидкого пошуку, не призначене для суцільного читання. До найважливіших різновидів довідкових видань відносяться словники, довідники, енциклопедії.

7) Для складання довідкового видання створюється редакційна колегія, яка повинна забезпечити цільову спрямованість, повноту і науковість видання, також організовуються галузеві редакції по окремих галузях знань. Складання словника вимагає злагодженої, організованої роботи великого кола фахівців.

8) Більшість сучасних енциклопедій одночасно призначені як для кваліфікованого читача і можуть використовуватись в його фаховій, науковій чи виробничій діяльності, так і для масового читача, який має на меті лише отримати коротку довідку з певного питання, що його цікавило.

9) Насамперед можна виокремити універсальні, галузеві, спеціальні та регіональні енциклопедії. Якщо універсальні енциклопедії подають інформацію з усіх галузей знань та практичної діяльності (фізики, сільського господарства, живопису, статистики тощо), то галузеві присвячені окремим галузям знань чи певним напрямам практичної діяльності людини.

10) Спеціалізовані енциклопедії викладають матеріали з певної теми (так, енциклопедія «Живий світ геральдики» розкриває одну локальну тему) або факти життя та діяльності видатного науковця, митця, літератора (прикладом може бути «Шевченківський словник»). Відповідно такі довідники називають тематичними та персональними енциклопедіями.

11) Енциклопедичний словник – довідкове видання, що містить відомості про предмети, явища та осіб, визначених тематикою, наукової галуззю або універсальних. Статті в ньому розташовані в порядку алфавітного сортування назв предметів. У свою чергу, лінгвістичний словник описує лексичні одиниці з точки зору мовознавства, дає роз'яснення значення і вживання слів.

12) Акумулюючи інформацію з різноманітних галузей знань та практичної діяльності, систематизуючи й узагальнюючи здобутки національної культури, енциклопедичні видання вимагають від усіх, причетних до їх підготовки та видавничого опрацювання, ґрунтовних фахових знань.

13) Довідкові видання не допускають неточних, не перевірених, спотворених фактичних даних. Отримуючи в руки довідкове видання, читач повинен бути впевнений в тому, що він користується достовірними відомостями і вони відповідають реаліям сьогоднішнього дня. Тому перевірка фактичної сторони матеріалу є частиною роботи редактора.

14) Для цього необхідно чітке розмежування логічних частин і різних елементів видання (основного тексту, ілюстрацій, коментарів до них, і в той же час відображення їх ролі та взаємозв'язків), оформлення текстового матеріалу (розрядка, курсив тощо), наявність ілюстративного матеріалу, який несе в собі рівне з текстом інформаційне навантаження, організація системи пошуку: зміст, покажчики, колонтитули тощо.

15) Необхідний компонент будь-якого довідника – апарат видання. У нього входять: вихідні відомості, зміст, передмова, перелік умовних позначень та скорочень, бібліографічний список, покажчики. Він потрібен для того, щоб відрізнити видання від іншого, швидко і безпомилково зорієнтуватись у його змісті, мати можливість ідентифікувати його і здійснити пошук в потоках і масивах інших видань.

SPEAKING SECTION.
PICTURE DESCRIPTION WORKSHEET

DISCUSSING FACTUAL INFORMATION

Where? There is a classroom/hall/lecture room at the university/library/school. It looks out-of-date/modern/poorly/well-equipped. The people in the picture belong to different/ the same ethnic group/(s)/nationalit(y)ies/origin. The people represent the same/ different age groups and gender.

When? The scene takes place during the /lecture/seminar/match/excursion/staff meeting. We can/cannot see in the picture, what is the weather like outside. The scene takes place in the early morning/at midday/in the evening/at midnight.

What else can be mentioned about the setting? What can you see at the background and at the foreground?

Consider the statements below *true, false* or *not stated*. Justify your choice:
1) There are 17 students in the classroom._____
2) There are 5 women and 11 men in the picture._____
3) None of the students in the classroom is wearing glasses._____
4) The teacher is in her early thirties._____
5) The teacher is wearing a dark-blue blouse._____
6) The students seem to be rather passive and uninvolved._____
7) The teacher is not speaking at the moment, being silent and waiting for students to speak out.
8) The teacher is rather dynamic, moving actively around the classroom.___

DISCUSSING CONCEPTUAL INFORMATION
1. Who could take the picture?
2. What is the possible title of the picture under consideration? How can it be related to Applied Linguistics area? What is the message of the picture?
3. Are the people in the picture posing for camera on purpose?
5. What do you think had happened before the picture was taken?
6. What will they do next after the picture having been taken?

PICTURE DESCRIPTION

Make up a list of 20 key-words covering factual and conceptual information of the picture considered. Describe the picture using the key-words, cover factual and conceptual information of the picture considered.

Unit 2. Dictionaries

PART 2

READING SECTION

I. Answer the questions below. Give your reasons. Provide examples from your own experience, if possible.

1. What are the subject areas of appled linguistics for lexicography?
2. Are dictionaries and glossaries of technical terms mostly thematic or alphabetical in organization and presentation?
3. What does a dictionary entry consist of?
4. What types of dictionaries are more convenient traditional or computer ones?
5. How do computer dictionaries differ from their paper versions?

II. Comment upon the following quotes related to the topic of the unit under consideration. Do you agree or disagree to the ideas suggested? Give your reasons.

1. "I was reading the dictionary. I thought it was a poem about everything." *(Steven Wright)*

2. "The dictionary is like a time capsule of all of human thinking ever since words began to be written down. And exploring where words have come from can increase your understanding of the words themselves and expand your understanding of how to use the words, and all of this change happens in your thinking when you read the words." *(Andrew Clements)*

3. "The dictionary is based on the hypothesis – obviously an unproven one – that languages are made up of equivalent synonyms." *(Jorge Luis Borges)*

4. "If I were allowed to take just one book to the proverbial desert island, it might be a dictionary." *(Steven Pinker)*

5. "Dictionary: a malevolent literary device for cramping the growth of a language and making it hard and inelastic." *(Ambrose Bierce)*

III. Study the topical vocabulary and read the article below. Find the contexts where the units under consideration are used in the text.

1	constituent (adj.)	[kənˈstɪtjʊənt]	складовий
2	extract (v.)	[ɪksˈtrækt]	вибирати; діставати
3	arrive (at)	[əˈraɪv]	досягати (чого-небудь)
4	disposal (n.)	[dɪsˈpəʊzəl]	право розпоряджатися; управління; використання
5	editor (n.)	[ˈɛdɪtə]	редактор
6	relevant (adj.)	[ˈrɛlɪvənt]	відповідний
7	frequently (adv.)	[ˈfriːkwəntli]	часто

8	equate	[ɪˈkweɪt]	вважати рівним; рівняти, прирівнювати; ототожнювати
9	core (adj.)	[kɔ:]	основний; центральний
10	suffice	[səˈfaɪs]	бути достатнім; вистачати; задовольняти
11	accordance	[əˈkɔ:dəns]	відповідність, злагода
12	sheer (adj.)	[ʃɪə]	єдиний; тільки один
13	vernacular (n.)	[vəˈnækjʊlə]	рідна мова; національна мова
14	avowedly (adv.)	[əˈvaʊɪdli]	відкрито, прямо, явно, гласно; за загальним визнанням
15	arbiter usus (n.)	[ˈɑ:bɪtə usus]	авторитетний суддя
16	controversy (n.)	[ˈkɒntrəvɜ:si]	суперечка, полеміка, розбіжність у думках
17	proscribe (v.)	[prəʊsˈkraɪb]	виганяти, висилати засуджувати, забороняти
18	contemporary (adj.)	[kənˈtɛmpərəri]	сучасний
19	permissive (adj.)	[pəˈmɪsɪv]	дозвільний, який дозволяє, який допускає
20	abdication (n.)	[ˌæbdɪˈkeɪʃ(ə)n]	складання повноважень, відмова
21	alleged (adj.)	[əˈlɛdʒd]	нібито наявний; такий, наявність якого стверджується; заявлений
22	rival (adj.)	[ˈraɪvəl]	який суперничає, конкуруючий
23	in terms of	[ɪn tɜ:mz ɒv]	з точки зору
24	indigenous (adj.)	[ɪnˈdɪdʒɪnəs]	місцевий, корінний
25	taxonomy (n.)	[tækˈsɒnəmi]	систематика, таксономія
26	dispersed (adj.)	[dɪsˈpɜ:st]	розсіяний; розосереджений
27	contiguous (adj.)	[kənˈtɪgjʊəs]	суміжний, дотичний, прилягаючий; сусідній; близький
28	pluricentric language (n.)	[ˌpluərɪˈsentrɪk ˈlæŋgwɪdʒ]	плюрицентрична мова (з декількома стандартними варіантами, наприклад, англійська з британським і американ-ським варіантами)
29	endonormative (adj.)	[ˌendəʊˈnɔ:mətɪv]	місцево визначений
30	distinct (adj.)	[dɪsˈtɪŋkt]	відмінний, який відрізняється; несхожий

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 & Ebbit, 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."

IV. 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 bee 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.

Unit 2. Dictionaries

V. 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	serving 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

VI. 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 *arbiter usus*, being concerned to avoid prescribing or proscribing usage.
5	Both areas draw *inter alia* 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	h	according to the nature of the translation, general or specialized, literary or scientific.

	prescriptive and normative in intention and approach,		
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.

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

glossaries *inter alia* 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.

Unit 2. 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.

VIII. 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	
9	все частіше	
10	єдина кількість термінів	
11	уникати заборони використання	
12	сучасна лексикографія	
13	дозвільна відмова	
14	отримати еквівалентне формулювання	
15	передбачувана відповідальність лексикографів	

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

№	Word	Synonym	Antonym
1	core (adj.)		
2	constituent (adj.)		
3	abdication (n.)		
4	range (n.)		
5	normative (adj.)		
6	variety (n.)		
7	prescribe (v.)		
8	codify (v.)		
9	effort (n.)		
10	remove (v.)		

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

establishment   controversy   tend   benefit   suffice   equate   frequently permissive   dispersed   expression

Unit 2. Dictionaries

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

1) The dictionaris range from general-purpose dictionaries of the second language and thesaurus and sinonim dictionaris of the first language to mono- and twolingual subject-specific tehnical dictionaris and glossaris.

2) Literary translators must to have an ability to extracting meaning from text in one language and to arrive at an equivalent formulation in other that could only benefit billingual lexicographers.

3) Other areas of applied linguistics is communication in the profeccions and languages for special purposes, both of whom have at their disposal a wast range of specialized, subject-specific referense works.

4) The authoritarian tradition is firmly established, and publishers still often appealing to it in their advertcing.

5) In some quarters this policy was been viewed as a permissive abdication of the alleged responsibility of lexicografers 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.

XII. Open the brackets, putting the infinitive form of the verb given into the necessary tense form (if needed).

1) Macro-structure refers to the list and organization of the lexical items .... (to enter) in the dictionary, the lemmas or headwords. 2) Lemma .... (to prefer) here as it is neutral on the morphological status of the items. 3) In practical terms the lemma list .... (to depend) on the projected size and scope of the dictionary. 4) It .... (to range) from reasonably comprehensive, as in large unabridged works, to highly selective, as in small pocket dictionaries.

5) Depending on size and intention, current one-volume defining dictionaries tend to emphasize the central core vocabulary of present-day standard usage and .... (to focus) as well on new words and senses and on terms from science and technology. 6) Different dictionaries have different policies on the information they .... (to regard) as lexically relevant and on the order in which they present it. 7) Decisions must .... (to make) on giving each item main lemma status or distinguishing between main lemmas and sub-lemmas.

8) In the latter case, lexicographers must determine on what grounds main lemmas .... (to distinguish) from sub-lemmas, how these are grouped or organized in nests or niches, and whether all or some of the sub-lemmas are supplied with a full or partial range of lexicographic information or whether they are simply listed as run-ons. 9) Decisions must also .... (to make) on the ordering of homographic lemmas and on the typography of the different types of lemma. 10) Here, as elsewhere, the chief macro-structural criterion must be userfriendliness: the user must be able to find the item .... (to look for) as quickly and easily as possible.

XIII. 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) Суворо 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.

6) Найбільш давньою прикладною сферою мовознавства можна вважати прикладну лексикографію. На сучасному етапі її розвитку цей напрямок прикладної лінгвістики, окрім розробки традиційних паперових словників, збагатився таким аспектом, як комп'ютерна лексикографія, що здатна продукувати такі електронні лексикографічні форми, якими є інформаційно-пошукові тезауруси.

7) Комп'ютерний варіант чи копія традиційного, укладеного людиною словника подають лише нову, комп'ютерну форму інформації, вміщеної в традиційних словниках, а отже, це лише трансформація форми уже готового продукту лексикографічного опрацювання мовного матеріалу традиційними, некомп'ютерними методами.

8) Комп'ютерний словник і комп'ютерний варіант традиційного словника становлять результати двох напрямків роботи у сучасній лексикографії: переведення у комп'ютерну форму вже існуючих словників, укладених людиною, і розроблення поняттєвого та процедурного апарату лінгвістичних алгоритмів для конструювання комп'ютерних словників нових типів.

9) Для забезпечення можливості виконання нових дослідницьких завдань створюються не комп'ютерні копії, а комп'ютерні версії традиційних словників зі своїми граматиками аналізу та синтезу текстів словникових статей.

10) Перевага комп'ютерних версій над комп'ютерними копіями полягає у можливості їхнього багаторазового й багатоаспектного використання, автоматичної переорганізації, доповнення чи стиснення вміщеної в них інформації про ті чи інші мовні об'єкти.

11) Інший напрямок досліджень з лексикографії становить конструювання словників з новими лінгвістичними об'єктами як одиницями реєстру, або одиницями опису й пояснення в таких словниках. Прикладом такого типу словників є створювані на основі одномовних тлумачних словників семантичні, або ідеографічні словники.

12) Комп'ютерна лексикографія з самого початку була задумана не як самостійна наукова галузь, а як автоматизація трудомістких процесів у лексикографії. Реалізація цього задуму поступово не лише накопичила досвід автоматичного аналізу мовного матеріалу з позицій лексикографії, а й привела до виникнення теоретичних засад автоматизації в лексикографії та автоматизованого укладання словників.

13) Вироблення концепції словника – сфера лексикографа, який використовує свій досвід та досвід своїх попередників.

14) Укладання реєстру словника теж може здійснюватися комп'ютером на базі укладеної електронної картотеки або списку лем, які трапилися в корпусі текстів. Вирішальне слово тут – за лексикографом, який визначає доцільність включення слова залежно від попередньо виробленої концепції майбутньою словника.

15) Розвиток комп'ютерних технологій сприяє переходу лексикографії на якісно новий рівень, значно прискорюючи лексикографічний процес і забезпечуючи високу ефективність роботи користувачів, зокрема фахівців різних галузей.

Unit 2. Dictionaries

SPEAKING SECTION.
PICTURE DESCRIPTION WORKSHEET.

DISCUSSING FACTUAL INFORMATION

Where? There is a classroom/hall/lecture room at the university/library/school. It looks out-of-date/modern. The people in the picture belong to different/the same ethnic group/(s)/nationalit(y)ies/origin.

When? The scene takes place during the lesson/conference/lecture/staff meeting. We can/cannot see in the picture, what is the weather like outside. The scene takes place in the early morning/at midday/in the evening/at midnight.

What else can be mentioned about the setting? What can you see in the background and the foreground?

Who and what? How many? There are _____ people in the _____. They are working _____. Every group is working around _____. There are _____ on the tables. They are working on _____ about the role of _____ in _____ linguistics. The group on the front consists of _____ people. They all are wearing _____ clothes. Four girls are making _____ to present their _____. The boy and _____ other girls are preparing the _____ to _____ the poster.

poster	eight	projects	casual	presentation	in groups	hall	foundlings
dictionaries	applied	support	three	laptops	the table		about seventy

DISCUSSING CONCEPTUAL INFORMATION
1. Who could take the picture?
2. Are the people in the picture posing for camera on purpose?

3. What is the possible title of the picture under consideration? How can it be related to Applied Linguistics area? What is the message of the picture?
4. What do you think had happened before the picture was taken?
5. What will they do next after the picture having been taken?

PICTURE DESCRIPTION
Make up a list of 20 key-words covering factual and conceptual information of the picture considered. Describe the picture using the key-words, cover factual and conceptual information of the picture considered.

LISTENING (AUDIOVISUAL COMPREHENSION) SECTION

LANGUAGE DESIGN (INTERVIEW WITH NOAM CHOMSKY)

I. Watch a video recording with a famous linguist Noam Chomsky being interviewed on the language design issues. Answer the following questions. Justify your answer relying upon the facts from the video material. (https://www.youtube.com/watch?v=MLk47AMBdTA)

1) Can any creative activities be realized without language mechanisms?
2) Why is the issue of language design strongly related to the issue of human evolution?
3) Why does the interviewee compare a language phenomenon to a snowflake?
4) Do languages look very different from one another or very similar to one another on the surface? Why?
5) In which terms does language, being an optimal communication system, follow the laws of nature?

II. Indicate whether the statements below are true/ false/not stated, justify your choice relying on the video materials.

1) From Noam Chomsky's perspective, there are three systems analogous to human language. ____
2) Humanity had its language emerged apparently ten thousand years ago. _____
3) We know a lot about the human visual system because of direct experimentation with cats and monkeys. _____
4) Language by its nature cannot be regarded as a computational system._____
5) There has been detectable evolution of human's cognitive and creative capacities in roughly the past 50.000._____
6) There is obviously no progress in carrying out invasive experiments with human beings in relation to language capacity. _____

7) Language capacity is viewed as a core of human sensibility, and a creative and cognitive capacity._____

8) All humans are pretty much identical with regard of the cognitive capacity, linguistic capacity and so, which means, that there's been essentially no detectable evolution. _____

9) Generally, human's language capacity is analogous to animal's communication capacity. _____

10) A snowflake in terms of its physical structure cannot be in anyway compared to language as a complex computational system._____

III. Fill in the gaps with appropriate words according to what is being delivered in the video-piece.

<…> There are animal ……… systems, but they're completely different in design and use in just about every …………... So, something strange happened, roughly, maybe, a hundred thousand years ago, not very long, and ………… emerged in humans, and the question then is, well, what kind of a system is it? On the …………. languages look very different from one another. So, if somebody walks into the room and starts speaking …………, I'm not going to understand a word. Though I will ………… that it's a language. I won't understand it, but I know it's not noise. No. As soon, as you look more deeply, you find that ……… languages are basically mould into a pretty similar …………., may be an identical design, the large parts of the ………… of what we hear, is just the sounds. But that's a very superficial part of language. The core of language is principles that ………… actually an infinite array of possible expressions, ………….. expressions, which have definite meanings. Now, all of that is well, beyond, what we can just …………., but I say looking at the texts, and when a child is learning a language, the ………. doesn't learn those things, there's no evidence for them. Almost, no …………… for them, nobody can teach them. <…>.

WRITING SECTION

Choose one of the topics below to write a for-and-against (argumentative) essay. Before writing study the tips.

1) Can humanity exist without language as means of communication?
2) Does language reflect the life of society or is it society that impacts the language system and structure?
3) Can artificially constructed language *(as Esperanto, for instance)* improve international communication and foster considerable progress of humanity?

Tips for writing

The Structure

1. *Introduction:* make some general statements about the topic, mention what the current situation is, without giving your opinion;

2. *The main body:* support your arguments with examples or reasons:
   - arguments for – mention about the advantages or reasons in favour;
   - arguments against (one paragraph) – give the disadvantages or reasons against;

Each paragraph should start with a topic sentence which summarizes the topic of the paragraph.

3. *Conclusion*: sum up the topic, make a short personal opinion or give a balanced consideration of the topic. Opinion clichés (*I think, in my opinion*, etc.) can only be used in the closing paragraph where you give your opinion on the topic.

Hints for writing

1) make a list of pros and cons;
2) start with a topic sentence;
3) use linking words to connect your ideas;
4) don't use informal language or contracted forms;
5) try to sound unemotional (*it seems that, therefore one feels that*…, etc.).

Useful language:

To list points:

Firstly, First of all, In the first place, To begin/start with, Secondly, Thirdly, Finally

To list advantages:

One/Another/A further/An additional (major) advantage of … is …The main/greatest/ first advantage of … is …

To list disadvantages:

One/Another/A further/An additional (major) disadvantage/drawback of …The main/ greatest/most serious/first disadvantage/drawback of …

Another negative aspect of …

To introduce points/arguments for & against:

One (very convincing) point/argument in favour of … / against …, A further common criticism of … / It could be argued that …,

It is often widely / generally claimed/suggested /argued/maintained/felt/believed/held that some/many/most people/experts/scientists/sceptics/critics claim/suggest/argue/ feel that …maintain/believe/point out/agree/hold that…advocate (+ing/noun)/support the view that …oppose the view that …are in favour of/against …of the opinion that/ convinced that …opposed to …

To add more points to the same topic:

in addition (to this), furthermore, moreover, besides, apart from, what is more, as well as, not to mention (the fact) that, also, not only …but also/as well/both … and, There is another side to the issue/question/argument of…

To make contrasting points:

on the other hand, however, still, yet, but, nonetheless, nevertheless, even, so, it may be said/argued/claimed that, …others/many people oppose this viewpoint (strongly) disagree …,claim/feel/believe this argument is incorrect/misguided although, though, even though, while, whilst, whereas, despite/inspite of (the fact that), regardless of the fact that…/While/Although…, it cannot be denied that …

SELF-REFLECTION SHEET ON THE UNIT 2 MATERIALS:

1) THE FOLLOWING 10 KEY-CONCEPTS OF THE UNIT CAN BE SINGLED OUT:
_____
_____

2) UNIT REVIEW: THE UNIT GENERALLY RUNS ABOUT:

_____
_____
_____
_____

3) WHAT ARE THE ADVANTAGES OF THE UNIT UNDER CONSIDERATION?
_____
_____
_____

4) WHAT ARE THE DISADVANTAGES OF THE UNIT UNDER CONSIDERATION?
_____
_____
_____

5) WHICH TASKS DO YOU CONSIDER TO BE MOST USEFUL IN THE UNIT?
_____
_____
_____

6) TAKE A LOOK AT THE TITLE OF THE NEXT UNIT. WHAT DO YOU EXPECT TO LEARN WHEN YOU START COVERING THE MATERIALS OF THE UNIT?
_____
_____
_____

UNIT 3
LANGUAGE CORPORA

PART 1

READING SECTION

I. Answer the questions below. Give your reasons. Provide examples from your own experience, if possible.

1. What does Empirical Linguistics study?
2. Do different social groups use characteristically elaborated or restricted language codes?
3. What is the role of new technologies in the fields of linguistics?
4. What is Corpus Linguistics?
5. What approaches can characterize Empirical Linguistics appropriately?

II. Comment upon the following quotes related to the topic of the unit under consideration. Do you agree or disagree to the ideas suggested? Give your reasons.

1. "Every linguistic sign is located on two axes: the axis of simultaneity and that of succession." *(Roman Jakobson)*
2. "It is common now to address theoretical issues through the examination of bodies of naturally occurring language use." *(Bybee)*
3. "Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency." *(Goldberg)*
4. "The patterns of a word can be defined as all the words and structures which are regularly associated with the word and contribute to its meaning. A pattern can be identified if a combination of words occurs relatively frequently, if it is dependent on a particular word choice, and if there is a clear meaning associated with it." *(Hunston & Francis)*
5. "[…] corpus linguistics is a whole system of methods and principles of how to apply corpora in language studies and teaching/learning, it certainly has a theoretical status. Yet theoretical status is not theory in itself." *(Mc Enery)*

III. Study the topical vocabulary and read the article below. Find the contexts where the units under consideration are used in the text.

1	corpus (*pl.* corpora) (n.)	[ˈkɔ:pəs]	корпус; збірник
2	sample (n.)	[ˈsɑ:mpl]	зразок, приклад
3	concordance (n.)	[kənˈkɔ:dəns]	узгодження, відповідність

4	pattern (n.)	[ˈpætən]	модель, зразок
5	assess (v.)	[əˈsɛs]	оцінювати
6	interact (v.)	[ˌɪntərˈækt]	взаємодіяти
7	forensic (adj.)	[fəˈrɛnsɪk]	судовий, судочинний
8	authentic (adj.)	[ɔːˈθɛntɪk]	автентичний, оригінальний
9	reference (n.)	[ˈrɛfrəns]	посилання; довідка
10	embody (v.)	[ɪmˈbɒdi]	включати; об'єднувати
11	observable (adj.)	[əbˈzɜːvəbl]	видимий, помітний, доступний для огляду
12	monolingual (adj.)	[ˌmɒnəʊˈlɪŋgwəl]	одномовний
13	invaluable (adj.)	[ɪnˈvæljʊəbl]	безцінний
14	evidence (n.)	[ˈɛvɪdəns]	свідчення
15	descriptive (adj.)	[dɪsˈkrɪptɪv]	описовий
16	authorship (n.)	[ˈɔːθəʃɪp]	авторство
17	refute (v.)	[rɪˈfjuːt]	спростовувати
18	denote (v.)	[dɪˈnəʊt]	позначати
19	persist (v.)	[pəˈsɪst]	наполягати
20	disapprove (v.)	[ˌdɪsəˈpruːv]	не схвалювати, спростовувати
21	recurrent (adj.)	[rɪˈkʌrənt]	періодичний, повторюваний
22	likelihood (n.)	[ˈlaɪklɪhʊd]	подібність, схожість
23	hypothesis (n.)	[haɪˈpɒθɪsɪs]	гіпотеза
24	circular (adj.)	[ˈsɜːkjʊlə]	круговий, що рухається по колу
25	generalization (n.)	[ˌdʒɛnərəlaɪˈzeɪʃən]	узагальнення
26	valid (adj.)	[ˈvælɪd]	дійсний; чинний
27	unreliable (adj.)	[ˌʌnrɪˈlaɪəbl]	ненадійний
28	attainable (adj.)	[əˈteɪnəbl]	досяжний
29	unaided (adj.)	[ʌnˈeɪdɪd]	без сторонньої допомоги
30	axis (n.)	[ˈæksɪs]	вісь

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>.

Unit 3. Language Corpora

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. <…>

IV. 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?

Unit 3. Language Corpora

V. 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;

VI. 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 (*langue*).
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 (*parole*).
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.

VII. 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 untypical ……………..in literature, advertising, or jokes.

2) In a real study one might have hundreds or thousands of ……….lines, 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 the ………….norms of general language use.

4) Nevertheless, with some striking exceptions, cognitive approaches have neglected corpus data on recurrent ……..patterns as evidence of cognitive structures.

5) Such recurrent co-occurrence patterns provide ……..of typical meaning and use.

6) Intuitive data can be circular: data and theory have the same source in the linguist who both proposes a …………………and invents examples to support or refute it.

7) Corpus methods therefore differ sharply from the view, widely held since the 1960s, that native speaker ……………gives 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 an ……….finding 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 in …….written language.

VIII. 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	

IX. 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.)		

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

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

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

6) 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 dezign and software, findings and descriptions, and implications and fapplications.

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.

XII. Fill the gaps with a necessary preposition from the box below.

with (2)	to (2)	of (9)	for	within (2)	from (2)
between	before	on	about	by (2)	in (2)

1) There was corpus study long _____ computers. _____ a historical perspective, Saussure's radical uncertainty _____ the viability _____ studying parole, followed _____ Chomsky's reliance _____ introspective data, were short breaks _____ a long tradition _____ observational language study. 2) Disregard _____ quantified textual data was never accepted _____ everyone. 3) Corder emphasizes the relevance _____ frequency studies _____ language teaching, and language corpora have always been indispensable _____ studying dead languages, unwritten languages and dialects, child language acquisition, and lexicography. 4) So, _____ both philological and fieldwork traditions, corpus study goes back hundreds _____ years, _____ a broad tradition _____ rhetorical and textual analysis. 5) Early concordances were prepared _____ texts _____ cultural significance. 6) The other main reason _____ studying large text collections, which again emphasizes the central concern _____ meaning, was the attempt to produce comprehensive dictionaries. 7) Modern lexicographers use better designed corpora, but the basic approach _____ semantic analysis is not fundamentally different _____ that _____ Cruden, Ayscough, Johnson, and Murray. 8) Other impressive quantitative corpus analyses, _____ the 1890s and the 1950s, were possible only _____ significant expense and personnel, and often had precise institutional and/or educational applications.

XIII. 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, бібліо-графії, 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.

6) Завдяки стрімкому розвитку в науковому лінгвістичному словнику з'явилися дуже близькі поняття: «електронні бібліотеки», «масив текстів», «колекція текстів», «електронний архів», «повна текстова база даних», які стали предметами наукових пошуків як вітчизняних, так і зарубіжних дослідників.

7) Лінгвістичний корпус – це масив текстів, зібраних в єдину систему, сформовану за певними ознаками (мовою, жанром, часом створення, автором тощо) і забезпечених пошуковою системою. Він може містити як письмові тексти, так і транскрипти радіо- і телепередач.

8) Залежно від цілей його створення в корпус можуть входити тексти конкретною мовою, одного або кількох авторів і літературних жанрів, написані в певний історичний період і т.д. Весь масив текстів є систематизованим. Це означає, що в корпусі зафіксоване розташування кожного слова в реченні щодо інших слів, а також враховується частота його використання у цьому корпусі.

9) Першим досить великим корпусом, тексти якого зберігалися на машинному носії, був Браунівський (Brown corpus 1960 – for American English). Його розробники У. Френсіс та Г. Кучера розглядали поняття корпусу як сукупність текстів, яка вважається репрезентативною для певної мови чи діалекту, що призначена для лінгвістичного аналізу.

10) Oxford Text Archive (OTA) – це архів електронних текстів та інших літературних і мовних ресурсів, які були створені, зібрані й розподілені з метою дослідження літературних та лінгвістичних питань в університеті Оксфорд. Він вважається одним із перших архівів цифрових навчальних текстових ресурсів для збору і поширення матеріалів з усіх доступних наукових центрів.

11) ОТА продовжує співпрацю з Оксфордським університетом обслуговування з використанням електронно-обчислювальних машин OUCS, який володіє відповідними науково-дослідними проектами, що здійснюються в Оксфордському електронному науково-дослідному центрі на факультеті лінгвістики, філології та фонетики Оксфордського університету. ОТА також керує розподілом британського національного корпусу (BNC).

12) Структура архіву містить 8 розділів: 1) власне архів (OEC); 2) рекомендації щодо завантаження ресурсів; 3) новини; 4) Оксфорд (для обмеженого кола користувачів); 5) електронні мовні ресурси; 6) проекти; 7) поради та 8) проблемні питання. Особливо значущим надбанням в ОТА є власне архів (OEC), в якому зосереджений найбільший у своєму роді текстовий корпус англійської мови, що містить понад два мільярди слів.

13) *TEI texts (Text Encoding Initiative texts)*. Ініціатива кодування тексту – консорціум, що згалом розвиває і підтримує стандарт для представлення текстів

у цифрову форму. Його головним результатом є набір керівних принципів, які визначають методи кодування для машинописних текстів в основному в гуманітарних, соціальних науках і лінгвістиці. Тексти в у цьому підрозділі доступні в різних форматах для читання, завантаження або посилання.

14) *Corpora*. Колекції мовних даних, що містять тексти з різних джерел, як правило, складені для цілей лінгвістичного дослідження. Підрозділ являє собою своєрідний каталог власних накопичених корпусних даних та корпуси інших університетів із відкритим (безкоштовним) чи обмеженим доступом до інформації.

15) Через OTA проходить безліч наукових документів, розмічених відповідно до останніх вимог кодування матеріалів та текстів, а це обов'язкове зазначення таких даних, як: а) назва документа; б) автор документа (якщо відомо); в) стать автора (якщо відомо); г) тип мови (наприклад, британська чи американська англійська); д) жанр джерела; е) рік; є) дата збору матеріалу архівом; ж) доступ до матеріалів; з) статистика документа.

SPEAKING SECTION. PICTURE DESCRIPTION WORKSHEET

DISCUSSING FACTUAL INFORMATION
Where? There is a classroom/computer lab/lecture room at the university/library/school. It looks out-of-date/modern. The people in the picture belong to different/the same ethnic group/(s)/nationalit(y)ies/origin.

When? The scene takes place during the lesson/working day/excursion/staff meeting. We can/cannot see in the picture, what is the weather like outside. The scene takes place in the early morning/at midday/in the evening/at midnight.

What else can be mentioned about the setting? What can you see in the background and the foreground?

Who and what? How many? There are _____ people in the _____. They all are dressed _____, the woman is wearing a _____ blouse, the man _____ is wearing a white _____ shirt and the other man is wearing a _____ t-shirt. They are working with _____. They are _____ and they are involved in the creating of _____.

applied linguists   casually   in the middle   computers   language corpora   khaki   computer lab   three   light sleeveless

DISCUSSING CONCEPTUAL INFORMATION
1. Who could take the picture?
2. What is the possible title of the picture under consideration? How can it be related to Applied Linguistics area? What is the message of the picture?
3. Are the people in the picture posing for camera on purpose?
4. What do you think had happened before the picture was taken?
5. What will they do next after the picture having been taken?

PICTURE DESCRIPTION
Make up a list of 20 key-words covering factual and conceptual information of the picture considered. Describe the picture using the key-words, cover factual and conceptual information of the picture considered.

Unit 3. Language Corpora

PART 2

READING SECTION

I. Answer the questions below. Give your reasons. Provide examples from your own experience, if possible.

1. Are you fammiliar with any corpora of English or Ukrainian languages?
2. Will any list of extant corpora be quickly out of date?
3. What are the basic principles of corpus design?
4. What is a row corpus?
5. What can be corpora used for?

II. Comment upon the following quotes related to the topic of the unit under consideration. Do you agree or disagree to the ideas suggested? Give your reasons.

1. "...corpora [are becoming] more and more the normal tools of linguistic enquiry." *(Elena Tognini-Bonelli)*

2. "Corpora offer an ideal instrument to observe and acquire socially established form/meaning pairings." *(Silvia Bernardini)*

3. "Corpora have been likened to the invention of telescopes in the history of astronomy." *(Susan Hunston)*

4. "...a corpus by itself can do nothing at all, being nothing more than a store of used language." *(Susan Hunston)*

5. "The essence of the corpus as against the text is that you do not observe it directly; instead you use tools of indirect observation, like query languages, concordancers, collocators, parsers, and aligners..." *(John McHardy Sinclair)*

III. Study the topical vocabulary and read the article below. Find the contexts where the units under consideration are used in the text.

1	computer-assisted (adj.)	[kəmˈpjuːtər-əˈsɪstɪd]	(проведений) за допомогою еом, машинний
2	significant (adj.)	[sɪgˈnɪfɪkənt]	значимий, несучий значення (про афікс); значущий
3	inherently (adv.)	[ɪnˈhɪərəntli]	за своєю суттю, у своїй основі; від природи
4	pejorative (adj.)	[ˈpiːdʒərətɪv]	зневажливий, пейоративний
5	shift (n.)	[ʃɪft]	переміщення, перестановка, перенесення; переміна; зміна
6	idiosyncratic (adj.)	[ˌɪdɪəsɪŋˈkrætɪk]	своєрідний, унікальний характерний, особливий

7	recur (v.)	[rɪˈkɜː]	відбуватися знову
8	extant (adj.)	[ɛksˈtænt]	існуючий; збережений; який дійшов до нас
9	distinction (n.)	[dɪsˈtɪŋkʃən]	розрізнення, розпізнавання; відмінність, розходження
10	onward (adv.)	[ˈɒnwəd]	уперед, далі
11	comparative (adj.)	[kəmˈpærətɪv]	порівняльний
12	underlie (v.)	[ˌʌndəˈlaɪ]	лежати в основі (чого-н.)
13	sermon(n.)	[ˈsɜːmən]	проповідь; повчання, нотація
14	gap (n.)	[gæp]	огріх, прогалина (у знаннях)
15	criteria (n.)	[kraɪˈtɪərɪə]	множ. від criterion – критерій, мірило, ознака
16	feasible (adj.)	[ˈfiːzəbl]	можливий, імовірний
17	pattern (n.)	[ˈpætən]	зразок, приклад
18	concordance (n.)	[kənˈkɔːdəns]	відповідність, конкорданція, алфавітний покажчик слів
19	sacrifice (n.)	[ˈsækrɪfaɪs]	жертва
20	assess (v.)	[əˈsɛs]	оцінювати, давати оцінку
21	raw (unannotated) corpus (n.)	[rɔː (ˌʌnˈænəʊteɪtɪd) ˈkɔːpəs]	не анотований корпус
22	lemmatized corpus (n.)	[lemməˈtaɪzd ˈkɔːpəs]	лематизований корпус англ. lemmatize – формувати гнізда; визначати головні слова і сортувати словосполучення
23	annotated corpus (n.)	[ˈænəʊteɪtɪd ˈkɔːpəs]	анотований корпус
24	raw text (n.)	[rɔː tɛkst]	необроблений текст
25	annotated(adj.)	[ˈænəʊteɪtɪd]	з коментарями, примітками, анотований
26	occurrence (n.)	[əˈkʌrəns]	випадок; частотність; поширеність
27	comprise (v.)	[kəmˈpraɪz]	включати, містити в собі, становити; охоплювати; складатися з; входити до складу
28	assumption (n.)	[əˈsʌmpʃ(ə)n]	припущення, допущення; вихідне положення
29	preconception (n.)	[ˌpriːkənˈsɛpʃən]	заздалегідь складена думка; упереджена думка, упередження
30	parsing (n.)	[ˈpɑːzɪŋ]	синтаксичний аналіз; аналіз, розбір

Unit 3. Language Corpora

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).

IV. 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.

Unit 3. Language Corpora

V. 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**	resoling (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

VI. 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**	*undergo* 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.

VII. 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 ……………………

VIII. 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	смислові 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	

IX. 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.)		

X. 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

XI. 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 hipotheses, but corpus data are a part of natural language use, and not produced for purposes of linguistic analisis.

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

3) The LOB corpus contain British datsa from 1961; Frown and FLOB contain American and British datas 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 speaken 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 betveen 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.

XII. Open the brackets, putting the infinitive form of the verb given into the necessary tense form (active or passive).

1) The main findings which .... (to result) from the "vastly expanded empirical base" .... (to concern) the association patterns which inseparably .... (to relate) item and context. 2) The implications for language teaching .... (to be), at one level, rather evident. 3) A well-known problem for even advanced language learners .... (to be) that they may speak grammatically, yet not sound native-like, because their language use .... (to deviate) from native speaker collocational norms. 4) Syllabus designers ought to know which words .... (to use) frequently in conventionalized combinations, and which .... (to use) rarely and in special contexts.

5) The importance of collocations for language learners .... (to emphasize) in the 1930s and 1940s by H. E. Palmer and A. S. Hornby. 6) More recently corpora .... (to use) to study how learners and native speakers .... (to differ) in their use of conventionalized expressions, and a major topic .... (to be) how to represent such information in learners' dictionaries. 7) Proposals also .... (to make) about the form of a "lexical syllabus." 8) This concept .... (to discuss) in detail by Corder, and .... (to revive) in corpus work by Willis and Lewis, although corresponding teaching materials .... (to adopt) only to a limited extent. 9) The shorthand label for this area .... (to be) phraseology: the identification of typical multi-word units of language use and meaning.

XIII. 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.

6) Корпус текстів – це не просто колекція відібраних за певною методикою і представлених в електронному вигляді текстів певних сфер вживання мови, але така колекція, яка категоризована як з боку інтегральних характеристик кожного тексту (наприклад, жанрових), так і з боку специфічних характеристик різних одиниць його організації (лексеми, словоформи, морфеми тощо).

7) Мовний корпус повинен бути організованим у вигляді деякої бази даних. Це полегшує його практичне використання і науковий аналіз.

8) Існування корпусів текстів дає можливість значно розширити і автоматизувати аналіз мовного матеріалу, який є найважливішою базою будь-якого лінгвістичного дослідження. Чим більше матеріалів аналізується, тим вища значущість висновків і рівень їх достовірності.

9) Корпус формується з реальних уривків писемного або усного мовлення, не передбачаючи модифікації мовленнєвої дійсності, що перетворює його на категорію емпіричну і дозволяє розглядати фактичний корпусний матеріал як емпіричну базу лінгвістичного дослідження.

10) Сучасне корпусне мовознавство має у своєму дослідницькому арсеналі значну кількість корпусів різних типів і розмірів. Таке розмаїття існуючих корпусів визначається широким спектром дослідницьких і прикладних завдань, для розв'язання яких вони створюються, та особливостями мовного матеріалу, покладеного в основу.

11) Лінгвісти і програмісти, що зайняті створенням національних корпусів, головним чином вирішують питання матеріального та технічного оснащення, для того щоб матеріали корпусу були доступні для роботи і задоволення запитів користувачів.

12) Надзвичайно ефективними для визначення рівня адекватності перекладу, удосконалення систем машинного перекладу і гармонізації термінології є паралельні корпуси, які складаються з оригінальних та перекладних текстів, що вирівняні по реченнях.

13) Корпуси усного мовлення включають транскрибовані версії реальних усних комунікативних ситуацій. Корпуси усного мовлення – це спеціальні колекції ретельно відібраних текстових уривків (слів, фраз, речень), вимовлених численними мовцями за різних акустичних умов.

14) Якщо необхідні для створення корпусу тексти існують лише у друкованому вигляді, їх конвертують в електронну форму за допомогою сканування. Це можуть бути рідкісні чи старі видання, що не мають цифрових версій. Звичайно, сканування потребує гарного технічного та програмного забезпечення.

15) Зазвичай, корпуси укладаються з метою можливості їх багаторазового використання якісно різними користувачами, тому, їхня структура та програмне забезпечення повинні бути максимально уніфіковані. Загальноприйняті формати представлення даних дозволяють у багатьох випадках використовувати єдине програмне забезпечення й обмінюватися корпусними даними. Стандартизація відносно корпусів, сумісність типів даних важливі також і для можливості порівняння різних корпусів

SPEAKING SECTION.
PICTURE DESCRIPTION WORKSHEET
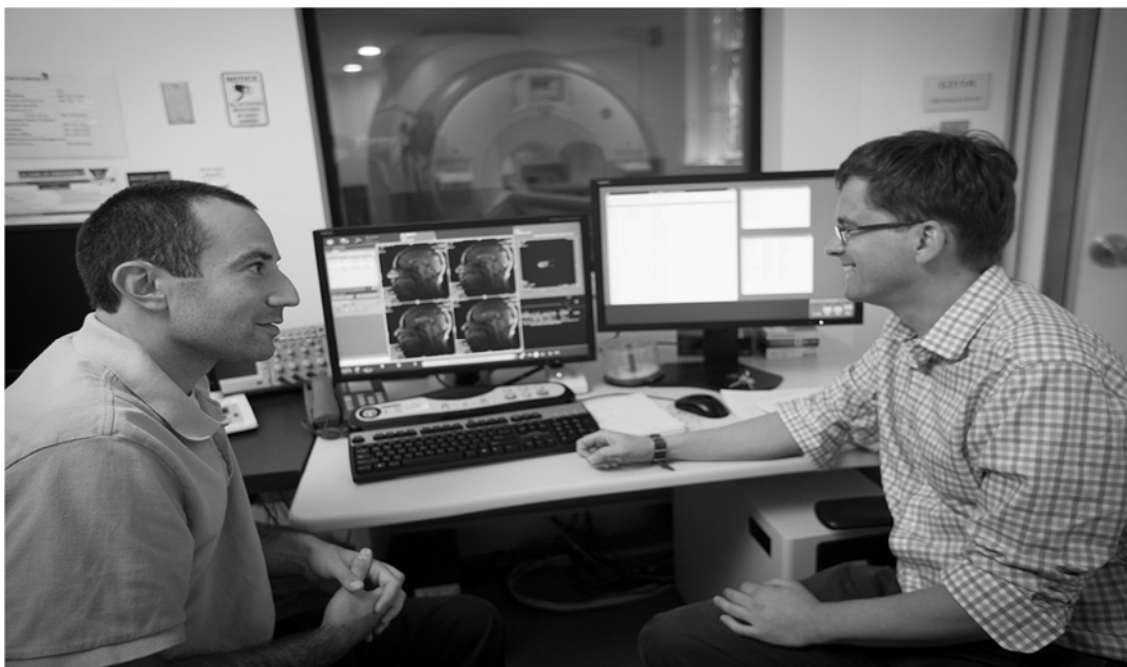
DISCUSSING FACTUAL INFORMATION
Where? There is a classroom/conference hall/lecture room/laboratory at the university/library/school/enterprise. It looks out-of-date/modern and well/poorly equipped. The people in the picture belong to different/the same ethnic group/(s)/nationalit(y)ies/ origin, they represent the same/different age and gender groups.

When? The scene takes place during the lesson/match/discussion session/job-interview/procedure of carrying out an experiment/excursion/staff meeting. We can/ cannot see in the picture, what is the weather like outside. The scene takes place in the early morning/at midday/in the evening/at midnight.

What else can be mentioned about the setting? What can you see in the background and the foreground?

Who and what? How many? There are ___ men in the _____ office. One of them is wearing_____, the other one is _____neither glasses nor _____. The men are obviously in their _____thirties. Both men are _____. The people in the picture are discussing_____ issues. There are two personal _____on the table. The men in the picture are _____. One of them is wearing a blue _____, the other one is wearing a checked_____. The discussion seems rather _____.

mid	two	glasses	sitting	wearing	a watch	computers	T-shirt
	laboratory	neurolinguistics	smiling	shirt	positive		

DISCUSSING CONCEPTUAL INFORMATION
1. Who could take the picture?
2. What is the possible title of the picture under consideration? How can it be related to Applied Linguistics area? What is the message of the picture?
3. Are the people in the picture posing for camera on purpose?
5. What do you think had happened before the picture was taken?
6. What will they do next after the picture having been taken?

PICTURE DESCRIPTION
Make up a list of 20 key-words covering factual and conceptual information of the picture considered. Describe the picture using the key-words, cover factual and conceptual information of the picture considered.

PART 3

READING SECTION

I. Answer the questions below. Give your reasons. Provide examples from your own experience, if possible.

1. Do you agree that a word may be considered as the main unit of language? Give your reasons.
2. Why do particular words appear to be more frequent than other? How can word frequency be measured?
3. Why are short words more frequently used than longer ones? How is that related to the issue of economy of effort for the speaker?
4. Should genres be taken into consideration when forming and analyzing word frequency lists? Give your reasons.
5. Why and to what extent can word frequency lists be limited? Give your reasons.

II. Comment upon the following quotes related to the topic of the unit under consideration. Do you agree or disagree to the ideas suggested? Give your reasons.

1. "Collocations are actual words in habitual company. A word in a usual collocation stares you in the face just as it is. Colligations cannot be of words as such. Colligations of grammatical categories related in a grammatical structure do not necessarily follow word divisions or even sub-divisions of words." *(John Rupert Firth)*

2. So difficult it is to show the various meanings and imperfections of words when we have nothing else but words to do it with. *(John Locke)*

3. A word is not a crystal, transparent and unchanged, it is the skin of a living thought and may vary greatly in color and content according to the circumstances and the time in which it is used. *(Oliver Wendell Holmes)*

4. "It's strange how a word, a phrase, a sentence, can feel like a blow to the head." *(Veronica Roth)*

5. "Prose consists less and less of words chosen for the sake of their meaning, and more and more of phrases tacked together like the sections of a prefabricated hen-house." *(George Orwell)*

III. Study the topical vocabulary and read the article below. Find the contexts where the units under consideration are used in the text.

1	salience (n.)	[ˈseɪljəns]	характерна особливість
2	figure (n.)	[ˈfɪgə]	цифра,число
3	pervasive (adj.)	[pɜːˈveɪsɪv]	проникаючий, всепроникний
4	rough (adj.)	[rʌf]	приблизний, нерівний, грубий

5	predictable (adj.)	[prɪˈdɪktəbl]	передбачуваний
6	descending (adj.)	[dɪˈsɛndɪŋ]	спадний
7	node (n.)	[nəʊd]	вузол (*іноді контекст.* вершина)
8	ascending (adj.)	[əˈsɛndɪŋ]	висхідний
9	remarkably (adv.)	[rɪˈmɑːkəbli]	надзвичайно, дивовижно
10	respective (adj.)	[rɪsˈpɛktɪv]	відповідний
11	bare (adj.)	[beə]	порожній, незаповнений
12	lexicalize (v.)	[ˈleksɪkəlaɪz]	лексикалізувати
13	disperse (v.)	[dɪsˈpɜːs]	розпорошувати, розсіювати; розходитися
14	positional (adj.)	[pəˈzɪʃənl]	позиційний
15	persistent (adj.)	[pəˈsɪstənt]	наполегливий
16	residual (adj.)	[rɪˈzɪdjʊəl]	залишковий
17	denotation (n.)	[ˌdiːnəʊˈteɪʃən]	позначення, денотація
18	average (adj.)	[ˈævərɪdʒ]	середньостатистичний
19	span (n.)	[spæn]	інтервал, коротка відстань між елементами; (*іноді контекст.* група, комплект)
20	randomly (adv.)	[ˈrændəmli]	випадковим чином
21	allegation (n.)	[ˌælɪˈgeɪʃ(ə)n]	твердження, заява (*іноді контекст.* голослівне твердження, не підтверджене доказами)
22	approximate (adj.)	[əˈprɒksɪmɪt]	приблизний
23	word-form (n.)	[wɜːd-fɔːm]	словоформа
24	set (n.)	[sɛt]	набір, комплекс
25	speculation (n.)	[ˌspɛkjʊˈleɪʃən]	міркування
26	conversely (adv.)	[ˈkɒnvɜːsli]	навпаки, протилежно, всупереч
27	boundary (n.)	[ˈbaʊndəri]	межа, розподільна лінія, кордон
28	distribution (n.)	[ˌdɪstrɪˈbjuːʃən]	розподіл, розповсюдження, дистрибуція
29	replicable (adj.)	[ˈreplɪkəbl]	повторюваний, відтворюваний
30	disambiguation (n.)	[ˌdɪsæmˈbɪgjueɪʃən]	усунення неоднозначності

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), Johannson 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).

Unit 3. Language Corpora

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 bottomof 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.

IV. 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?

V. 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;

VI. 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 *persistent* 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. |

VII. 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 a …………..shows 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).

VIII. 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	

IX. 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.)		

X. 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

XI. 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 languauge 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 specifical collocate.

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.

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

to (5)   about   of (8)   in (3)   with (2)
between (2)   from (2)   on (2)   for   by

A central aim is to make more explicit the semantic and pragmatic features .... multi-word units. .... example, *enduring, persistent*, and *haunting* are all rough synonyms, but they co-occur .... nouns .... different semantic fields. Characteristic combinations .... modifier plus noun include: enduring peace; haunting music; persistent headaches. We can also generalize .... semantic preferences. .... adjective-noun constructions, *persistent* is often used .... medical conditions, and *haunting* is usually used .... music, words, and images. Louw (1993) was the first important article .... how such attitudes are conveyed.

A model .... extended lexical units proposed .... Sinclair (1998) combines these increasingly abstract relations: collocation, colligation, semantic, and discourse prosody. We can also specify: strength .... attraction .... node and collocates;

position .... node and collocate, variable or fixed; and distribution, wide occurrence .... general English or .... broad varieties, or restricted .... specialized text-types.

In summary, work .... extended lexical units has redrawn the lexis– grammar boundary. Only a few units are fixed phrases; most are recurrent combinations .... grammatical constructions .... words .... restricted lexical fields, but .... considerable lexical variation. Relations correspond .... the classic distinctions .... syntax (how language units relate .... one another), semantics (how linguistic signs relate .... the external world), and pragmatics (how linguistic signs relate .... their users).

XIII. 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 малоймовірне.

5) Основні процедури, які доступні досліднику при здійсненні аналізу корпусу текстів, містять: а) пошук заданого слова, словосполучення в корпусі; б) висновок результатів пошуку з урахуванням оточення в окремому полі; в) підрахунок кількості прикладів вживання слова в корпусі; г) сортування результатів пошуку за необхідними параметрами. Всі дані процедури швидко і точно виконуються за допомогою комп'ютерної програми складання конкордансу (пошуку відповідностей).

6) Перед початком роботи з програмою на головній сторінці контролера інструментів користувачу пропонуються три кнопки основних інструментів і декілька клавіш для уточнення налаштувань. Клавіша "Concord" укладає конкордаси, "KeyWords" знаходить у текстах ключові слова, а "WordList" створює списки слів у тексті або колекції текстів. Для вдалого використання програмного забезпечення спочатку необхідно завантажити чи створити власний корпус текстів.

7) Після впорядкування та оформлення матеріалу за допомогою "WordSmith" можна створювати список частотності вживання слів у одному тексті чи в корпусі текстів. За допомогою порівняння власного корпусу даних із BNC отримуємо список ключових слів. Поряд із кожним ключовим словом розміщені різні цифри, які містять інформацію про те, як часто вживається кожне слово у вихідному тексті (текстах) і наскільки ця частотність відрізняється від частотності його вживання у референтному корпусі.

8) Одним із основних підходів до аналізу мовних даних у корпусній лінгвістиці є конкорданси. Загалом *конкорданс* – це спеціалізована лінгвістична прикладна програма, за допомогою якої здійснюється автоматична вибірка заданих мовних одиниць з електронних текстів, проводиться дослідження корпусу за обраним словом, словосполученням чи фразою.

9) Найбільш репрезентативний та швидкий спосіб обробки інформації за допомогою "WordSmith" – це укладання конкордансу. Конкорданс – це список усіх уживань слова, перед і після якого є слово-розділювач, такий як знак пунктуації, пробіл тощо. Фундаментальні конкорданси становлять скарбницю знань про вживання в текстах тієї чи іншої мовної одиниці, і дослідницьких, що підпорядковані розв'язанню конкретного завдання.

10) Дослідивши великий обсяг інформації, що міститься в корпусі, можна отримати повне уявлення щодо досліджуваного явища і певної мови в цілому. Велика кількість створених корпусів дає змогу отримати дані аналізу писемного чи спонтанного мовлення, мовлення певної вікової, гендерної, соціальної чи етнічної групи, інформацію про особливості певного діалекту.

11) Розвиток інформаційного суспільства та суспільства знань спричинив бурхливий прогрес у галузі комп'ютерних технологій опрацювання природної мови, який поставив нові завдання перед лінгвістикою щодо вивчення різних властивостей мовної системи. Завдання, які стоять у цій царині перед усіма ділянками мовознавства, тепер потребують вивчення різнопланових мовних явищ, структур, одиниць, відношень тощо не на окремих, хоча й показових, прикладах, а в їх повному, репрезентативному обсязі.

12) Залежно від технічних можливостей конкорданс може надавати інформацію про частотність вживання і сполучення тієї або іншої мовної одиниці, а також дає змогу звертатися до конкретного тексту, в якому був знайдений приклад, і демонструє слова, словосполучення чи фрази в центрі комп'ютерного екрану, разом зі словами, що стоять перед і після них, ліворуч і праворуч.

13) Вибране слово, що видається в центрі екрану, відоме як "вузлове". Лінії конкордансу видають інформацію хаотично, але її можна сортувати, щоб вона надходила за алфавітним порядком, чи групами, які вибрані та організовані для ілюстрації певної особливої поведінки заданого слова чи фрази.

14) Термін "типовий" (typical) використовується в корпусній лінгвістиці щодо найбільш характерних випадків як дистрибуції мовних одиниць, так і значень мовних одиниць. Прототиповими вважають такі мовні засоби, частотність яких за інтуіцією носіїв чи користувачів мови мала б бути високою, але, як

показують дослідження з корпусної лінгвістики, вони не є настільки частотними як передбачалось.

15) Поняття прототиповості було введене науковцями Дереком Дейві та Девідом Крісталом. Термін "прототиповий" (prototypical) відображає розбіжності між типовими випадками вживань мовної одиниці і типовими значеннями та даними про частотність. Розбіжності між типовим й прототиповим виявляються на основі різних жанрів, регістрів, соціолінгвістичного варіювання мови.

SPEAKING SECTION. PICTURE DESCRIPTION WORKSHEET

DISCUSSING FACTUAL INFORMATION

Where? There is a classroom/hall/lecture room at the university/library/school. It looks out-of-date/modern. The people in the picture belong to different/the same ethnic group/(s)/nationalit(y)ies/origin.

When? The scene takes place during the lesson/match/excursion/staff meeting. We can/cannot see in the picture, what is the weather like outside. The scene takes place in the early morning/at midday/in the evening/at midnight.

What else can be mentioned about the setting? What can you see in the background and the foreground?

Who and what? How many? There are ___ people in the _____, a _____ and his_____. The teacher has a ___ style and is wearing a suit jacket, a ___ and light ___. He is wearing ___ as well. The students are ____and have a ___ style. They are

___notes and ___ to the teacher. The people in the picture are discussing_____ linguistics issues. ___analysis is a tough issue for clear_____.

Fashion-conscious	classroom	thirteen	applied	casual	classic	taking
teacher	comprehension	trousers	students	tie	glasses	listening	discourse

DISCUSSING CONCEPTUAL INFORMATION
1. Who could take the picture?
2. What is the possible title of the picture under consideration? How can it be related to Applied Linguistics area? What is the message of the picture?
3. Are the people in the picture posing for camera on purpose?
5. What do you think had happened before the picture was taken?
6. What will they do next after the picture having been taken?

PICTURE DESCRIPTION
Make up a list of 20 key-words covering factual and conceptual information of the picture considered. Describe the picture using the key-words, cover factual and conceptual information of the picture considered.

LISTENING (AUDIOVISUAL COMPREHENSION) SECTION

CORPORA AND SPOKEN LANGUAGE

I. Watch the talk by Prof. McCarthy and answer the following questions. Justify your answer relying upon the facts form the video material. (https://www.youtube.com/watch?v=UKQcVE9d67s&t=371s)

1. Why were spoken corpora smaller and difficult to build in the early 1990s?
2. What is one of the standard techniques that corpus linguists use to count frequency?
3. What is the evidence that the most common vocabulary of writing and speaking is different one compared with the other?
4. Why does Prof. McCarthy confront the tendency to consider the expression *you know* as an inappropriate to be taught?
5. Why are the words like *well* and *right* incredibly frequent in the spoken corpus and may be absent in the frequency list of the written corpus?

II. Indicate whether the statements are true or false, correct the false ones.

1. Small sized recording equipment and modern software made spoken corpora easier to collect. _____

2. English Profile is a very big project aimed to build language corpora for the Common European Framework of Reference. _____

3. Prof. McCarthy and Dr. Paula Buttery compared two hundred most frequent words in the British national corpus spoken segments with a similar top two hundred words in the British national corpus written segment. _____

4. The aim of Prof. McCarthy and Ronald Hart's research was to compare the 50 most frequent words in the CANCODE corpus and in a same sized corpus of general written English. _____

5. According to the results of the top 50 comparison there is a number of words which appear only in the spoken top 50. _____

6. *Right* is used to start the conversation or to show that you are not interested in the topic. _____

7. *Well* can express the signal that the person is going to change the direction of the projected discourse. _____

8. The word *just* can make the statement friendlier, softer and less direct. _____

9. These little words that we get from spoken corpora show that there is a thick skill which can be called the skill of interaction. _____

10. Words like *right*, *well*, *know* and *think* prove that we need to pronounce things correctly and talk about our experiences. _____

III. Fill in the gaps with appropriate words and phrases according to Prof. McCarthy's talk.

My particular area of interest is conversation and _____, and there are a couple of reasons for this. Firstly, conversation is the most _____ way that we use language in every day of our lives. We talk to our friends, our colleagues, our neighbors, family and _____ an awful lot of language every day, much more than we write. Secondly, a good reason for collecting _____ is that it is actually quite difficult to be _____ about how we speak; it's much easier to be objective about how we write. When we write we can _____, we can change things, we can cut and paste, we can do all sorts of things, we can _____ on what we're doing. When we speak we have no such opportunities for reflection, we don't have time to sit back and _____ how we speak. Therefore, it's not surprising that a lot of the ways in which we _____ language are based on _____ that come from the written language.

WRITING SECTION

Choose one of the topics below to write a comparison essay. Before writing study the tips.

1) Corpora of spoken language vs. corpora of written language.
2) English vs. Ukrainian corpora.
3) Raw vs. annotated corpora.

<u>Tips for writing</u>
    1. In a comparison essay (or a Compare and Contrast essay) you should critically analyze any two subjects, finding and pointing out their similarities and/or differences.
    2. The <u>essay format</u> of a comparative essay is just like that of any other essay, and it should have an introduction, body paragraphs, and a conclusion.
    3. The <u>introduction</u> paragraph is mainly about giving the reader a glimpse of the essay in terms of its content and expectation of the reader. It should therefore be simple and clear. It should also have a thesis statement that highlights the purpose of the essay and what is going to be discussed not forgetting a simple conclusion.
    4. In the <u>body paragraphs</u> you are just transferring what you had while identifying the things about the subject you are going to compare and organizing then in the preferred structure.
    5. The <u>conclusion</u> is mean to give the reader an overview of the essay. The conclusion should be a product of the identified similarities and differences and not of personal preferences.
    6. The topic sentence should introduce the reader to what the paragraph handles. Discussion of the aspect is done in the middle of the paragraph. The last part of the paragraph should carry a low-level conclusion about the aspect discussed in the paragraph.

SELF-REFLECTION SHEET ON THE UNIT 3 MATERIALS:

1) THE FOLLOWING 10 KEY-CONCEPTS OF THE UNIT CAN BE SINGLED OUT:
_____
_____

2) UNIT REVIEW: THE UNIT GENERALLY RUNS ABOUT:

_____
_____
_____
_____

3) WHAT ARE THE ADVANTAGES OF THE UNIT UNDER CONSIDERATION?
_____
_____
_____

4) WHAT ARE THE DISADVANTAGES OF THE UNIT UNDER CONSIDERATION?
_____
_____
_____

5) WHICH TASKS DO YOU CONSIDER TO BE MOST USEFUL IN THE UNIT?
_____
_____
_____

6) TAKE A LOOK AT THE TITLE OF THE NEXT UNIT. WHAT DO YOU EXPECT TO LEARN WHEN YOU START COVERING THE MATERIALS OF THE UNIT?
_____
_____
_____

UNIT 4
DISCOURSE ANALYSIS

PART 1

READING SECTION

I. Answer the questions below. Give your reasons. Provide examples from your own experience, if possible.

1. Why is the term "discourse" considered to be polysemantic?
2. How can the notion of discourse be defined?
3. What causes the interdisciplinary nature of discourse analysis?
4. Why does discourse analysis have such a very important part to play in the work that applied linguistics does?
5. What is the role of non-verbal components in oral and written communication?

II. Comment upon the following quotes related to the topic of the unit under consideration. Do you agree or disagree to the ideas suggested? Give your reasons.

1. "Every discourse, even a poetic or oracular sentence, carries with it a system of rules for producing analogous things and thus an outline of methodology." *(Jacques Derrida)*

2. "There is nothing in discourse that is not to be found in a sentence." *(Roland Barthes)*

3. "We are all full of discourses that we only half understand and half mean." *(Rae Armantrout)*

4. "Where wise actions are the fruit of life, wise discourse is the pollination." *(Bryant McGill)*

5. "The altitude of any relation is possible through discourse." *(Unarine Ramaru)*

III. Study the topical vocabulary and read the article below. Find the contexts where the units under consideration are used in the text.

1	discourse (n.)	[dɪsˈkɔːs]	дискурс; висловлення
2	unconsciously (adv.)	[ʌnˈkɒnʃəsli]	несвідомо
3	patterning (n.)	[ˈpætənɪŋ]	структурування, компонування
4	circumstance(n.)	[ˈsɜːkəmstəns]	обставина; умова; подробиця, деталь
5	mundane (adj.)	[ˈmʌndeɪn]	світський
6	deliberately (adv.)	[dɪˈlɪbərɪtli]	обдумано; зважено
7	raison d'être (french).	[ˌreɪzɵ̃ ˈdɛtrə]	підстава; сенс існування

8	multi-disciplinary	[ˌmʌltɪˈdɪsɪˌplɪnəri]	багатогалузевий
9	boundary (n.)	[ˈbaʊndəri]	кордон; межа
10	divine (adj.)	[dɪˈvaɪn]	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.)	[ˌɪntərˈækʃən]	взаємодія; вплив один на одного
17	conviction (n.)	[kənˈvɪkʃən]	переконання, погляди
18	affiliation (n.)	[əˌfɪlɪˈeɪʃ(ə)n]	простежування джерел, походження; встановлення зв'язку (з чим-небудь)
19	palpable (adj.)	[ˈpælpəbl]	очевидний, явний
20	clause (n.)	[klɔːz]	речення (частина складного речення)
21	prosodic (adj.)	[prəˈsɒdik]	просодичний, інтонаційний
22	evaluate (v.)	[ɪˈvæljʊeɪt]	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.)	[ˌɪnkəˌmɛnʃərəˈbɪlɪti]	несумірність; непорівнянність
27	complementary (adj.)	[ˌkɒmplɪˈmɛntəri]	комплементарний, взаємодоповнюючий
28	embedded (adj.)	[ɪmˈbɛdɪd]	включений, вкладений; вбудований
29	instantiate (v.)	[ɪnˈstænʃieɪt]	проілюструвати конкретним прикладом
30	subsequently (adv.)	[ˈsʌbsɪkwəntli]	згодом, потім, пізніше

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 (Schдffner & 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."

Unit 4. Discourse Analysis

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.

Unit 4. Discourse Analysis

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":

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).

IV. 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 concepts of language are involved in the more detailed definition of discourse analysis?
8) What kind of person is Linguist 2?
9) What is Linguist 3 mainly interested in?
10) How does Linguist 4 see the text, the even, and the drama?

V. 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 |

VI. 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 *performance* being enacted, the *drama* 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 *event*	**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 *framework of knowledge and power* 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.

VII. 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.

VIII. 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	

IX. 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.)		

X. 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

XI. 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 came to the fore, and this is where issues to do with language and linguistics representation comes into phocus."

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.

XII. Fill the gaps with a necessary article *(a/an, the, # zero article).*

<…> In the context of the spoken language skills, ……. importance of strategic competence in …… learner's negotiation of meaning is readily apparent: their strategies for coping with ………potential or actual breakdown need to be developed, and this can be facilitated, though not without difficulty (Hedge, 2000), through appropriate design and management of …… communication tasks. In teaching *written language skills*, recognition of the interactional and socially situated nature of ……… task focuses attention on contextualization: in the case of the reading skill, contextualization of the reader, their purpose in reading ….. particular text, and what they bring to it in terms of background knowledge and expectations; in …… case of the writing skill, contextualization of ……. writer, their purpose in writing, and the way in which they construct their reader in terms of social role (e.g., membership of …… particular discourse community), reading purpose, background knowledge, and expectations. Both reading and writing in …… second language are complex skills, capable of causing great difficulties to learners: writing especially, because the output is ….. product (text) that, in addition to being satisfactory in terms of …….content, needs to meet reader expectations in terms of register and generic features (overall organization, metadiscourse features, use of …..cohesion, etc.), and also attain ……. adequate standard of linguistic accuracy. <…>

XIII. 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 багатозначнім, 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.

6) Дискурс-аналіз – цілісний комплекс, що включає в себе: філософські передумови, які стосуються ролі мови у соціальних структурах світу, теоретичні моделі дослідження певної проблеми, методологію того, як вибрати підхід до дослідження проблеми, і специфічні прийоми аналізу цієї проблеми.

7) Мета дискурс-аналізу полягає у тому, щоб окреслити процеси структурування соціальної реальності, під час яких відбувається закріплення за тими чи іншими знаками певних значень, встановлюються, відтворюються і змінюються відносини ідентичності.

8) Дискурсивна психологія – розглядає дискурс як ситуативне використання мови та мовлення у повсякденній практиці спілкування.

9) Дослідження дискурсу в комунікативно орієнтованому напрямі дає можливість вивчити особливості відображення людиною навколишнього світу, соціального буття, проаналізувати реальний вияв взаємодії когнітивно- номінативної та комунікативної функцій мови в конкретних соціокультурних ситуаціях.

10) 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.

11) Дискурс існує у двох станах: статичному та динамічному. У статичному стані дискурс тлумачать як результат, продукт мовленнєво- мисленнєвої діяльності, у динамічному – як процес його творення, сприйняття й розуміння.

12) На основі типу носія інформації виділяють такі різновиди сучасного дискурсу як друкований дискурс, радіопередача, телефонна розмова, спілкування за допомогою автовідповідача чи месенджера, листування електронною поштою.

13) У структурі дискурсу розрізняють – макроструктуру (глобальну структуру) і мікроструктуру (локальну структуру). Під макроструктурою дискурсу розуміють його членування на великі складові. Мікроструктура дискурсу характеризується членуванням дискурсу на мінімальні складові, що відносять до дискурсивного рівня.

14) Засоби вербально-семіотичного рівню є базовими для визначення параметрів інтерпретаційних рівнів (змістовно-смислового і мотиваційно-прагматичного), які, у свою чергу, піддаються поясненню на заключній, когнітивно-концептуальній, ланці аналізу.

15) Одним із центральних положень у визначенні дискурсу є його протиставлення як процесу системі мови, аналіз конкретного висловлення як індивідуального перетворення мови на дискурс.

SPEAKING SECTION.
PICTURE DESCRIPTION WORKSHEET.

DISCUSSING FACTUAL INFORMATION

Where? There is a classroom/conference room/lecture room at the university/ library/school. It looks out-of-date/modern. The people in the picture belong to different/the same ethnic group/(s)/nationalit(y)ies/origin.

When? The scene takes place during the lesson/conference/excursion/staff meeting. We can/cannot see in the picture, what is the weather like outside. The scene takes place in the early morning/at midday/in the evening/at midnight.

What else can be mentioned about the setting? What can you see in the background and the foreground?

Who and what? How many? There are _____ people in the _____, a _____, a presenter and _____ of the conference. The head of the section committee is wearing a _____. She is wearing _____ as well. She is _____ the topic of the next presentation. The presenter is wearing a _____ dress. She has a _____ to provide her presentation with _____. The participants have a _____ style. They are _____ to the head of the section committee. Some of the participants

are _____ notes. The conference is _____ to the modern studies in _____. The people in the picture will be listening to the _____ on the topical issues of _____.

listening	presentation	laptop	light dress	glasses
dark	head of the section	committee	applied linguistics	
figures	conference room	taking announcing	participants	
discourse analysis	devoted	about forty	casual	

DISCUSSING CONCEPTUAL INFORMATION
1. Who could take the picture?
2. What is the possible title of the picture under consideration? What is the message of the picture?
3. Are the people in the picture posing for camera on purpose?
5. What do you think had happened before the picture was taken?
6. What will they do next after the picture having been taken?

PICTURE DESCRIPTION
   Make up a list of 20 key-words covering factual and conceptual information of the picture considered. Describe the picture using the key-words, cover factual and conceptual information of the picture considered.

PART 2

 READING SECTION

I Answer the questions below. Give your reasons. Provide examples from your own experience, if possible.

1. Discuss some of the social norms that guide conversational interaction.
2. Identify some of the ways in which language varies based on the cultural context.
3. Explain the role that accommodation and code-switching play in communication.
4. Discuss cultural bias in relation to specific cultural identities.
5. What can you do now to be more aware of how verbal communication can reinforce cultural biases?

II Comment upon the following quotes related to the topic of the unit under consideration. Do you agree or disagree to the ideas suggested? Give your reasons.

1. "Pragmatics studies the factors that govern our choice of language in social interaction and the effects of our choice on others." *(David Crystal)*

2. "Because language and society are so closely linked, it is possible, in some cases, to encourage social change by directing attention towards linguistic reflections of aspects of society that one would like to see altered." *(Peter Trudgill)*

3. "Language is a process of free creation; its laws and principles are fixed, but the manner in which the principles of generation are used is free and infinitely varied. Even the interpretation and use of words involves a process of free creation." *(Noam Chomsky)*

4. "Language furnishes the best proof that a law accepted by a community is a thing that is tolerated and not a rule to which all freely consent." *(Ferdinand De Saussure)*

5. "There are four ways, and only four ways, in which we have contact with the world. We are evaluated and classified by these four contacts: what we do, how we look, what we say, and how we say it." *(Dale Carnegie)*

III. Study the topical vocabulary and read the article below. Find the contexts where the units under consideration are used in the text.

1	hierarchy (n.)	[ˈhaɪərɑ:ki]	Ієрархія
2	interactionality (n.)	[ˌɪntərəkʃəˈnælɪtɪ]	здатність до взаємодії
3	resort (n.)	[rɪˈzɔ:t]	інстанція; звернення; надія
4	ultimately (adv.)	[ˈʌltɪmɪtli]	в кінцевому рахунку, в кінцевій позиції; зрештою
5	spontaneous (adj.)	[spɒnˈteɪnɪəs]	спонтанний, безпосередній; неви-мушений

6	decline (v.)	[dɪˈklaɪn]	зменшуватися, знижуватися, занепадати
7	self-contained (adj.)	[sɛlf-kənˈteɪnd]	окремий, самостійний; недружній, нетовариський
8	sociolinguistics (n.)	[ˌsəʊsiəʊlɪŋˈgwɪstɪks]	соціолінгвістика
9	accomplishment (n.)	[əˈkɒmplɪʃmənt]	виконання, досягнення
10	mutual (adj.)	[ˈmjuːtjʊəl]	взаємний, обопільний
11	maintenance (n.)	[ˈmeɪntənəns]	підтимання, утримання; захист
12	rapport (n.)	[ræˈpɔː]	взаєморозуміння, контакт
13	breakdown (n.)	[ˈbreɪkˌdaʊn]	невдача (в комунікативному процесі); розподіл, розбиття (елементів цілого)
14	trajectory (n.)	[ˈtrædʒɪktəri]	траєкторія
15	negotiate (v.)	[nɪˈgəʊʃɪeɪt]	вести перемовини
16	contribution (n.)	[ˌkɒntrɪˈbjuːʃən]	внесок; співробітництво
17	reveal (v.)	[rɪˈviːl]	відкривати, виявляти, з'ясовувати
18	consequence (n.)	[ˈkɒnsɪkwəns]	наслідок
19	capture (v.)	[ˈkæptʃə]	захопити, привернути
20	establishment (n.)	[ɪsˈtæblɪʃmənt]	установлення, налагодження
21	underpin (v.)	[ˌʌndəˈpɪn]	підкріпляти, бути базисом
22	latter (adj.)	[ˈlætə]	останній
23	contrive (v.)	[kənˈtraɪv]	вигадувати, вимислювати; винаходити
24	concern (n.)	[kənˈsɜːn]	відношення; стурбованість, занепокоєння
25	cohesion (n.)	[kəʊˈhiːʒən]	єдність, узгодженість
26	coherence (n.)	[kəʊˈhɪərəns]	послідовність; злагодженість
27	simultaneous (adj.)	[ˌsɪməlˈteɪnɪəs]	одночасний, одномоментний
28	transitivity (n.)	[ˌtrænzəˈtɪvəti]	перехідність (грам. категорія)
29	indeterminate (adj.)	[ˌɪndɪˈtɜːmɪnɪt]	невизначений, невирішений; сумнівний
30	predisposition (n.)	[ˌpriːdɪspəˈzɪʃən]	Схильність

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	- pragmatics (including speech act theory and - politeness theory); - conversation analysis;
Contexts and cultures	- ethnography of communication; - interactional sociolinguistics;
Functions and structures	- systemic-functional linguistics (SFL); - Birmingham school discourse analysis; - text-linguistics;
Power and politics	- 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 exical 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." <…>

IV. 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 puspose of IRF analysis? How do you understand the notion of the *accomplished interaction*?

V. 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; |

VI. 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: *Initiating move* (from the teacher), *Responding move* (from the pupil), *Feedback move* (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.

VII. 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) Discourse ………..is sometimes defined as the analysis of language 'beyond the sentence'.

2) Some discourse analysts consider the larger discourse ………..in 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 the ………of 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 properly ……………the text you are reading.

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

8) When speakers have different ……………..about 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 its …………….meaning every time it's used.

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

VIII. 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	смисловий потенціал висловлення	

IX. 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.)		

X. 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		

XI. 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 proccess, joint action in the making, and in conseqeunce most dificult to capture and analize without losing sight by its esence.

XII. Open the brackets, putting the infinitive form of the verb given into the necessary tense form (if needed).

1) Since the beginnings of communicative language teaching (CLT) and especially the teaching of English for specific (academic and professional) purposes, second language teaching and learning ………. (to come) to be understood increasingly in terms of discourse and the evaluation of students' communicative competence". 2) Hymes' concept of communicative competence ………. (to appropriate) for language teaching purposes in a series of evolutionary reformulations so as to include grammatical, pragmatic, sociolinguistic, discourse. 3) Defining the goals of language teaching in terms of communicative competence .... (to lead)naturally to "an integrative view wherein the over-arching perspective of language as discourse will affect every part of the syllabus, ………. (to include) any conventional system components and functional/ speech act components, however they are treated, whether as a series of layers of language, or as realizations within general specifications of discourse strategies". 4) Within such a perspective, learner needs, syllabus aims and content, and task goals and procedures ………. (to specify) primarily in discourse terms; materials (text or audio/video) are selected and ……… to meet criteria of communicative authenticity. (to present) 5) Tests …………… (to construct) to recreate as closely as possible the conditions under which language will be used in real communication in the defined target situation.

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

1) Дискурсивний аналіз тексту – це аналіз дискурсу як результату (текст) і про-цесу 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 водночас 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).

6) Сучасна лінгвістична прагматика тісно пов'язана з філософією мови, соціолінгвістикою, психолінгвістикою, теорією мовленнєвих актів, функціональним синтаксисом, лінгвістикою тексту, дискурсивним аналізом, конверсаційним аналізом, етнографією мовлення, когнітивною лінгвістикою, дослідженнями у сфері штучного інтелекту, комунікативною лінгвістикою та деякими іншими напрямами сучасної науки про мову.

7) Дослідники говорять про існування *соціопрагматики,* яка вивчає соціальні й культурні умови використання мовних засобів для виконання певних функцій, *контрастивної прагматики,* яка досліджує вияви прагматичного чинника, насамперед функції і побудову мовленнєвих актів, у різних мовах, *міжкультурної прагматики,* предметом дослідження якої є подібність, відмінність і варіативність у вираженні прагматичних значень у різних культурах.

8) У концепціях інших дослідників згадані напрями прагматики розглядаються як складники саме лінгвістичної прагматики. Зважаючи на таку невизначеність, мовознавці жартують, що лінгвістична прагматика нагадує кошик для сміття: що не визнається семантикою і не охоплюється синтаксисом, те ≪скидається≫ в прагматику.

9) У 70–80-х рр. XX ст. проблематика лінгвопрагматики визначалася вивченням впливів контексту на функціонування мовних одиниць у комунікації, передусім мовленнєвих актів, вивченням неконвенціональних засобів мовленнєвої діяльності людей, схем інтерпретаційної діяльності учасників інтеракції та деякими іншими проблемами.

10) Ця схема засвідчує розмежування авторами "лінгвістичної" і "нелінгвістичної" прагматики, яке значною мірою буде поставлене під сумнів на пізніших етапах розвитку мовознавчої думки, коли найістотнішим чинником у розширенні лінгвопрагматичної проблематики стане врахування всіх без винятку складників комунікативного акту.

11) Таким чином, найважливішими напрямами дослідження лінгвістичної прагматики слід уважати сфери, пов'язані з суб'єктом мовлення (ілокутивні сили, мовленнєві тактики, правила ведення діалогу, різні настанови, референцію мовця, прагматичні пресупозиції, емпатію тощо), адресатом мовлення (правила виведення непрямих смислів, перлокутивні ефекти, типи мовленнєвого реагування тощо), пов'язані зі стосунками між учасниками спілкування (форми мовленнєвого спілкування, соціально-етикетний бік мовлення та ін.).

12) Незважаючи на значну розбудову свого проблемного поля, до цього часу залишається дискусійним, на наш погляд, найважливіше питання: що слід уважати *предметом* лінгвістичної прагматики. Ця невизначеність не в останню чергу викликана значним розширенням об'єктів дослідження комунікативної лінгвістики, зокрема введенням у лінгвістичний обіг таких її засадничих категорій, якими є дискурси і мовленнєві жанри.

13) *Мікропрагматика* як прагматика одиниць мовного коду в спілкуванні включає в своє проблемне поле питання, пов'язані із суб'єктивним чинником, наявним у семантико-прагматичній структурі лексичних і граматичних одиниць; дискурсивними словами, особливостями їх функціонування та системної організації; імплікаціями та імплікатурами дискурсу; пресупозиціями у висловленні; інференціями; проблемами суб'єктивності референції та анафори.

14) *Макропрагматика,* або прагматика категорій міжособистісного спілкування, звернена до таких понять: інтенціональність, емпатія і прагматика фокусу й погляду учасників комунікації; прагматичні аспекти стратегій і тактик спілкування; принципи, максими і постулати спілкування; прагматика ввічливості.

15) *Мегапрагматика* зосереджена на проблемах специфіки соціального контексту і його впливу на всі складники комунікативного акту, зокрема, дискурси. До цього типу прагматики варто долучити прагматичні аспекти конвертаційного аналізу, а також прагматику міжкультурної комунікації в її зв'язках із проблемами організації дискурсів, мовленнєвих актів і мовленнєвих жанрів.

SPEAKING SECTION.
PICTURE DESCRIPTION WORKSHEET

DISCUSSING FACTUAL INFORMATION

Where? There is a classroom/conference hall/office/canteen/lounge-bar at the university/school/library/enterprise/firm. It looks out-of-date/modern and well/poorly equipped. The people in the picture belong to different/the same ethnic group/(s)/ nationalit(y)ies/origin. They represent the same/different age and gender groups.

When? The scene takes place during the negotiation procedure/match/excursion/ staff meeting. We can/cannot see in the picture, what is the weather like outside. The scene takes place in the early morning/at midday/in the evening/at midnight.

What else can be mentioned about the setting? What can you see in the background and the foreground?

Consider the statements below *true, false* or *not stated*. Justify your choice:
1) The people in the picture are in their late forties.____
2) The woman in the picture is single and the man is obviously married.____
3) It is spring outside._____
4) The man is holding his right hand in his trousers pocket.____
5) The woman is wearing a long straight skirt.____
6) The people in the picture are both wearing expensive watches._____
7) The people in the picture are using actively body language._____
8) It is obvious that the man and the woman in the picture are strogly arguing about something._____

Unit 4. Discourse Analysis

DISCUSSING CONCEPTUAL INFORMATION
1. Who could take the picture?
2. What is the possible title of the picture under consideration? How can it be related to Applied Linguistics area? What is the message of the picture?
3. Are the people in the picture posing for camera on purpose?
5. What do you think had happened before the picture was taken?
6. What will they do next after the picture having been taken?

PICTURE DESCRIPTION
Make up a list of 20 key-words covering factual and conceptual information of the picture considered. Describe the picture using the key-words, cover factual and conceptual information of the picture considered.

 LISTENING (AUDIOVISUAL COMPREHENSION) SECTION

LANGUAGE AND DISCOURSE ANALYSIS

I. Watch a video recording about the general view of language and discourse analysis issue. Answer the following questions. Justify your answer relying upon the facts form the video material. (https://www.youtube.com/watch?v=JZ8bkus3vis)

1) What is linguistics about from David Crystal's perspective?
2) What is the difference between phonetics and phonology? How may discourse analysis be related to these branches of linguistics?
3) Why does the speaker in the video recording consider Leo Spitzer's definition of the *discourse analysis* concept to be fairly not brilliant? Do you think the concept under consideration covers only the idea of the examination of any significant semiotic event?
4) How are the concepts of *signifier* and *signified* viewed from Ferdinand de Saussure's perspective? Provide your own examples relying on the scholar's scheme.
5) What is semiosis? Why is this linguistic phenomenon important for the process of discourse analysis?

II. Indicate whether the statements below are true/ false/not stated, justify your choice relying on the video materials.

1) Psycholinguistics is looking at how social conventions, norms, values, the rules, governing our behavior, are getting inscribed into the language that we speak. ____

2) David Spitzer is known to be the founder of discourse analysis. _____

3) Phonology is the study that lets us differentiate between sounds._____

4) Discourse analysis originates from literature._____

5) David Crystal, a prominent linguist, has been mentioned twice in the video recording. _____

6) Syntax is the examination of how meaningful structures are put together out of different words. _____

7) From Ferdinand de Saussure's perspective, discourse analysis is the examination of any significant semiotic event._____

8) If we had a phrase like *the cat sat on the mat* which is syntactically correct and functional, then a phrase *the cat purred on the map* would surely differ from the first one and be unequally functional. _____

9) Semiosis is viewed as the interpretation of meaning making. _____

10) The speaker in the video recording claims the gender category to be insignificant for the culture we live in, and thus not being inscribed within the language system._____

III. Fill in the gaps with appropriate words according to what is being delivered in the video-piece.

<...> Then there are………… and phonetic areas of linguistics these to do with the study of sounds and how sound is used to make meaning, so phonetics is the study of how we can ………. between sounds, what's the minimum difference that can be used for meaning and then phonology is the set of ……….., used by a particular language so the human voice can ……….. many different phonetic ………….., but in any one language we only have a limited set of phonological ones. Syntax is the examination of how…………. structures are put together out of different words so if we had phrases like the famous syntactic one is *the cat sat on the mat*, that's ……… correct you know works it's functional and then we could have also had *the cat purred on the map* that would ………. be functional. You could even have something like *the cat juggled on the mat*, which was it might not make sense it's still …………… legit, you know, works because *juggled* is the past ……… of the verb. So, the fact that *the camp juggled on the map* wouldn't or would perhaps be seen as nonsense that's a kind of ………… thing, so what we're looking at there is how whole …………… has meaning or stance in defiance of …………… on me and so, that's that looks at what the meaning of a whole group of words might be so, a sentence or phrase, or …….…….. Pragmatics looks at the meanings that aren't present, but are somehow indicated by a syntactic ….…..... <...>.

WRITING SECTION

Choose one of the topics below to write an opinion essay. Before writing study the tips.
1) Computer programs can analyze various discourse materials appropriately. *(for instance, can essays be analyzed and evaluated/assessed by computers)?*
2) Discourse analysis may be viewed as a tool of linguistics methodology only.
3) Discourse itself covers only oral and written texts and is not related anyhow to sounds, visual images or music that a particular text piece may contain *(TV-shows, radio programmes, movies – if these can be viewed as discourse samples).*

Tips for writing

1. Decide whether you agree or disagree with the title. Try to think of at least two or three good reasons to support your opinion, including examples of why you think the alternative point of view is wrong.

2. Introduce each paragraph with a topic sentence, outlining the main ideas. Do not write about advantages or disadvantages or points for or against.

3. Write in formal style.

4. Do not use colloquial expressions, short forms or emotive vocabulary.

5. Organise your essay into clear paragraphs:
   a) an introductory paragraph in which you state the topic and your opinion;
   b) a main body which consists of several paragraphs, each presenting a separate viewpoint supported by reasons;
   c) a conclusion in which you restate your opinion using different words.

SELF-REFLECTION SHEET ON THE UNIT 4 MATERIALS:

1) THE FOLLOWING 10 KEY-CONCEPTS OF THE UNIT CAN BE SINGLED OUT:
_____
_____

2) UNIT REVIEW: THE UNIT GENERALLY RUNS ABOUT:

_____
_____
_____
_____

3) WHAT ARE THE ADVANTAGES OF THE UNIT UNDER CONSIDERATION?
_____
_____
_____

4) WHAT ARE THE DISADVANTAGES OF THE UNIT UNDER CONSIDERATION?
_____
_____
_____

5) WHICH TASKS DO YOU CONSIDER TO BE MOST USEFUL IN THE UNIT?
_____
_____
_____

6) TAKE A LOOK AT THE TITLE OF THE NEXT UNIT. WHAT DO YOU EXPECT TO LEARN WHEN YOU START COVERING THE MATERIALS OF THE UNIT?
_____
_____
_____

VIDEO TRANSCRIPTS

THE JOY OF LEXICOGRAPHY
https://www.youtube.com/watch?v=J4VzuWmN8zY&t=4s

Now, have any of y'all ever looked up this word? You know, in a dictionary? (Laughter) Yeah, that's what I thought. How about this word? Here, I'll show it to you. Lexicography: the practice of compiling dictionaries. Notice – we're very specific – that word "compile." The dictionary is not carved out of a piece of granite, out of a lump of rock. It's made up of lots of little bits. It's little discrete – that's spelled D-I-S-C-R-E-T-E – bits. And those bits are words.

Now one of the perks of being a lexicographer – besides getting to come to TED – is that you get to say really fun words, like lexicographical. Lexicographical has this great pattern: it's called a double dactyl. And just by saying double dactyl, I've sent the geek needle all the way into the red. (Laughter) (Applause) But "lexicographical" is the same pattern as "higgledy-piggledy." Right? It's a fun word to say, and I get to say it a lot. Now, one of the non-perks of being a lexicographer is that people don't usually have a kind of warm, fuzzy, snuggly image of the dictionary. Right? Nobody hugs their dictionaries. But what people really often think about the dictionary is, they think more like this. Just to let you know, I do not have a lexicographical whistle. But people think that my job is to let the good words make that difficult left-hand turn into the dictionary, and keep the bad words out.

But the thing is, I don't want to be a traffic cop. For one thing, I just do not do uniforms. And for another, deciding what words are good and what words are bad is actually not very easy. And it's not very fun. And when parts of your job are not easy or fun, you kind of look for an excuse not to do them. So if I had to think of some kind of occupation as a metaphor for my work, I would much rather be a fisherman. I want to throw my big net into the deep, blue ocean of English and see what marvelous creatures I can drag up from the bottom. But why do people want me to direct traffic, when I would much rather go fishing? Well, I blame the Queen. Why do I blame the Queen? Well, first of all, I blame the Queen because it's funny. But secondly, I blame the Queen because dictionaries have really not changed.

Our idea of what a dictionary is has not changed since her reign. The only thing that Queen Victoria would not be amused by in modern dictionaries is our inclusion of the F-word, which has happened in American dictionaries since 1965. So, there's this guy, right? Victorian era. James Murray, first editor of the Oxford English Dictionary. I do not have that hat. I wish I had that hat. So he's really responsible for a lot of what we consider modern in dictionaries today. When a guy who looks like that, in that hat, is the face of modernity, you have a problem. And so, James Murray could get a job on any dictionary today. There'd be virtually no learning curve.

And of course, a few of us are saying: okay, computers! Computers! What about computers? The thing about computers is, I love computers. I mean, I'm a huge geek, I love computers. I would go on a hunger strike before I let them take away Google

Book Search from me. But computers don't do much else other than speed up the process of compiling dictionaries. They don't change the end result. Because what a dictionary is, is it's Victorian design merged with a little bit of modern propulsion. It's steampunk. What we have is an electric velocipede. You know, we have Victorian design with an engine on it. That's all! The design has not changed.

And OK, what about online dictionaries, right? Online dictionaries must be different. This is the Oxford English Dictionary Online, one of the best online dictionaries. This is my favorite word, by the way. Erinaceous: pertaining to the hedgehog family; of the nature of a hedgehog. Very useful word. So, look at that. Online dictionaries right now are paper thrown up on a screen. This is flat. Look how many links there are in the actual entry: two! Right? Those little buttons, I had them all expanded except for the date chart. So there's not very much going on here. There's not a lot of clickiness. And in fact, online dictionaries replicate almost all the problems of print, except for searchability. And when you improve searchability, you actually take away the one advantage of print, which is serendipity. Serendipity is when you find things you weren't looking for, because finding what you are looking for is so damned difficult.

So – (Laughter) (Applause) – now, when you think about this, what we have here is a ham butt problem. Does everyone know the ham butt problem? Woman's making a ham for a big, family dinner. She goes to cut the butt off the ham and throw it away, and she looks at this piece of ham and she's like, "This is a perfectly good piece of ham. Why am I throwing this away?" She thought, "Well, my mom always did this." So she calls up mom, and she says, "Mom, why'd you cut the butt off the ham, when you're making a ham?" She says, "I don't know, my mom always did it!" So they call grandma, and grandma says, "My pan was too small!" (Laughter)

So, it's not that we have good words and bad words. We have a pan that's too small! You know, that ham butt is delicious! There's no reason to throw it away. The bad words – see, when people think about a place and they don't find a place on the map, they think, "This map sucks!" When they find a nightspot or a bar, and it's not in the guidebook, they're like, "Ooh, this place must be cool! It's not in the guidebook." When they find a word that's not in the dictionary, they think, "This must be a bad word." Why? It's more likely to be a bad dictionary. Why are you blaming the ham for being too big for the pan? So, you can't get a smaller ham. The English language is as big as it is.

So, if you have a ham butt problem, and you're thinking about the ham butt problem, the conclusion that it leads you to is inexorable and counterintuitive: paper is the enemy of words. How can this be? I mean, I love books. I really love books. Some of my best friends are books. But the book is not the best shape for the dictionary. Now they're going to think "Oh, boy. People are going to take away my beautiful, paper dictionaries?" No. There will still be paper dictionaries. When we had cars – when cars became the dominant mode of transportation, we didn't round up all the horses and shoot them. You know, there're still going to be paper dictionaries, but it's not going to be the dominant dictionary. The book-shaped dictionary is not going to be the only shape dictionaries come in. And it's not going to be the prototype for the shapes dictionaries come in.

So, think about it this way: if you've got an artificial constraint, artificial constraints lead to arbitrary distinctions and a skewed worldview. What if biologists could only study animals that made people go, "Aww." Right? What if we made aesthetic judgments about animals, and only the ones we thought were cute were the ones that we could study? We'd know a whole lot about charismatic megafauna, and not very much about much else. And I think this is a problem. I think we should study all the words, because when you think about words, you can make beautiful expressions from very humble parts. Lexicography is really more about material science. We are studying the tolerances of the materials that you use to build the structure of your expression: your speeches and your writing. And then, often people say to me, "Well, OK, how do I know that this word is real?" They think, "OK, if we think words are the tools that we use to build the expressions of our thoughts, how can you say that screwdrivers are better than hammers? How can you say that a sledgehammer is better than a ball-peen hammer?" They're just the right tools for the job.

And so people say to me, "How do I know if a word is real?" You know, anybody who's read a children's book knows that love makes things real. If you love a word, use it. That makes it real. Being in the dictionary is an artificial distinction. It doesn't make a word any more real than any other way. If you love a word, it becomes real. So if we're not worrying about directing traffic, if we've transcended paper, if we are worrying less about control and more about description, then we can think of the English language as being this beautiful mobile. And any time one of those little parts of the mobile changes, is touched, any time you touch a word, you use it in a new context, you give it a new connotation, you verb it, you make the mobile move. You didn't break it. It's just in a new position, and that new position can be just as beautiful.

Now, if you're no longer a traffic cop -- the problem with being a traffic cop is there can only be so many traffic cops in any one intersection, or the cars get confused. Right? But if your goal is no longer to direct the traffic, but maybe to count the cars that go by, then more eyeballs are better. You can ask for help! If you ask for help, you get more done. And we really need help. Library of Congress: 17 million books, of which half are in English. If only one out of every 10 of those books had a word that's not in the dictionary in it, that would be equivalent to more than two unabridged dictionaries.

And I find an un-dictionaried word – a word like "un-dictionaried," for example – in almost every book I read. What about newspapers? Newspaper archive goes back to 1759, 58.1 million newspaper pages. If only one in 100 of those pages had an un-dictionaried word on it, it would be an entire other OED. That's 500,000 more words. So that's a lot. And I'm not even talking about magazines. I'm not talking about blogs – and I find more new words on BoingBoing in a given week than I do Newsweek or Time. There's a lot going on there.

And I'm not even talking about polysemy, which is the greedy habit some words have of taking more than one meaning for themselves. So if you think of the word "set," a set can be a badger's burrow, a set can be one of the pleats in an Elizabethan ruff, and there's one numbered definition in the OED. The OED has 33 different

numbered definitions for set. Tiny, little word, 33 numbered definitions. One of them is just labeled "miscellaneous technical senses." Do you know what that says to me? That says to me, it was Friday afternoon and somebody wanted to go down the pub. (Laughter) That's a lexicographical cop out, to say, "miscellaneous technical senses."

So, we have all these words, and we really need help! And the thing is, we could ask for help – asking for help's not that hard. I mean, lexicography is not rocket science. See, I just gave you a lot of words and a lot of numbers, and this is more of a visual explanation. If we think of the dictionary as being the map of the English language, these bright spots are what we know about, and the dark spots are where we are in the dark. If that was the map of all the words in American English, we don't know very much. And we don't even know the shape of the language. If this was the dictionary – if this was the map of American English – look, we have a kind of lumpy idea of Florida, but there's no California! We're missing California from American English. We just don't know enough, and we don't even know that we're missing California. We don't even see that there's a gap on the map.

So again, lexicography is not rocket science. But even if it were, rocket science is being done by dedicated amateurs these days. You know? It can't be that hard to find some words! So, enough scientists in other disciplines are really asking people to help, and they're doing a good job of it. For instance, there's eBird, where amateur birdwatchers can upload information about their bird sightings. And then, ornithologists can go and help track populations, migrations, etc.

And there's this guy, Mike Oates. Mike Oates lives in the U.K. He's a director of an electroplating company. He's found more than 140 comets. He's found so many comets, they named a comet after him. It's kind of out past Mars. It's a hike. I don't think he's getting his picture taken there anytime soon. But he found 140 comets without a telescope. He downloaded data from the NASA SOHO satellite, and that's how he found them. If we can find comets without a telescope, shouldn't we be able to find words?

Now, y'all know where I'm going with this. Because I'm going to the Internet, which is where everybody goes. And the Internet is great for collecting words, because the Internet's full of collectors. And this is a little-known technological fact about the Internet, but the Internet is actually made up of words and enthusiasm. And words and enthusiasm actually happen to be the recipe for lexicography. Isn't that great? So there are a lot of really good word-collecting sites out there right now, but the problem with some of them is that they're not scientific enough. They show the word, but they don't show any context. Where did it come from? Who said it? What newspaper was it in? What book?

Because a word is like an archaeological artifact. If you don't know the provenance or the source of the artifact, it's not science, it's a pretty thing to look at. So a word without its source is like a cut flower. You know, it's pretty to look at for a while, but then it dies. It dies too fast. So, this whole time I've been saying, "The dictionary, the dictionary, the dictionary, the dictionary." Not "a dictionary," or "dictionaries." And that's because, well, people use the dictionary to stand for the whole language. They use it synecdochically. And one of the problems of knowing a word like "synecdochically" is that you really want an excuse to say "synecdochically." This whole talk has just been an excuse to get

me to the point where I could say "synecdochically" to all of you. So I'm really sorry. But when you use a part of something – like the dictionary is a part of the language, or a flag stands for the United States, it's a symbol of the country – then you're using it synecdochically. But the thing is, we could make the dictionary the whole language. If we get a bigger pan, then we can put all the words in. We can put in all the meanings. Doesn't everyone want more meaning in their lives? And we can make the dictionary not just be a symbol of the language – we can make it be the whole language.

You see, what I'm really hoping for is that my son, who turns seven this month – I want him to barely remember that this is the form factor that dictionaries used to come in. This is what dictionaries used to look like. I want him to think of this kind of dictionary as an eight-track tape. It's a format that died because it wasn't useful enough. It wasn't really what people needed. And the thing is, if we can put in all the words, no longer have that artificial distinction between good and bad, we can really describe the language like scientists. We can leave the aesthetic judgments to the writers and the speakers. If we can do that, then I can spend all my time fishing, and I don't have to be a traffic cop anymore. Thank you very much for your kind attention.

LANGUAGE DESIGN (INTERVIEW WITH NOAM CHOMSKY)
https://www.youtube.com/watch?v=MLk47AMBdTA)

To look into the question of language design, it's useful to think of how human beings evolved. We don't know a great deal about it, but we know some things. So, for example, it's fairly clear from the archaeological record, that modern humans, modern *Homo sapiens,* cognitively modern Homo sapiens developed quite recently in evolutionary time, and maybe within the last roughly hundred thousand years, which is a flick of an eye, that's when you get the enormous increase, explosion of indications of the creative activity, complex family structures, symbolism and so on, all of this develops roughly in that period, and interestingly, there has been no detectable evolution of these capacities in roughly the past 50.000 years.

That's the period since our ancestors left Africa, a small number of them, and pretty quickly spread over the world. So, all humans are pretty much identical with regard of the cognitive capacity, linguistic capacity and so, which means, that there's been essentially no detectable evolution. So, there's small window there, where something happened, and it's generally assumed by paleoanthropologist people who study these topics. That must have been the emergence of language, because it's hard to imagine any of these basically creative activities without language, and that language does provide the mechanisms for them.

So, it seems as though the core of human sensibility, and a creative and cognitive capacity, is the development of this completely unique capacity. There's nothing analogous to it anywhere in the animal world. There are animal signaling systems, but they're completely different in design and use in just about every dimension. So, something strange happened, roughly, maybe, a hundred thousand years ago, not very

long, and language emerged in humans, and the question then is, well, what kind of a system is it?

On the surface languages look very different from one another. So, if somebody walks into the room and starts speaking Swahili, I'm not going to understand a word. Though I will recognize that it's a language. I won't understand it, but I know it's not noise. No. As soon, as you look more deeply, you find that these languages are basically mould into a pretty similar design, may be an identical design, the large parts of the length of what we hear, is just the sounds. But that's a very superficial part of language. The core of language is principles that determine actually an infinite array of possible expressions, structured expressions, which have definite meanings. Now, all of that is well, beyond, what we can just observe, but I say looking at the texts, and when a child is learning a language, the child doesn't learn those things, there's no evidence for them. Almost, no evidence for them, nobody can teach them.

We don't know, what they are. These are just part of our nature, the core principle, so-called syntactic principles that form expressions, and that provide specific interpretations for them, that's apparently just all part of our nature. And then there are various ways of externalizing it in sound, or in sign which is about the same, but it. But that's a kind of a superficial manifestation of an internal uniformity, and the really exciting, and it almost has to be this way. If you think about the way, the system developed, apparently all of a sudden, in evolutionary terms which meant that there were very limited selectional pressures, so, it probably was designed as a computational….it is a computational system, so many explanation for this array of capacities computational systems have certain optimal characteristics. That some are more efficient than others, and there's every reason to believe that. This developed pretty suddenly as an optimal communication system essentially following laws of nature very much.

The way a snowflake assumes a very complex form, and not because of experience or training, but just, because that's the way the laws of physics work, and there's every reason to believe that language is something like this. Now, to try to show it, is a trivial matter. You have to try to show that the superficial variety of languages, actually reduces to principles of a common character which approach notions of optimal design. And there has been, I think, no notable progress in that process, is a long way to go to try to demonstrate it for, but then, of course, then one wants to go beyond the tried, maybe ultimately to discover the neural basis for whatever this unique capacity is; and it's a very hard problem the study for humans.

So, we know a lot about the human visual system, because of direct experimentation with cats and monkeys. We allow ourselves to do direct experimentation, you know, sticking electrodes into the brain and so on. Controlled experiments, but we don't do it with humans, and humans have about the same visual system as cats and monkeys, so we know about the human visual system, you can't do that for language, there are no analogous systems.

So, you can stand, watch, study other animals, we're unique in this respect, and invasive experiments with human beings are, of course, barred. So, it's a very complex and intricate matter to try to find clever ways of getting around the barriers to learn

something about these topics. And some progress is being made. I think, we can look forward to a period, when there will be convergence of various modes of inquiry into design of language neuro basis acquisition of our possible varieties of language, and so on. That's crucial task for the future which in fact is directed to the core of human nature. The core of cognitive of human nature.

The most intriguing question, I think, is the one that I have basically just mentioned, there's reason to believe that the core of human intellectual nature, cognitive nature, is a computational system which probably has something like the properties of a snowflake. It simply had to develop this way, given biological and physical law and special circumstances. And the most intriguing question is, to try to see, if that's true, but if it is, to show that it's true.

CORPORA AND SPOKEN LANGUAGE
https://www.youtube.com/watch?v=UKQcVE9d67s&t=371s

Then, in the early 1990s, I moved to the University of Nottingham and began work on spoken corpora with my colleague Ronald Carter. And we built the CANCODE spoken corpus, C-A-N-C-O-D-E, Cambridge and Nottingham Corpus of Discourse in English. By that time written copra were taping indeed, tens of millions of words. But spoken corpora have always opened necessity been somewhat smaller because of the immense costs involved in recording, transcribing etc been spoken data. And it's also quite difficult to obtain the appropriate ethical permissions and so on. Tang code nonetheless is 5 million words of everyday data which is extremely useful. Also now spoken corpora are much easier to collect. We have miniaturized audio recording equipment; the software is much better, more capable of handling different kinds of data. So in a way, there has been a revolution and the spoken corpus has indeed come of age.

My particular area of interest is conversation and conversational corpora, and there are a couple of reasons for this. Firstly, of course, conversation is for most of us the most typical way that we use language in every day of our lives. We talk to our friends, our colleagues, our neighbors, family etc and expend an awful lot of language every day much more than we write. But secondly, a good reason for collecting spoken corpora is that, that it is actually quite difficult to be objective about how we speak, it's much easier to be objective about how we write. When we write we can revise, we can change things, we can cut and paste, we can do all sorts of things, we can reflect on what we're doing. When we speak we have no such opportunities for reflection, it's in real time conversational speech at least. There are other kinds of more formal speaking where we do have time to think but everyday conversation is real-time, online, face-to-face so we don't have time to sit back and reflect, and observe how we speak. Therefore it's not surprising that a lot of the ways in which we describe language and as a result the ways in which we teach it and what we teach are based on notions that come from the written language.

So the question is "What do we learn by using spoken corpora?" "What's the difference?" "What are the different insights?" and "If there are different insights?" "What is their relevance for language teaching, for TESOL?"

As you probably know, one of the standard techniques that corpus linguists use to count frequency. We can create frequency lists which tell us what are the most frequent items words or phrases or grammatical patterns in the language and equally what are the rare or hardly ever used words and phrases and patterns. So, frequency lists are quite useful.

A couple of years ago my English Profile colleague Dr. Paula Buttery, and the English Profile incidentally is a very big project looking into learner English within the framework of the Common European Framework of Reference (and I would invite you to visit the website of the English Profile). My colleague Dr. Paula Buttery and I did an investigation where we took the top two thousand, top couple of thousand most frequent words in the British National Corpus spoken segments and compared them with a similar top two thousand words in the British National Corpus written segment. Now we found, of course, that there was a great deal of overlap, there's no surprise, there. There's hardly anything that you can write which you cannot say and vice versa. But the overlap came to about two thirds, about sixty-five percent and that meant that there was a good third, about thirty-five percent, of the words which were unique to either the spoken list or the written list which was fascinating in itself. So there is evidence that the most common vocabulary of writing and speaking is different one compared with the other.

Certainly, my colleague Ronald Hart and I did a comparison of the top 50 words, the 50 most frequent words in our CANCODE corpus and in a same sized Corpus of General Written English consisting of newspapers and books, and magazines, and so on typical everyday written texts. And we found that there were a number of words in the spoken list, the spoken top 50, which didn't appear in the written top 50 at all. These included words like *right* and *well*, and the verbs *know* and *think*. K-N- 0-W, *know*, is obviously very very frequent because people are all the time saying *you know*. Now there's a tendency sometimes and I've had to confront this tendency in lectures and conferences and things that I've written. There is a tendency to think of expressions like *you know* as being rather lazy uses of language, sloppy, not the sorts of things we would want to teach as language teachers. But believe me, everybody uses them all the time. Educated speakers, because of all ages, genders, geographical backgrounds, everything, we do it all the time. Why? Because we are constantly reaching out to our interlocutors, the people we are talking to. We are constantly monitoring their conversation, we want to give out this signal that "Well, you and I where we share the same world, we share the same life experiences, we're on the same wavelength, *you know*". And this is important, this reaching out this, this creation of interaction, not just what we are talking about. Of course, we may be talking about our jobs, our holidays, our friends, anything, but what is really much more important is how we do it and how we try to create this relationship with the other speakers, to create this interaction

between us. That's one of the things that leaps out of the frequency lists if you compare spoken and written language.

And if we think of words like *right* and *well,* then, obviously, these are very useful words for organizing our conversations. We use *right,* perhaps, if we want to end the conversation *"Right,* see you tomorrow". *"All right,* let's talk about something else" a way of organizing our topics in the conversation. So, again not as what we're talking about but how we talk about it.

Well is an interesting word. People often say to me "Do you live in Cambridge?" and I usually answer *"Well,* near Cambridge". Because I happen to live in a village about eight miles outside of Cambridge. Why do I say *well*? because I've been asked a yes/no question "Do you live in Cambridge?" but I can't answer it as a yes-no question so I have to give out a signal that says "I'm going to take the discourse in a different direction, I can't take it in the direction that you have projected".

So, these words like *well* and *right* they're very important organizational words, they are strategic ways of organizing and managing our conversations. So, it's not surprising that they are incredibly frequent in the spoken corpus and may be absent until quite considerably way down in the frequency list of the written corpus.

If you like, their small words with very big meanings are the small words, for example, the word *just* which is in the top 50 spoken but not written. I could say to you "Can I ask you a question?" But I can make it much softer, much less in your face, much friendlier by saying "Can I *just* ask you a question?" So, it's these little, what I call the interactive words, this repertoire of vocabulary in the most frequent spoken lists that we get from spoken corpora which create interaction, and then they suggest to us that there is indeed a thick skill which we can call the skill of interaction. Over and above being able to pronounce things correctly, over and above being able to talk about our experiences, we need to have the appropriate way of doing it, the appropriate way of engaging with the other person or persons who are there in real time in front of us, and I think this is one of the very big things that I've got from studying spoken corpora.

LINGUISTICS AND DISCOURSE ANALYSIS
https://www.youtube.com/watch?v=JZ8bkus3vis

Before looking at discourse analysis, I'm going to just quickly review the discipline that it fits into. So, discourse analysis comes to us from linguistics. David Crystal tells us that linguistics is the science of language and linguists are the people who try to understand, why human language is the way it is. So, linguists study the history and acquisition of language, and its structure and use. So, I'm just going to pick up on that structure use distinction a little bit, so, on the sort of the structural side we have these different areas of linguistics, so, structural linguistics is interested in the formal properties of language, so, includes things like word structure, which we find studied in morphology and so, that's like how you make a participle from a root word you know, so, *run – running* so, that kind of add an *ING* word structure.

Then there are phonological and phonetic areas of linguistics. These are to do with the study of sounds and how sound is used to make meaning, so, phonetics is the study of how we can differentiate between sounds, what's the minimum difference that can be used for meaning. And then phonology is the set of sounds, used by a particular language, so, the human voice can produce many different phonetic distinctions, but in any one language we only have a limited set of phonological ones. Syntax is the examination of how meaningful structures are put together out of different words, so, if we had phrases like the famous syntactic one is *the cat sat on the mat*, that's syntactically correct, you know, works it's functional and then we could have also had *the cat purred on the map* that would equally be functional. You could even have something like *the cat juggled on the mat*, which was it might not make sense, it's still syntactically legit, you know, works, because *juggled* is the past tense of the verb.

So, the fact that *the camp juggled on the map* wouldn't or would perhaps be seen as nonsense, that's a kind of semantics thing, so, what we're looking at there, is how whole phrase has meaning or stance in defiance of conventions on me, and so, that's that looks at what the meaning of a whole group of words might be so, a sentence, or phrase, or clause. Pragmatics looks at the meanings that aren't present, but are somehow indicated by a syntactic arrangement. So, if you said to somebody *are you putting the cattle on* and really meant *will you make me a cup of tea,* that's kind of pragmatic usage. On the other side, here we've got this kind of used things, so, there's psycholinguistics, how we develop an acquired language, historical linguistics, how languages change and develop over time, varieties of language…. so *slang, creoles and pigeons*.

These are …errr.. fascinating areas of study but one does not terribly relevant to what we are doing, but this side of the thing is. So, sociolinguistics is looking at how social conventions, norms, values, the rules governing our behavior getting inscribed into the language that we speak. So, we live in a culture where gender distinction is very important and we can see that inscribed in our language in terms of we have different verbal forms and different pronoun forms for males and females.

Then there's these two things here, which, these are more relevant to us against discourse analysis, the thing we're actually interested in today's conversation analysis. I will register that as a special case discourse analysis, and people be upset by that, but I've got to shorten this discourse analysis, look really at how sociolinguistic value gets inscribed in the language as it is being used, and it pretty much uses things like semantics and pragmatics. I borrow some bits of this syntactics as well in order to do that in a wiry. So, we're going to look at the overlap between these different things. This is Leo Spitzer, and many people think of him as the founder, the granddad of discourse analysis, that's given the people like to have a starting place for anything Leo Spitzer could have started anywhere, but really you know we've been looking at discourse and working out, how it means, what it means, way back, you know, go back to the Greeks, beyond that I think any time, people who've used language, they've wanted to know how it works.

So, from Leo's perspective discourse analysis is the examination of any significant semiotic event which is not a brilliant definition to be fair, because it was dividing the

thing that we didn't know discourse analysis in reference to something we've probably never even heard of significant semiotic event, so what is this, what semiotic? Well, way back in the day semiotics was the study of how signs function in the construction of meaning and it, you know, if we talk to three four hundred years ago, people would have talked about the semiotic subnets and or agriculture meaning so, how you interpret material traces of illness, so the doctor sees the spots on the skin and interprets it in terms of measles.

So, if we think about that in a more formal way by the time I get to the nineteenth century, we're looking at these two gentlemen here, this is Charles Sanders Percy, who was an American, I may not be saying his name entirely correctly, and there's something unusual other ways of his name, but I don't quite know what it is. He's a pragmatist American philosopher, he was a person who inspired some of the great names of American philosophical tradition, particularly, Jewry and also to some extent the James brothers. William does anyway, there is a way he was an important philosopher from America, and he says that a sign is something that stands in for something else it's some respect or capacity. And once he sort of made that claim the thing, that's most interesting leaders are the respects and capacities of the site.

So, how kind of thing standing for something else, so, he says signs can be iconic, so they can be similar in some way to the thing they stand in for. So, these are signs for about, so if somebody says *ding-dong* or if somebody drew that picture or present it to you on a screen and it made you think of *Bell,* these are then iconically representing the bell, and that's iconic, because they are similar to the object is standing for in some way, and so this is supposed to sound a bit like the noise a bell makes. This is supposed to look a bit like about all, right, he also says you were indexical signs.

So, here we can get the Bell in the same way, so the things being stood always is about, but this time it's the noise made by these guys he rattles this thing. If you hear a ringing sound – *clang clang clang* noise – then you have a tendency to think there must be a Bell nearby so, the sound indicates it is an index of the presence of a bat. And it could be a logical relationship as well, so, if somebody says they are a brother that indicates that they have a sibling in their immediate family, and a brother or sister themselves. So, it can be hard to see how somebody could be an only one and brother. And then Pierce also says that there are symbolic signs where the standing in four is achieved just by convention. So, we use the sound [b] earlier than the word *"bell",* but we could equally as well use the word *"chime"* or any other word at all, you know, we could use the word *"sandwich".*

There's no reason why *bell* has anything to do with *bells* any more than any other sound we could think. Right, so this is the other dimension of its either side of a coin, if you like, and this comes from Ferdinand de Saussure. He was a famous Swiss linguist, famously could speak wide variety of languages with considerable fluency and thought and wrote on the nature of the sign, but in a different dimension, so, he's not some interest in different types as there's the internal mechanics. So, he would basically agree with persuade and he says that, you know, yet one thing stands in for another, but he says with these two parts of the sign there. So, there's the thing that

does the standing in for which he calls a signifier, that would be landmarks or sounds, or gestures, that we read here or observe.

And then there are the signified, that's the things that are being stood in for, so, here we got a bunch of sounds *lion, leo, Simba,* all of which can be used to stand in for this. So, those are the two halves of the sign, but the other thing that he says it's important to secure, observes that this relationship is arbitrary and it doesn't just mean that, you know, it's not just that we could say *sandwich* and mean that, it's not just that we could say we all have *lion,* or we could say *Leo,* or we could say *Simba* and mean the same thing. It's also far between that we group particular objects into the same class.

So, we have a specific class of objects that we call *lion* and it excludes *leopards*, but it could include them that's an arbitrary distinction we drew there. And the people get upset about this appetizer. They say, well, look, there is it, there's a natural difference between a lion and a leopard. And yes, maybe, there is, but also there are natural distinctions within the group that we call lions. So, you get lions from certain types, you know, so, the places which are quite different to others, they may be able to reproduce, but you can also get different big cats to reproduce as one of the dart lions and produce hybrids. So, the idea here is that there isn't this kind of uniform, homogeneous, naturally, occurring object, it's an arbitrary class that we've put together.

So, personally, you're both taking different approaches to signs, but what their work brings out is that there is something arbitrary and conventional in the way the signs work. And this is going to be important for discourse analysis, because in discourse analysis we aim at finding this arbitrary and conventional dimension of making meaning, *semiosis.*

So, why are we doing that? Well we're so familiar with semiotic processes that they go on around us we often don't see the arbitrary in the conventional and so, we can be in a way hijacked by it. And just as the fish is always in the water, and therefore, loses sight of the water, we very often lose sight of the fact that we're constantly engaged in semiosis, the interpretation of meaning making. So, we behave as though some of the meanings that we work with aren't arbitrary and conventional, that the world is not necessarily the way we think it is, but it could be chopped up in other way, so, we could use different sounds to mean the same objects, or we could actually divide the object that we're referring to up in different ways.

So, this arbitrary must tend to disappear, if rooms, just as quickly as we start to interpret science. But we can sometimes recover these assumptions about this arbitrariness by listening carefully to the language, and that's what discourse analysis does. So, wherever we hear somebody catching something up as natural or common sense, that is usually the case of people did send them natural religion or traditional illusion, sorry, the religion nature and common sense are sometimes the alibis for the arbitrary and the conventional. And we do this because arbitrary conventions in language, that we have forgotten and turned into natural or common sensual distinctions, are often used to disguise the arbitrary and conventional distribution of power. For example, 200 years ago my ancestors went to North Africa and stole people from there, and forced them to work on plantations in appalling circumstances, so that my ancestors could enjoy and sugar, and cotton, and coffee, and stuff like that.

And then, whenever people have engaged in slavery, other people have always criticized it, and whenever those criticisms arose, the answer would very often be dressed up in terms of nature. And common sense, and also religion back in the day, so that it was argued, that said that the religious discourse might say something like, well, they were, they were mired in sin, that they hadn't heard the good news of Jesus and being mired in sin and caught up in, that they were going to go to Hell…. So, we went there, and we took them good news, and where they refused to listen we forced them to accept it, because that was our responsibility to bring them to God. And …..you, as soon as you start to say it's okay to force people to believe what you believe, it's a very short step to forcing them to work on your plantation, it seems.

Then we also had people arguing from nature and they would say, well, black people are just naturally not as able to look after themselves as white people, so the white person has to do the looking after, and left what the black person owns the white person, a debt of responsibility and so, they then have to do. I take on responsibility for their good management, so, they owe me their labor kind of thinking. And then, there were common-sense arguments people would say, well, the fact, that we can turn them into slaves, means that just common sense we will turn them into slaves. And there's something horrible being said about human nature that people will explode one another, being dressed up as common sense.

So, the three discourses and the written one hasn't been maintained as much not nature and common sense, but those three discourses have been used to disguise the arbitrary assumption of power. Just a quick and wander off into thinking about language then. So, this is Martin Heidegger and not as popular characters certainly used to be, and certainly give him some of his motive to a fairly questionable news, political activities in the mid 20-th centuries dubious, but he argues that language is like a hammer. And so, when we're using it to do stuff we don't think about it, too much, if you think about the hammer, you'll hit yourself with it, but when it breaks down, when it goes wrong, you really notice how it works. It's only when it starts to fail you that you really begin to think about what it does, when it's useful. And then there's this guy, he's a bit later than Heidegger, this is Rowland Bart, he's a French major journalist, cultural commentator, and academic, and he suggests that the photographs, and by this he means any sign, but he was particular interested in how photographs function in science.

So, he says a photograph is always invisible. We never see the photograph, what does make any sense. Because it's a perfect illustration, I already said, this is Roland Barthes, and it's not wrong advice, a photograph falls apart, but it's very hard for us to see the photograph, as a photograph we look through it, and believe we see the person, of course, we don't see the person that is not Roland Barthes at all, and just as it is a photograph, it's a photograph of a particular place and pertinent, so, if we really-really thought about it, it's just a bunch of colored lights and shades, isn't it, but this is Bart's point. The sign disappears as soon as we see it, because what we do is interpreted, as it's how we forget that the discourse analysis tries to focus on and that'll do for now I think.

SELF-ASSESSMENT FINAL TEST

The following self-assessment test will help you prepare for the exam and estimate your current knowledge of the subject. To calculate the amount of points you deserve, compare your answers with the key given below; every correct answer gives you 2 points. Here is the suggested rating scale:

0–28 – poor "F"
30–48 – poor "FX"
50–58 – satisfactory "E"
60–68 – satisfactory "D"
70–78 – good "C"
80–88 – good "B"
90–100 – excellent "A"

1) In a dictionary phonetic and grammatical information is word-related and thus essentially……..
   a) lexical   c) extralinguistic
   b) grammatical   d) syntactic

2) In theory linguistic or lexical information may be distinguished from ……… information.
   a) intralinguistic   c) factual
   b) extralinguistic   d) conceptual

3) Prepositions, determiners, or conjunctions and discourse-marking chunks such as you know, I mean, etc. are regarded as ………..words.
   a) function   c) supraphrasal
   b) notional   d) lemmatized

4) A reference work that stores and classifies such factual information as proper names of people and places, biographical data, and descriptions of historical events, political, social, and cultural institutions, etc. on all or some branches of knowledge or a single subject area is generally known as ……..
   a) bilingualized dictionary   c) encyclopedia
   b) semi-bilingual dictionary   d) bridge dictionary

5) 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 ……….
   a) handbook   c) vocabulary
   b) glossary   d) encyclopedia

6) A set of words for communication and knowledge acquisition is viewed as…….
   a) vocabulary
   b) glossary
   c) dictionary
   d) encyclopedia

7) A structured set of texts for storage and processing (a monolingual corpus, a multilingual corpus, a translation corpus (texts and their translations), etc.) is viewed as…..
   a) discourse corpus
   b) text corpus
   c) language corpora
   d) storage of lexical units

8) A style level in a language within a specific communicative situation is viewed as……
   a) discourse marker
   b) register
   c) language level
   d) speech level

9) The set of forms belonging to a particular word-class or member of a word-class is viewed as……..
   a) parameter
   b) paradigm
   c) syntagm
   d) level

10) A controlled vocabulary organizing semantic metadata for information storage and retrieval is viewed as…….
   a) encyclopedia
   b) glossary
   c) dictionary
   d) thesaurus

11) 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" in the following field of linguistic knowledge:
   a) pragmatics
   b) semantics
   c) semiotics
   d) proxemics

12) What language unit does Hoey define 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."?
   a) sentence
   b) utterance
   c) text
   d) dialogue

13) Within the discourse theory the fundamental form of a text is ….
   a) monologue
   b) utterance
   c) dialogue
   d) polylogue

14) In different discourse analysis issues discourse is viewed as….
   a) interaction
   c) text
   b) non-verbal communication
   d) utterance

15) Birmingham school of discourse analysis is famous for introducing…
   a) a hierarchical model of discourse structure
   b) a linear model of discourse structure
   c) a chaotic model of discourse structure
   d) an anarchical model of discourse structure

16) What field of linguistics aims at "replicable analysis that accounts for our ability to interpret what participants intend to convey in everyday communicative practice"?
   a) communicative linguistics
   c) interactional sociolinguistics
   b) discourse linguistics
   d) interactional psycholinguistics

17) Language is seen not as an autonomous system but as part of the wider socio-cultural context, as "social semiotic" in the plane of the following field of knowledge:
   a) linguopragmatics
   c) theory of intercultural communication
   b) systemic-functional linguistics
   d) semantics

18) 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, deals with….
   a) theories of grammatical competence and communicative competence
   b) theory of grammatical competence only
   c) theory of communicative competence only
   d) theory of intercultural communication

19) 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 is offered by:
   a) theory of intercultural communication
   c) ethnopragmatics
   b) ethnography of communication
   d) ethnolinguocultural studies

20) What 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?
   a) politeness theory
   c) face-to-face theory
   b) impoliteness theory
   d) deixis and reference theory.

21) The investigation of the structure and patterning of discourse (human speech) which contrasts explicitly with analyses of written language or of contrived examples in linguistic works is viewed as………
   a) lemmatization   c) discourse analysis
   b) text analysis   d) decoding

22) A process of adding items to a lexicon, for example words, set phrases and word patterns is viewed as……..
   a) lexicalization   c) decoding
   b) lemmatization   d) encoding

23) The techniques for examining and structuring conversations or any type of social interaction which involves spoken language are viewed as……..
   a) discourse analysis   c) convergence
   b) conversation analysis   d) divergence

24) A term referring to the environment in which an element (sound, word, phrase) occurs is viewed as……..
   a) context   c) content
   b) semantics   d) displacement

25) The process of studying digital media (texts, pictures, audio, video) and communication patterns is a systematic manner is viewed as………
   a) content analysis   c) contextualization
   b) discourse analysis   d) verbalization

26) To produce by assembling information collected from other sources means …
   a) to conduct   c) to lemmatize
   b) to compile   d) to translate

27) A word of more specific meaning than a general or superordinate term applicable to it is a …..
   a) hyponym   c) lemma
   b) homonym   d) synonym

28) A concentration on or interest in one particular area or subject is
   a) constraint   c) bias
   b) utility   d) exaggeration

29) Modern thesauruses are either arranged alphabetically or ……
   a) topically   c) have a quick search
   b) thematically   d) have an alphabetical index

30) Some scholars distinguish theory, also known as …………… or dictionary research, from practice as lexicography proper.
   a) hyperlexicography
   b) metalexicography
   c) dictionary making
   d) dictionary compiling

31) A word or phrase defined in a dictionary or entered in a word list is a ………..
   a) idiom
   b) noun
   c) hyponym
   d) lemma

32) The state of being checked, restricted, or compelled to avoid or perform some action is …..
   a) bias
   b) compilation
   c) constraint
   d) edition

33) Corpora solve the problem of observing ………...
   a) patterns of language use
   b) parts of speech
   c) grammar structures
   d) lexical units development

34) Valid applications of corpus studies depend on the design of corpora, the observational methods of analysis, and ………….
   a) the amount of words in the corpus
   b) the comparison with other corpora
   c) the language of the corpus
   d) the interpretation of the findings

35) …………… data are often especially good at distinguishing words with related propositional meanings, but different connotations and patterns of usage.
   a) Spoken
   b) Concordance
   c) Written
   d) Modern

36) A sociolinguist might use a corpus of audio-recorded conversations to study ..
   a) relations between social class and accent
   b) slips of the tongue
   c) the frequency of different phrases
   d) relations between spoken and written languages

37) Corpus data are essential for accurately describing language use, and have shown how lexis, grammar, and ……………. interact.
   a) syntax
   b) semantics
   c) pragmatics
   d) phonetics

38) What does the LOB corpus stand for?
     a) the universities of London, Oslo, & Bergen
     b) the universities of Lancaster, Oslo, & Bergen
     c) the universities of Lancaster, Oslo, & Bristol
     d) the universities of Lancaster, Ottawa, & Bergen

39) The rise and fall of the voice in speaking is ….
     a) emphasizing   c) stress
     b) connotation   d) intonation

40) What must a corpus which claims to be a balanced sample of language use represent?
     a) variables of tone, vocabulary, and topic
     b) a huge amount of nouns and verbs
     c) variables of demography, style, and topic
     d) a huge amount of sentences patterns

41) What must a corpus which claims to be a balanced sample of language use include?
     a) only written texts
     b) only spoken texts
     c) texts which have different length
     d) texts which illustrate a wide range of subject fields

42) What data does the FLOB corpus contain?
     a) American data from 1980
     b) American and British data from 1991
     c) American and British data from 2000
     d) British data from 1980

43) What does the abbreviation ICE stand for?
     a) International Corpora of English
     b) Indian Corpora of English
     c) International Complex of English
     d) Interactional Corpora of English

44) What does part-of-speech tagging allow a corpus?
     a) to be searched for lexical constructions
     b) to be searched for grammatical constructions
     c) to be searched alphabetically
     d) to be searched quickly

45) What data does the ICE contain?
     a) regional varieties of American English
     b) regional varieties of English, such as Indian and African

c) regional varieties of English, such as Irish and Welsh

d) regional varieties of English, such as Indian and Australian

46) A collection of linguistic data, either written text or transcriptions of recorded speech is …….
a) an online dictionary   c) a thesaurus
b) a linguistic corpus   d) compilation

47) A noticeable arrangement or conjoining of linguistic elements (such as words) is ……
a) a collocation   c) a text
b) a sentence   d) a lexical unit

48) A string of two or more uninterrupted word-forms which occur more than once in a text or corpus is …
a) a collocation   c) a text
b) a sentence   d) a phrase

49) A field that uses computer programs to process large amounts of data pertaining to natural language is …..
a) computer language processing   c) corpus linguistics
b) natural language processing   d) computer linguistics

50) A study of language as expressed in bodies (corpora) of written text; originated in the 1970s to advance discourse analysis is …….
a) applied linguistics   c) sociolinguistics
b) lexicography   d) corpus linguistics

THE KEY TO THE SELF-ASSESSMENT FINAL TEST

1) a	11) a	21) c	31) d	41) d
2) b	12) c	22) a	32) c	42) b
3) a	13) c	23) b	33) a	43) a
4) c	14) a	24) a	34) d	44) b
5) d	15) a	25) a	35) b	45) d
6) a	16) c	26) b	36) a	46) b
7) b	17) b	27) a	37) b	47) a
8) b	18) a	28) c	38) b	48) d
9) b	19) b	29) d	39) d	49) a
10) d	20) a	30) b	40) c	50) d

CONCISE GLOSSARY OF APPLIED LINGUISTICS TERMS

1	*accent*	1) Strictly speaking this refers to the pronunciation of a dialect, i.e. it is a reference to the collection of phonetic features which allow a speaker to be identified regionally or socially. It is frequently used to indicate that a given speaker does not speak the standard form of a language. The term is used in German to refer to grammatical features as well. 2) The stress placed on a syllable of a word or the type of stress used by a language (pressure or pitch);
2	*algorithm*	specification for performing calculation, data processing and automated reasoning tasks;
3	*analytic*	A term used for a language which tends to use free morphemes to indicate grammatical categories. Examples are Modern English and French to a certain extent. Other languages, such as Chinese or Vietnamese, are very clearly analytic and approach a relationship of one word per morpheme;
4	*application programming interface*	set of tools and resources in an operating system in order to create software applications;
5	*applied linguistics*	studies the uses to which linguistic insights can be put, especially in second language teaching;
6	*arbitrariness*	An essential notion in structural linguistics which denies any necessary relationship between linguistic signs and their referents, e.g. objects in the outside world;
7	*artificial intelligence*	design of machines capable of intelligent behavior, meaning behavior capable of achieving objectives; field originated in the 1960s, and including computational linguistics (originated in the 1950s);
8	*augmented reality*	technology superimposing a computer-generated image on a user's view of the real world; "augmentation" of the real-world environment with

computer-generated perceptual information (visual, auditory, sensory, olfactory);

| 9 | *base* | A free lexical word to which one or more endings can be added. A base can itself consist of more than one morpheme whereas a root contains only one; |

| 10 | *big data* | data sets that are too complex for standard data-processing application software, for example big data obtained by social media mining from user-generated content on social media sites and apps; |

| 11 | *bilingualism* | The ability to speak two languages with native-like competence. In every individual case one language will be dominant. Lay people often use the term if someone can simply speak a second language well; |

| 12 | *character encoding* | encoding of textual data with an encoding system such as Unicode; |

| 13 | *chatbot* | web or mobile interface used by a human being to ask questions through text, sound or video, and retrieve information from hard-coded answers or from a larger content base using machine learning; |

| 14 | *cloud database* | database on a cloud computing platform; |

| 15 | *code* | algorithm used to convert information (letter, word, sound, image, gesture) into another form of representation for communication and storage; |

| 16 | *competence* | According to Chomsky in his *Aspects of the theory of syntax* (1965) this is the abstract ability of an individual to speak the language which he/she has learned as native language in his/her childhood. The competence of a speaker is unaffected by such factors as nervousness, temporary loss of memory, speech errors, etc. These latter phenomena are entirely within the domain of *performance* which refers to the process of applying one's competence in the act of speaking. Bear in mind that competence also refers to the ability to judge if a sentence is |

grammatically well-formed; it is an unconscious ability;

17	*compiled language*	programming language whose implementations are compilers (and not interpreters);
18	*compiler*	program that transforms computer code written in one programming language into another programming language;
19	*computational linguistics*	branch of linguistics which processes natural languages using computer science and mathematics for analysis and synthesis of language and speech; originated in the 1950s with machine translation; includes applications such as spell and grammar checkers, speech synthesis, speech recognition, virtual assistants and smart speakers;
20	*computational science*	multidisciplinary field using computing capabilities for science;
21	*computational semantics*	study of computing capabilities for semantics;
22	*computer-assisted translation / cat*	language translation in which a human translator uses specific software to support and facilitate the translation process; includes translation memory, language search engines, terminology management, alignment, interactive machine translation and augmented translation;
23	*computer science*	study of computers (hardware, software, networks, internet) and computing concepts;
24	*computer vision*	theory behind the artificial systems that extract data from digital images or videos in order to process, analyze and understand such data;
25	*contact*	A term which refers to a situation in which speakers of two languages or varieties are continually in contact with each other, either due to geographical or social closeness or both. The mutual influence which results from such contact

can and does lead to changes in the structure – or at least in the lexicon – of one or both languages;

26	*content analysis*	process of studying digital media (texts, pictures, audio, video) and communication patterns is a systematic manner;
27	*context*	A term referring to the environment in which an element (sound, word, phrase) occurs. The context may determine what elements may be present, in which case one says that there are 'co-occurrence restrictions' for instance 1) /r/ may not occur after /s/ in a syllable in English, e.g. */sri:n/ is not phonotactically permissible in English; 2) the progressive form cannot occur with stative verbs, e.g. *We are knowing German* is not well-formed in English;
28	*convention*	An agreement, usually reached unconsciously by speakers in a community, that relationships are to apply between linguistic items, between these and the outside world or to apply in the use of rules in the grammar of their language;
29	*convergence*	In a general sense a process whereby two languages or varieties come to resemble each other more and more. In historical linguistics the term is often used to refer to a situation whereby two causes are taken to have led to a certain effect, e.g. where a feature in a present-day dialect is taken to derive from both substrate interference and language-internal developments;
30	*conversation analysis*	The techniques for examining and structuring conversations or any type of social interaction which involves spoken language;
31	*conversational interface*	interface that uses natural language processing (NLP) and natural language understanding (NLU) to run a conversation with a human being, for example a voice assistant;

32	*cross-linguistic*	Refers to phenomena which occur in several different languages or in investigations which draw on data from diverse languages;
33	*corpus*	Any structured and principled collection of data from a particular language – usually in electronic form, i.e. on disk – which has been compiled for the purpose of subsequent analysis. The number of corpora available has increased greatly since the spread of the personal computer in the 1980's. The most famous corpus for historical forms of English is the *Helsinki Corpus of English*;
34	*corpus linguistics*	study of language as expressed in bodies (corpora) of written text; originated in the 1970s to advance discourse analysis;
35	*code switching*	Moving from one language to another within a single sentence or phrase. This is a phenomenon found among bilinguals who feel it is appropriate to change languages (or dialects in some cases) – perhaps to say something which can only be said in the language switched to. Code-switching is governed by fairly strict rules concerning the points in a sentence at which one can change over;
36	*correctness*	An extra-linguistic notion, usually deriving from institutions in society like a language academy or a major publishing house, which attempts to lay down rigid rules for language use, especially in written form. Notions of correctness show a high degree of arbitrariness and are based on somewhat conservative usage, intended to maintain an unchanging standard in a language – a complete fiction;
37	*creativity*	An accepted feature of human language – deriving from the phenomenon of sentence generation – which accounts for speakers' ability to produce and to understand a theoretically infinite number of sentences;

38	*Creole*	A term used to describe a pidgin after it has become the mother tongue of a certain population. This development usually implies that the pidgin has become more complex grammatically and has increased its vocabulary in order to deal with the entire set of situations in which a native language is used. A well-known example is Tok Pisin, a creole spoken in Papua New Guinea and which has official status there;
39	*data analysis*	process of inspecting, cleaning, transforming and modeling data to find useful information;
40	*data modeling*	process of creating a data model for an information system by applying formal techniques;
41	*data processing*	collecting, storing, visualizing, searching, querying, analyzing, updating, sharing and transferring data;
42	*decoding*	process of converting code symbols back into information, for example information expressed in a plain natural language;
43	*dialect*	A traditional term referring to a variety of a language spoken in a certain place. There are urban and rural dialects. The boundaries between dialects are always gradual. The term *dialect* is used to denote a geographically distinct variety of a language. Two major points in this connection should be noted: 1) 'dialect' does not refer to the social or temporal aspect of language and 2) the term 'dialect' makes no reference to the standard variety of a language. In connection with the latter point it is important to stress that the standard of a language is nothing more than a dialect which achieved special political and social status at some stage in the past and which has been extensively codified orthographically;
44	*dictionary*	A reference work which offers varied information – usually arranged in alphabetical order – about words in a language, such as their spelling, pronunciation, meaning and possibly historical origins, additional

shades of meaning, typical combinations (collocations) and status vis à vis the standard of the language concerned;

45	*discourse analysis*	The investigation of the structure and patterning of discourse (human speech). It contrasts explicitly with analyses of written language or of contrived examples in linguistic works;
46	*displacement*	the referent may be removed in time or space in regards to human language;
47	*duality of patterning*	A structural principle of human language whereby larger units consist of smaller building blocks, the number of such blocks being limited but the combinations being almost infinite. For instance all words consist of combinations of a limited number of sounds, say about 40 in either English or German. Equally all sentences consist of structures from a small set with different words occupying different points in the structures allowing for virtually unlimited variety;
48	*encoding*	process of converting information into code symbols for communication and storage;
49	*error*	A characteristic mistake made by learners of a second language, usually traceable to a structural feature of their native language;
50	*hierarchy*	Any order of elements from the most central or basic to the most peripheral, e.g. a hierarchy of word classes in English would include nouns and verbs at the top and elements like adjectives and adverbs further down with conjunctions and subordinators still further down. The notions of top and bottom are intended in a metaphorical sense;
51	*hypercorrection*	A kind of linguistic situation in which a speaker overgeneralises a phenomenon which he/she does <u>not</u> have in his/her native variety. For example if a speaker from northern England pronounces *butcher*

/butʃə/ with the vowel in *but*, i.e. as /bʌtʃə/, then this is almost certainly hypercorrection as he/she does not have the *but*-sound in his/her own dialect and, in an effort to speak 'correct' English, overdoes it. The same applies to native speakers of Rhenish German when they pronounce *Kirschen* like *Kirchen* when they are talking to speakers of High German;

52	*idealisation*	A situation where the linguist chooses to ignore details of language use for reasons of greater generalisation;
53	*innateness hypothesis*	In language acquisition studies, the notion that children are born with a predisposition to learn language. It contrasts explicitly with the notion that knowledge of language is gained by experience (a view typical of behaviourism in psychology);
54	*idiolect*	The language of an individual as opposed to that of a group;
55	*interference*	The transfer of certain phenomena from one language to another where they are not considered grammatical. This may happen on an individual level (during second language learning, for example) or collectively in which case it often leads to language change;
56	*intuition*	A term referring to unconscious knowledge about his/her native language which a speaker has. Intuition is used frequently when speakers are asked to judge the grammaticality of sentences;
57	*language*	A system which consists of a set of symbols (sentences) – realised phonetically by sounds – which are used in a regular order to convey a certain meaning. Apart from these formal characteristics, definitions of languages tend to highlight other aspects such as the fact that language is used regularly by humans and that it has a powerful social function;

58	*language acquisition*	The process by which children acquire knowledge about their native language in their early childhood. Acquisition is distinguished from learning which refers to gaining knowledge of a second language in later life;
59	*language acquisition device*	A postulated pre-disposition for learning language which all humans are born with and which enables any child to learn any language in a remarkably short period of time. According to this view, the LAD consists of the structural features which are common to all languages and specific to none;
60	*language change*	A process by which developments in a language are introduced and established. Language change is continual in every language and it is largely regular. However, the rate of language change is different among different languages. It depends on a number of factors, not least on the amount of contact and informational exchange with other linguistic communities on the one hand (this tends to further change) and the degree of standardisation and universal education in the speech community on the other hand (this tends to hamper change);
61	*language contact*	A situation in which speakers of two languages intermingle. The causes of this range from invasion and deportation to voluntary emigration to a new country. The results of this intermingling depend on external factors such as the relative status of the two linguistic groups and on internal factors such as the typological similarity of the languages involved, i.e. whether their grammatical structures are comparable or not;
62	*language death*	The process by which a language ceases to exist. It is characterised by the switch over to some other language which surrounds the dying language and which is a superstratum to it, e.g. English vis à vis Manx on the Isle of Man in the middle of the present century;

63	*law*	A formulation of an ordered or predictable relationship between forms. Such laws can be diachronic or synchronic. An example of the former is Grimm's Law which states (simply) that Indo-European voiceless stops changed to corresponding fricatives at the beginning of Germanic. A synchronic law would be the devoicing of obstruents at the end of words (and syllables) in German. A law is taken to be virtually without exception;
64	*learnability*	a speaker of a language can learn any of a wide variety of languages;
65	*lemma*	dictionary form used for a set of words, for example "run" for the set of words "run", "runs", "ran" and "running";
66	*level*	A reference to a set of recognisible divisions in the structure of natural language. These divisions are largely independent of each other and are characterised by rules and regularities of organisation. Traditionally five levels are recognised: phonetics, phonology, morphology, syntax, semantics. Pragmatics may also be considered as a separate level from semantics. Furthermore levels may have subdivisions as is the case with morphology which falls into inflectional and derivational morphology (the former is concerned with grammatical endings and the latter with processes of word-formation). The term 'level' may also be taken to refer to divisions within syntax in generative grammar;
67	*lexeme*	The smallest (abstract) unit which is recognised as semantically independent in the lexicon of a language. A lexeme subsumes a set of forms which are related semantically, e.g. the lexeme walk unites the various forms *walk*, *walks*, *walked*, *walking*;
68	*lexical*	1) Pertaining to the vocabulary of a language and/ or information which is deposited in the mental lexicon of the speaker. 2) Irregular, 'quirky', not

conforming to a given pattern. This second use implies that a form cannot be derived by rule and hence it must be learned as an indivisible whole during language acquisition and stored in the lexicon in its full, unalterable form;

69	*lexical diffusion*	A type of language change in which a certain feature spreads slowly rather than establishing itself at once. Cases of lexical diffusion are characterised by incompleteness, otherwise it is not recognisable afterwards and is a case of normal change which affects the entire vocabulary. The lexical diffusion type of change usually ceases before it can cover all theoretical instances in a language, e.g. the lowering of short /u/ in the Early Modern English period which does not apply to instances before [ʃ] and after a labial stop: *bush*, *push*;
70	*lexical resource*	database offering one or several dictionaries (monolingual, bilingual, multilingual);
71	*lexicalization*	process of adding items to a lexicon, for example words, set phrases and word patterns;
72	*lexicography*	practice of compiling, writing and editing general or specialized dictionaries; study of the semantic relationships in the vocabulary (lexicon) of a natural language;
73	*lexicon*	The vocabulary of a language. It can refer to the book form of a dictionary (usually with an alphabetic listing of words) or the assumed lexicon which speakers possess mentally. The precise nature and organisation of this mental lexicon is much debated in linguistic literature as it is generally assumed to be radically different in organisation from a conventional dictionary;
74	*linguistic corpora*	collection of linguistic data, either written text or transcriptions of recorded speech;
75	*linguistic competence*	when a speaker knows the syntactic rules of a language;

76	*linked data*	structured data that are interlinked for more or better results in semantic queries;
77	*Locale*	set of parameters that defines a user's language (language identifier) and region (region identifier) in a user interface;
78	*metalanguage*	The language which is used to discuss language; see also *object language*;
79	*mistake*	An instance of incorrect usage in a foreign language which is apparently random;
80	*natural language processing*	field that uses computer programs to process large amounts of data pertaining to natural language;
81	*natural language understanding*	subfield of natural language processing for machine reading comprehension; includes search engines optimization, news gathering, text categorization, voice activation, large-scale content analysis, automated customer service and online education;
82	*natural-language user interface*	computer-human interface in which linguistic components (verbs, phrases, etc.) act as UI (user interface) controls for creating, selecting and modifying data in software applications;
83	*overextension*	A phenomenon in first language acquisition where the child uses a narrow term in a very general sense, e.g. calling all males 'papa';
84	*paradigm*	The set of forms belonging to a particular word-class or member of a word-class. A paradigm can be thought of as a vertical list of forms which can occupy a slot in a syntagm. Pronounced [ˈpærədaim];
85	*parameter*	Any aspect of language which can obtain a specific value in a given language, e.g. canonical word-order which can have the verb in a declarative sentence either before the subject, after the subject or after both subject and object. Contrast *principle* in this respect;

86	*parole*	A term deriving from Ferdinand de Saussure and which refers to language as it is spoken, contrast this with *langue*;
87	*parsing*	analyzing a string of symbols from large-scale empirical data in order to annotate the syntactic and/or semantic sentence structure and create a parsed corpus (or treebank);
88	*performance*	The actual production of language as opposed to the knowledge about the structure of one's native language which a speaker has internalised during childhood (*see* Competence);
89	*pidgin*	A language which arises from the need to communicate between two communities. Historically, and indeed in almost all cases, one of the communities is socially superior to the other. The language of the former provides the base on which the latter then creates the pidgin. A pidgin which has become the mother language of a later generation is termed a creole. Pidgins are of special interest to the linguist as they are languages which have been created from scratch and because they are not subject to the normalising influence of a standard. Classically pidgins arose during trade between European countries and those outside of Europe. The lexicon of a pidgin is usually taken from the lexifier language (the European one in question) and its grammar may derive from native input (such as the languages of West Africa during the slave trade with the Caribbean and America) or may take elements from the lexifier language or may 'invent' its own structures going on an innate blueprint which many linguists assume speakers have from birth. The further development of a pidgin is a creole, although this stage does not have to be reached if there is no necessity to develop a native language;
90	*presupposition*	Any information which is taken for granted in a discourse situation, for instance the sentence *Did*

you enjoy your breakfast? assumes that the interlocutor already had breakfast;

91	*prevarication*	speakers can intentionally make utterances that are false or meaningless in regards to human language;
92	*productivity*	novel messages may be sent; not just a fixed repertoire of calls;
93	*proposition*	A statement which can be assessed as being true or false, e.g. *The sun is shining* contains the proposition that 'the celestial body at the centre of the solar system is casting its light directly on the surface of the earth' and in any given situation this statement is either true or false;
94	*quantifier*	Any term which serves to indicate an amount such as *all*, *some*, *a few*, or the set of numerals in a language;
95	*reflexiveness*	speakers can use a language to talk about language;
96	*register*	A style level in a language. When we speak we automatically locate ourselves on a specific stylistic level. This can vary depending on the situation in which we find ourselves. For example when talking to the postman one would most likely use a different register than when one is holding a public address;
97	*second language teaching*	This is probably the main area of applied linguistics. There are many views on how a second language is *learned*, above all in comparison with the relative perfection of first language *acquisition*. Research here tends to concentrate on developing models to explain the process and ideally they should be applied to the actually teaching of foreign languages to improve results;
98	*sentiment analysis*	process that uses natural language processing (NLP) and text analysis to identify, extract, quantify and study subjective information such as users' reviews and surveys;

99	*signifiant*	A linguistic item which signifies something; contrasts with *signifié* which is what is signified. The term derives from Ferdinand de Saussure;
100	*sign language*	A communication system in which people use their hands to convey signals. In recent years sign language has been the object of linguists' attention and has come to be regarded as a fully-fledged system comparable to natural language with those individuals who are congenitally deaf and who learn sign language from childhood;
101	*sociolect*	A variety of a language which is typical of a certain class. Sociolects are most common in urban areas. In history, sociolects may play a role, e.g. in the formation of the English standard, Received Pronunciation, which derives from a city dialect (that of London in the late Middle Ages) but which has long since become a sociolect (Cockney being the dialect of London nowadays);
102	*sociolinguistics*	The study of the use of language in society. Although some writers on language had recognised the importance of social factors in linguistic behaviour it was not until the 1960's with the seminal work of Labov that the attention of large numbers of linguists was focussed on language use in a social context. In particular the successful explanation of many instances of language change helped to establish sociolinguistics as an independent sub-discipline in linguistics and led to a great impetus for research in this area;
103	*speech community*	Any identifiable and delimitable group of speakers who use a more or less unified type of language;
104	*standard*	A variety of a language which by virtue of historical accident has become the leading form of the language in a certain country. As a result of this, the standard may be expanded due to the increase in function which it experiences due to its position in society. There is nothing inherently superior

about a standard although nearly all speakers of a community accept that it has highest prestige;

105	*text corpus*	structured set of texts for storage and processing; can be for example a monolingual corpus, a multilingual corpus, a translation corpus (texts and their translations), a parallel corpus (texts alongside their translations), or a comparable corpus (texts covering the same contents);
106	*textlinguistics*	The investigation of the structure and style of texts, of pieces of language which consist of more than a single sentence;
107	*text processing*	creation and manipulation of electronic text, for example reformatting or content change (search and replace, select and move, etc.);
108	*thesaurus*	listing of words grouped according to similarity of meaning; controlled vocabulary organizing semantic metadata for information storage and retrieval;
109	*variety*	A term used to refer to any variant of a language which can be sufficiently delimited from another variant. The grounds for such differentiation may be social, historical, spatial or a combination of these. The necessity for a neutral term such as *variety* arose from the loaded use of the term *dialect*: this was not only used in the sense defined above, but also with the implication that the linguistically most interesting varieties of a language are those spoken by the older rural population. This view is understandable given the origin of dialectology in the 19th century, that is in the heydey of historical linguistics. Nowadays, sociolinguistic attitudes are prevalent and the need for a term which can include the linguistic investigation of urban populations from a social point of view became evident;
110	*vernacular*	The indigenous language or dialect of a community. This is an English term which refers to purely spoken forms of a language;

111	*virtual reality*	replacement of the user's real-world environment with a computer-generated simulation of a three-dimensional environment that be accessed with electronic equipment, for example a helmet with a screen or gloves with sensors;
112	*vocabulary*	set of words for communication and knowledge acquisition; can be for example reading vocabulary, listening vocabulary, speaking vocabulary, writing vocabulary, native language vocabulary, second language vocabulary and foreign language vocabulary;
113	*utterance*	Any stretch of spoken speech, a sentence or phrase with emphasis on the characteristics of the spoken medium in contrast either with the written form or with more abstract forms of a linguistic analysis.

REFERENCE LIST

1. Бацевич Ф. С. Прагматичні перформативи: спроба обгрунтування / Ф. С. Бацевич // Мовознавство. К.: Наукова Думка, 2008. № 1. С. 31–34.
2. Бацевич Ф. С. Смисл: сутність і сфери вияву в мові / Ф. С. Бацевич // Вісн. Львів. ун-ту. Сер. філол. Вип. 34. Ч. 1. Львів: Видавничий центр ЛНУ ім. Івана Франка, 2004. С. 346–353.
3. Жуковська В. В. Вступ до корпусної лінгвістики: навчальний посібник / В. В. Жуковська. Житомир: Вид-во ЖДУ ім. Івана Франка, 2013. 142 с.
4. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика: підручник / Є. А. Карпіловська. Донецьк: ТОВ «Юго-Восток, Лтд», 2006. 188 с.
5. Апресян Ю. Д. Интегральное описание языка и системная лексикография / Ю. Д. Апресян // Апресян Ю. Д. Избранные труды: в 2 т. М.: Языки русской культуры, 1995. Т. 2. С. 136.
6. Столнейкер Р. С. Прагматика / Р. С. Столнейкер // Новое в зарубежной лингвистике. М.: Прогресс, 1985. Вып. 16. С. 419.
7. Atwell E., Demetriou G., Hughes J., Schiffrin A., Souter C., Wilcock S. A Comparative Evaluation of Modern English Corpus Grammatical Annotation Schemes / E. Atwell, G. Demetriou, J. Hughes, A. Schiffrin, C. Souter, S. Wilcock // ICAME Journal. Bergen: De Gruyter Open, 2000. No 24. P. 7–24.
8. Beaugrande R. de Reconnecting Real Language with Real Texts: Text Linguistics and Corpus Linguistics / R. Beaugrande // International Journal of Corpus Linguistics. Amsterdam: John Benjamins Publishing Company, 1999. N. 4 (2). p. 243–59.
9. Biber D., Conrad S., Reppen R. Corpus Linguistics: [Textbook] / D. Biber, S. Conrad, R. Reppen. Cambridge: Cambridge University Press, 1998. 639 p.
10. Biber D., Johansson S., Leech G., Conrad S., Finegan E. Longman Grammar of Spoken and Written English: [Textbook] / D. Biber, S. Johansson, G. Leech, S. Conrad, E. Finegan. London: Longman, 1999. 1203 p.
11. Brown P., Levinson S. Politeness: Some Universals in Language Usage: [Monograph] / P. Brown, S. Levinson. Cambridge: Cambridge University Press, 1987. 345 p.
12. Cameron D. Working with Spoken Discourse: [Monograph] / D. Cameron. London: Sage, 2001. 206 p.
13. Davies A., Elder K. The Handbook of Applied Linguistics: [Monograph] / A. Davies, K. Elder. Oxford: Blackwell, 2004. 866 p.
14. Fillmore C. J. Corpus Linguistics or Computer-aided Armchair Linguistics: [Monograph]. In J. Svartvik (ed.), Directions in Corpus Linguistics (pp. 35–60) / C. J. Fillmore. Berlin: Mouton de Gruyter, 1992. 487 p.
15. Grundy P. Doing Pragmatics: [Textbook] / P. Grundy. London: Edward Arnold, 2000. 287 p.
16. Gumperz J. Interactional Sociolinguistics: a Personal Perspective [in D. Schiffrin, D. Tannen, H. E. Hamilton (edc.)] / The Handbook of Discourse Analysis (pp. 215–28) / J. Gumperz. Oxford: Blackwell, 2001. 874 p.

17. Hallan N. Paths to Prepositions? A Corpus-Based Study of the Acquisition of a Lexico-grammatical Category [in J. Bybee, P. Hopper (edc.)] // Frequency and the Emergence of Linguistic Structure (pp. 91–120) / N. Hallan. Amsterdam: Benjamins, 2001. 499 p.
18. Halliday M.A.K., Hasan R. Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective / M. A. K. Halliday, R. Hasan. Victoria: Deakin University Press, 1985. 126 p.
19. Hartmann R. R. K. Teaching and Researching Lexicography: [Monograph] / R. R. K. Hartmann. Harlow: Pearson Education Ltd, 2001. 211 p.
20. Hartmann R. R. K., James G. Dictionary of lexicography: [Monograph] / R. R. K. Hartmann, G. James. London: Routledge, 1998. 192 p.
21. Hunston S. Corpora in applied linguistics / S. Hunston. Cambridge: Cambridge University Press, 2002. 241 p.
22. Kennedy G. An introduction to corpus linguistics / G. Kennedy. London: Longman, 1998. 328 p.
23. Kilgarriff A. Putting frequencies in the dictionary / A. Kilgarriff // International Journal of Lexicography, 1997. N. 10 (2). p. 135–55.
24. Landau S. I. Dictionaries. The art and craft of lexicography (2nd edn.) / S. I. Landau. Cambridge: Cambridge University Press, 2001. 496 p.
25. Leech G. Principles of pragmatics / G. Leech. London: Longman, 1983. 264 p.
26. Leech G., Rayson P., Wilson A. Word frequencies in written and spoken English / G. Leech, P. Rayson, A. Wilson. London: Longman, 2001. 320 p.
27. Levinson S. Pragmatics / S. Levinson. Cambridge: Cambridge University Press, 1983. 436 p.
28. McArthur T. Worlds of reference: lexicography, learning and language from the clay tablet to the computer / T. McArthur. Cambridge: Cambridge University Press, 1986. 244 p.
29. Mey J. Pragmatics: an introduction / J. Mey. Oxford: Blackwell, 1993. 392 p.
30. Pennycook A. Incommensurable discourses? / A. Pennycook // Applied Linguistics, 1994. N. 15 (2). p. 115–38.
31. Sampson G. English for the computer / G. Sampson. Oxford: Clarendon, 1995. 167 p.
32. Saville-Troike M. Ethnography of communication: an introduction (3rd edn.) / M. Saville-Troike. Oxford: Blackwell, 2003. 215 p.
33. Schiffrin D. Approaches to discourse / D. Schiffrin. Oxford: Blackwell, 1994. 187 p.
34. Searle J. R. Speech acts: an essay in the philosophy of language / J. R. Searle. Cambridge: Cambridge University Press, 1969. 203 p.
35. Sinclair J. Corpus concordance collocation / J. Sinclair. Oxford: Oxford University Press, 1991. 179 p.
36. Sledd J., Ebbit, W. R. Dictionaries and that dictionary: a casebook on the aims of lexicographers and the targets of reviewers / J. Sledd, W. R. Ebbit. Glenview: Scott, Foresman & Co, 1962. 273 p.
37. Sperber D., Wilson D. Relevance: communication and cognition (2nd edn.) / D. Sperber, D. Wilson, Oxford: Blackwell, 1995. 338 p.

38. Thomas J. Meaning in interaction / J. Thomas. London: Longman, 1995. 224 p.
39. Widdowson H. G. Discourse analysis: a critical view / H. G. Widdowson // Language and Literature, 1995. N. 4 (3). p. 157–72.
40. Wilson E. O. Consilience: the unity of knowledge / E. O. Wilson. London: Little, Brown & Co, 1998. 382 p.

On-line resources

1. A Glossary of Applied Linguistics | Taylor & Francis Group [Electronic resource]. Retrieved from https://www.taylorfrancis.com/books/9781135468309
2. Computational linguistics – short glossary [Electronic resource]. Retrieved from https://marielebert.wordpress.com/2018/10/17/glossary/
3. Corpus Linguistics Glossary – Kent State University [Electronic resource]. Retrieved from https://www.kent.edu/appling/corpus-linguistics-glossary
4. Cambridge Dictionary – Cambridge University Press [Electronic resource]. Retrieved from: https://dictionary.cambridge.org/
5. Macmillan Dictionary. Free English Dictionary and Thesaurus Online [Electronic resource]. Retrieved from: https://www.macmillandictionary.com/
6. Merriam Webster On-Line Dictionary [Electronic resource]. Retrieved from: http://www.merriam-webster.com.
7. Oxford English Dictionary [Electronic resource]. Retrieved from: http://www.oed.com