



Міністерство освіти і науки України  
Сумський державний університет

**А. В. Прокопенко,  
Л. І. Гарцунова**

**ОСНОВИ ІНФОРМАТИКИ  
ТА ПРИКЛАДНОЇ ЛІНГВІСТИКИ  
ЗІ ЗМІСТОВИМ МОДУЛЕМ:  
КОПРАЙТИНГ**

Конспект лекцій

Суми  
Сумський державний університет  
2020

Міністерство освіти і науки України  
Сумський державний університет

# **ОСНОВИ ІНФОРМАТИКИ ТА ПРИКЛАДНОЇ ЛІНГВІСТИКИ ЗІ ЗМІСТОВИМ МОДУЛЕМ: КОПРАЙТИНГ**

Конспект лекцій

для студентів спеціальності 035 «Філологія»  
денної та заочної форм навчання

Затверджено  
на засіданні кафедри  
германської філології  
як конспект лекцій  
із дисципліни «Основи інформатики  
та прикладної лінгвістики  
зі змістовим модулем: копірайтинг».  
Протокол № 9 від 07.04.2020 р.



Суми  
Сумський державний університет  
2020

Основи інформатики та прикладної лінгвістики зі змістовим модулем: копірайтинг : конспект лекцій / укладачі : А. В. Прокопенко, Л. І. Гарцунова. – Суми : Сумський державний університет, 2020. – 108 с.

Кафедра германської філології

<b>ЗМІСТ</b>		<b>С.</b>
<b>Вступ</b> .....		5
<b>ОСНОВИ ІНФОРМАТИКИ ТА ПРИКЛАДНОЇ ЛІНГВІСТИКИ</b>		
<b>Лекція 1.</b> Прикладна лінгвістика як галузь мовознавства...		6
<b>Лекція 2.</b> Інформація та інформатика.....		11
<b>Лекція 3.</b> Комп'ютерна лінгвістика як прикладна лінгвістична дисципліна.....		20
<b>Лекція 4.</b> Лексикографія як один із важливих напрямків прикладної лінгвістики.....		27
<b>Лекція 5.</b> Електронні корпуси текстів і переклад.....		34
<b>Лекція 6.</b> Природний і штучний інтелект.....		42
<b>Лекція 7.</b> Системи машинного перекладу.....		48
<b>Лекція 8.</b> Системи автоматизованого перекладу.....		54
<b>КОПРАЙТИНГ</b>		
<b>Лекція 9.</b> Вступ. Основні поняття копірайтингу. Переваги копірайтингу.....		67
<b>Лекція 10.</b> Розуміння процесу копірайтингу. Робочі сфери копірайтера. Кліше.....		73
<b>Лекція 11.</b> Елементи реклами.....		78
<b>Лекція 12.</b> Перше речення. Середовище копірайтера.....		84
<b>Лекція 13.</b> Емоції в рекламі.....		88
<b>Лекція 14.</b> Деякі принципи редагування рекламного тексту.....		93
<b>Лекція 15.</b> Концепти, які необхідно враховувати під час створення реклами.....		98
<b>Лекція 16.</b> Психологічні тригери. Рівень якості. Сім кроків до написання чудового рекламного тексту.....		104

## ВСТУП

Конспект лекцій призначений для лекційних занять із дисципліни «Основи інформатики та прикладної лінгвістики зі змістовим модулем: копірайтинг» для студентів спеціальності 035 «Філологія» денної та заочної форм навчання. Дисципліна «Основи інформатики та прикладної лінгвістики зі змістовим модулем: копірайтинг» буде цікавою для студентів, які прагнуть отримати знання щодо основних принципів роботи хмарних сервісів, принципу дії ТМ-інструментів, особливостей процесу локалізації та копірайтингу, а також набути навичок необхідних для ефективного використання засобів сучасних інформаційних технологій у майбутній професійній діяльності.

Основною метою вивчення дисципліни є формування системи знань про прикладну лінгвістику, а також про основні поняття і процес копірайтингу.

# ЛЕКЦІЯ 1

## ПРИКЛАДНА ЛІНГВІСТИКА ЯК ГАЛУЗЬ МОВОЗНАВСТВА

**1. Теоретична та прикладна (практична) лінгвістика. Об'єкт, предмет і завдання прикладної лінгвістики.**

**2. Основні наукові напрями і методи прикладної лінгвістики.**

**3. Актуальні проблеми прикладної лінгвістики. Зв'язок лінгвістики з іншими науками.**

**1. Теоретична лінгвістика** досліджує мовні закони та формулює їх як теорії, акумулює та ставить під сумнів висновки щодо мови. Вона буває:

– *емпіричною (дескриптивною)*: описує реальне мовлення, цікавиться мовною поведінкою та знаннями людей через те, що вони говорять і як вони це роблять. У дескриптивній лінгвістиці немає поняття «правильно-неправильно»;

– *нормативною (прескриптивною, регулятивною)*: вказує, як «потрібно» говорити і писати, визначає стандартний варіант мови для функціонування в головних сферах життя держави: освіта, уряд, торгівля, право, спілкування між людьми, розділеними великими відстанями, з великими культурними відмінностями та відстанню в часі. Функціонує через створення граматик, словників і підручників зі стилю.

**Прикладна лінгвістика** є повноправним розділом мовознавства, спрямованим на розв'язання практичних завдань різних галузей науки й техніки, повсякденного життя людини, суспільства на підставі теоретичного доробку досліджень мови й мовлення.

**Об'єктом** прикладної лінгвістики є практичні завдання, для розв'язання яких потрібно залучення доробку мовознавства, а **предметом** – способи їхнього розв'язання.

Прикладна лінгвістика також розробляє методи розв'язання практичних завдань, пов'язаних із використанням

мови. Визначимо **коло основних завдань** прикладної лінгвістики:

1. Переклад з / на іноземну мову.
2. Навчання іноземної мови (методики навчання), лінгводидактика.
3. Комунікація за допомогою технічних засобів.
4. Створення штучних мов.
5. Інформаційний пошук.
6. Анотація і реферування текстів.
7. Складання словників (практична лексикографія).
8. Упорядкування, стандартизація та уніфікація науково-технічної термінології.
9. Організація бібліографічної інформації.
10. Ефективний друкарський набір.
11. Транскрибування усного мовлення.

Сьогодні впровадження нових інформаційних технологій в усі сфери людського спілкування прикладна лінгвістика розвивається в напрямку автоматизації основних завдань, а саме:

1. Машинний переклад + машинні словники.
2. Комп'ютерна лінгводидактика (CALL – Computer Assisted Language Learning).
3. Комп'ютерна лінгвістика (всі додатки лінгвістики в комп'ютерних середовищах).
4. Математична лінгвістика (розроблення формальних моделей мов).
5. Автоматичне оброблення природних мов:
  - а) розпізнавання і синтез мови,
  - б) автоматизація інформаційних робіт,
  - в) автоматичні системи інформаційного пошуку.
6. Квантитативна лінгвістика (частотний аналіз текстів).

Термін «прикладна лінгвістика» з'явився наприкінці 20-х років ХХ століття. Термін «прикладна лінгвістика» багатозначний. У вітчизняній і західній лінгвістиці він має абсолютно різні інтерпретації. У західній лінгвістиці (*applied*

*linguistics, angewandte Linguistik*) він зв'язується насамперед у викладанням іноземних мов, зокрема методику викладання, особливості опису граматики для навчальних цілей, викладання мови як рідної та іноземної тощо.

У СРСР термін «прикладна лінгвістика» став широко вживатися в 50-ті рр. ХХ ст. у зв'язку з розробленням комп'ютерних технологій і появою систем автоматичного оброблення інформації (автоматизованих систем управління, систем інформаційного пошуку, автоматизованих систем оброблення тексту та ін.).

2. Одним з найбільш поширених **напрямів прикладної лінгвістики** є традиційний напрям використання природної мови (як рідної, так і іноземних) у процесі комунікації людей. Цей напрям має ще назву конкретного мовознавства, тобто такого, що вивчає окремі (конкретні) мови (україністика, русистика тощо). Функціонування кожної мови як засобу спілкування вивчається в межах функціональної лінгвістики, яка, на відміну від конструктивної лінгвістики, що вивчає мову, надає перевагу дослідженню мовлення.

Загальні проблеми прикладної лінгвістики досліджували вітчизняні такі мовознавці: Н. Бардіна (прикладна лінгвістика), Є. Карпіловська (комп'ютерна лінгвістика), О. Гринбаум (комп'ютерні аспекти стилеметрії), Л. Бондарко, Л. Вербицька (синтаксичні структури і стилістична діагностика), Р. Кобрин (лінгвістика в картографії), Г. Мартиненко, С. Чебанов (стилеметрія), З. Партико, Е. Савенкова, О. Селіванова, Б. Коваль, Н. Городецький та інші.

Серед зарубіжних дослідників питаннями прикладної лінгвістики займалися: Х. Фергюсон, В. Відовсон, М. Холідей, У. Чейф, Т. Гівон, Дж. Гринберг (функціональна та типологічна теорії лінгвістики), У. Лабов, Д. Хаймз, Дж. Гамперц, Дж. Фішман, Л. Мілрой (антропологічна лінгвістика, соціолінгвістика), В. Грабе, Р. Каплан (корпусна лінгвістика).

Розвиток лінгвістичних досліджень останніх десятиліть характеризується міждисциплінарною взаємодією з іншими



науками (кібернетика, інформатика, семіотика, соціологія, психологія тощо). Інтегрованість зусиль учених зумовила розвиток таких комплексних наук, як соціолінгвістика, психолінгвістика, нейролінгвістика, когнітивна лінгвістика, математична лінгвістика, кібернетична лінгвістика, комп'ютерна лінгвістика тощо.

Кожну з цих наук доцільно розглядати як окремий напрям прикладної лінгвістики, який здатен дати суттєвий поштовх розвитку новітніх технологій, науково-технічному прогресу загалом.

До **методів** прикладної лінгвістики відносять:

- метод класифікації (виявлення тієї сітки параметрів, яка дозволяє охопити всі релевантні (у теорії) властивості мовних структур);
- метод моделювання.

**Метод моделювання** використовується в тих випадках, коли безпосереднє вивчення об'єкта моделювання щодо того чи іншого ускладнено.

Прикладна лінгвістика використовує як метод класифікації, так і метод моделювання. Однак оскільки завдання прикладної лінгвістики зосереджені у сфері оптимізації функцій мови, а оптимізація визначається конкретним завданням, то найважливішою властивістю методів прикладної лінгвістики є **оптимізація**. Оптимізація – такий опис (модель) проблемної області, при якому ця область зберігає в результатуючому поданні тільки ті суттєві властивості, які необхідні для даної практичної задачі. Іншими словами, якщо для теоретичного дослідження передбачається повний опис проблемної області з усіма її труднощами і тощо, то прикладний оптимізований опис має бути задовільним тільки для цього конкретного завдання.

3. Якщо вибрати ключові слова, що характеризують новітні галузі практичного життя людства, то це, звичайно, будуть терміни «комп'ютерні технології», «комунікація», «суспільство». Приклад об'єднання цих концептів – Інтернет,

який став проявом нового інформаційного середовища ХХІ століття. Гіпертекстові технології, безпосередньо пов'язані з експлуатацією і розвитком інтернету, претендують на те, щоб стати найважливішою сферою діяльності лінгвістів.

З комп'ютерними технологіями пов'язана проблематика, що має істотний лінгвістичний потенціал – комп'ютерний дизайн тексту і його компонентів, зокрема шрифту. Наявне програмне забезпечення дозволяє виводити на монітор комп'ютера всю смугу друкованого тексту книги, журналу, газети, рекламного буклета та компоувати текст і образотворчі елементи оригінал-макету в єдине ціле. У цьому разі текст є елементом зображення, а зображення – частиною тексту. Чисто практичні принципи побудови оригінал-макету видання повинні підтримуватися дослідженнями в галузі лінгвістики і семіотики, які виявляють закони поєднання змісту тексту, зображення, семантики гарнітури (шрифту) і способів виділення компонентів тексту (курсив, розрядка, відступ тощо).

Наразі з'явився відчутний попит на лінгвістів, що працюють у галузі реклами і публічної політики. Ці два напрямки об'єдналися в діяльності різних піарівських (від англ. Public Relations) агентств, які надають послуги з просування на «політичному ринку» як окремих політиків, так і політичних партій і рухів. Тут мається на увазі організація передвиборчих кампаній, зв'язки з громадськістю, моніторинг стану суспільної свідомості тощо. Перспективи розвитку в нашій країні лінгвістичних додатків у рекламі і практичній політиці дуже значні.

Триває підвищення інтересу до систем машинного перекладу, не стоїть на місці сучасна лексикографія, яка приваблює комп'ютерні способи оброблення і зберігання мовних даних. Створюються автоматичні словники для різних типів користувачів. Активні дослідження в практичній соціології стимулюють розвиток соціолінгвістики.

У широкому сенсі прикладна лінгвістика – це діяльність щодо застосування наукових знань про будову та

функціонування мови в нелінгвістичних наукових дисциплінах та в різних сферах практичної діяльності людини, теоретичне осмислення цієї діяльності [1].

### **Контрольні питання**

1. В чому полягає відмінність між теоретичною і прикладною лінгвістикою?
2. Які цікаві факти з історії прикладної лінгвістики вам відомі?
3. У чому полягають напрями і завдання прикладної лінгвістики?
4. Які методи використовуються у прикладній лінгвістиці?
5. Які проблеми прикладної лінгвістики є актуальними в наш час?

### **Список використаної літератури**

1. Баранов А. Н. Введение в прикладную лингвистику : учеб. пос. / А. Н. Баранов. – Москва : Наука, 2001. – 194 с.
2. Волошин В. Г. Комп'ютерна лінгвістика : навч. посібник / В. Г. Волошин. – Суми : Університетська книга, 2004. – 382 с.

## **ЛЕКЦІЯ 2**

### **ІНФОРМАТИКА ТА ІНФОРМАЦІЯ**

1. Семіотика. Знак і теорія знаків. Мова як мовна система.
2. Інформація. Формалізація інформації внутрішня і зовнішня.
3. Інформатика. Кодування інформації в комп'ютері. Системи кодування.
4. Формати файлів.

1. Семіотика («*sema*» – *значення*) – це наука про загальну теорію знака, що досліджує будь-які знакові системи як засоби позначення і передання значення або інформації. Як наука семіотика оформилася лише в 1974 році.

Лінгвістику цікавлять загальні положення семіотики про знаки, їхні розпізнавальні ознаки, способи класифікації знаків, комбінація їх у систему для конкретних завдань.

Знак матеріальний і має спрямоване значення.

У знаку виділяють 2 основні аспекти (або «плани»):

- план висловлювання (матеріал, як виглядає);
- план змісту (щось позначає, сенс).

Знак – це член певної знакової системи. Існує кілька основних типів знакових систем [2]:

1) натуральні або природні; 2) іконічні, 3) системи художніх образів; 4) системи мовних знаків; 5) системи письмових знаків; 6) формалізовані або кодові знакові системи.

Наприклад, всім відомі такі знакові системи як дорожні, телеграфні коди, сигнальні (прапорці), креслярські позначення, шифри, комп'ютерні піктограми тощо.

У мові виділяють 3 семіотичних аспекти знака:

1. Синтактика (вивчає відносини між знаками).
2. Семантика (відносини між знаком і значущим).

3. Прагматика (відносини між знаками і тими, хто їх використовує).

Повернемося до визначення семіотики. Показавши, що мова – це деяка система знаків, і кажучи, що мова – це засіб комунікації (основне визначення в прикладній лінгвістиці), або засіб передання значення (визначення семіотики), ми знаходимо зв'язок лінгвістики і семіотики. Крім того, це означає, що мова є засобом передання інформації, тим самим ми встановлюємо зв'язок із такою перспективною наукою, як інформатика.

2. Отже, **інформацію** в найширшому значенні цього слова можна визначити як сигнал у будь-якій формі про якийсь об'єкт або явище, сигнал, заразом такий, що служить виразним

їхнім образом, символом , що дозволяє розпізнати такий об'єкт або явище, однозначно вирізнити їх з-поміж інших подібних і неподібних. Вочевидь, саме таке трактування найповніше відповідає значенню етимона цієї запозиченої з латини лексеми, адже «*īnfōrmātio*» означає «роз'яснення, пояснення; виклад, поняття».

Український учений А. О. Білецький у своїй оригінальній *лінгвосеміотичній теорії* запропонував детальну класифікацію типів інформації, врахувавши різноманітні чинники, умови її творення та різновиди застосування (табл. 1):

**Таблиця 1 – Типи інформації за класифікацією А. О. Білецького**

<b>Ознака</b>	<b>Тип інформації</b>
1. Джерело створення	антропогенна – фізіогенна
2. Форма	кодована (системна) – некодована (одинична)
3. Зв'язок із ситуацією	ситуаційна – екстраситуаційна
4. Спосіб оформлення	мовна (вербальна) – немовна (екстравербальна)
5. Сфера призначення	побутова (узуальна) – спеціальна
6. Характер змістового навантаження	логічна – естетична

У теорії А. О. Білецького також вичерпно окреслений спектр функцій, виконуваних мовною системою як спеціалізованою системою обміну інформацією або різновидом інформаційної системи. Таких функцій А. О. Білецький виділив сім:

– **трансформаційна** (від лат. *transformo* «перетворюю, змінюю форму») – перетворення інформації;

– **консервативна** (від лат. *cōnservo* «зберігаю») – зберігання інформації;

- **транслятивна** (від лат. *trānslātio* «перенос, переміщення») – передавання інформації;
- **дименсійна** (від лат. *dīmēnsio* «вимірювання, розмір») – вимірювання інформації;
- **утилітарна** (від лат. *ūtilitās* «користь, придатність») – використання інформації;
- **естетична** (від грецьк. *αἰσθητικός* «чуттєво сприйманий») – вибір для оформлення та передавання інформації засобів, оптимальних для її сприйняття;
- **метасемантична** (від грецьк. *μετα* «зверх, понад» та *σημασι* – *κόσ* «означальний») – відображення в оформленні інформації додаткового змісту, який впливає на відчуття людини, стимулює в неї певні реакції на інформацію (пор., наприклад, різні способи оформлення прохання: наказ, уклінне прохання, дружнє спонукування до дії, нейтральний заклик, побажання тощо).

Для комп'ютерного опрацювання інформація має бути подана у придатній, доступній для «розуміння» комп'ютера формі. Отже, істотного значення набуває форма запису інформації, або, інакше, мова її запису. Розрізняють:

- 1) **формалізацію внутрішню**, або **формалізацію системи мови, створення моделей мовних об'єктів**;
- 2) **формалізацію зовнішню**, або **формалізацію моделей мовних об'єктів**, створення моделей, тобто такої форми їхнього подання, яка була б доступна комп'ютеру та іншим технічним засобам опрацювання інформації.

**Інформаційна модель** є структурованим і поданим у формалізованому вигляді описом окремого об'єкта або й предметної галузі загалом, який дозволяє одержати інформацію про них за допомогою комп'ютера. Лінгвістичні моделі становлять різновид інформаційних моделей [1].

3. Під **інформатикою** мають на увазі науку про закономірності запису, зберігання, перероблення, передання та використання інформації за допомогою технічних засобів.

Фахівці з інформатики, визначаючи **предмет дослідження** цієї наукової галузі як «технологію побудови, аналізу та використання людино-комп'ютерного (програмного) знання», основне її завдання вбачають у розробленні засобів і методів побудови, аналізу та узагальнення інформаційних моделей предметних галузей. Головним **завданням інформатики** є вивчення закономірностей, відповідно до яких відбувається створення, перетворення, збереження, передання і використання інформації всіх видів, зокрема із застосуванням сучасних технічних можливостей.

Основні **поняття в інформатиці** – це код, система умовних знаків або символів; алфавіт – набір знаків коду, текст – послідовність знаків заданого повідомлення.

В інформатиці існує такий термін, як «кодування інформації» – це репрезентування повідомлення у формі, зручній для передання заданим каналом. Більшість кодів для кодування інформації засноване на тих чи інших системах числення. **Система числення** – це спосіб запису чисел за допомогою заданого набору спеціальних знаків і відповідні їйому правила виконання дій над числами.

Будь-якому дискретному повідомленню, букві повідомлення можна приписати будь-який порядковий номер. Водночас передання і зберігання інформації – це робота з числами. Отже, з технічного погляду, інформація – це сигнал або елементарна подія «так» – «ні», яка вимірюється в двійкових одиницях 0 і 1. Ці знаки називаються двійковими цифрами, по-англійськи – *binary digit*, або, скорочено *bit* (*біт*). У сучасній обчислювальній техніці біти заведено об'єднувати у вісімки, які називаються **байтами**: 1 байт = 8 біт. Поряд із бітами і байтами використовують і більші одиниці вимірювання інформації:

1 Кбайт = 210 байт = 1024 байт;

1 Мбайт = 210 Кбайт = 1024 Кбайт = 220 байт;

1 Гбайт = 210 Мбайт = 1024 Мбайт = 230 байт;

1 Тбайт = 210 Гбайт = 1024 Гбайт = 240 байт;

1 Пбайт = 210 Тбайт = 1024 Тбайт = 250 байт.

У двійковій системі числення є лише дві цифри – 0 і 1. Вага розряду числа залежить від положення цього розряду в числі, а розряди двійкового числа домножуються на двійку у відповідному ступені:

$$10\ 101_2 = 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$$

$$10\ 111_2 = 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$$

$$111_2 = 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0$$

Окрім двійкової системи числення, в інформатиці використовують:

- вісімкову систему числення із цифрами від 0 до семи;
- шістнадцяткову систему числення із цифрами від 0 до F (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A[10], B[11], C [12], D[13], E[14], F[15]).

**Швидкість передавання інформації** вимірюється кількістю бітів, що передаються за одну секунду. Швидкість передавання біт за одну секунду називається 1 бодом. Похідні одиниці швидкості передачі називаються Кбод, Мбод і Гбод:

$$1\ \text{Кбод (один кілобод)} = 210\ \text{бод} = 1024\ \text{біт/с};$$

$$1\ \text{Мбод (один мегабод)} = 220\ \text{бод} = 1024\ \text{Кбод};$$

$$1\ \text{Гбод (один гігабод)} = 230\ \text{бод} = 1024\ \text{Мбод}.$$

**Інформаційним об'ємом повідомлення** називається кількість бітів у цьому повідомленні. Підрахунок інформаційного об'єму повідомлення є суто технічним завданням, оскільки за такого підрахунку зміст повідомлення не відіграє жодної ролі.

Для того щоб весь світ однаково кодував текстові дані, потрібні єдині таблиці кодування. На теперішній час існує кілька систем кодування. Найпоширенішими є такі: ASCII, Windows1251, KOI8, ISO.

Для англійської мови Інститут стандартизації США (ANSI – American National Standard Institute) ввів у дію систему кодування ASCII (American Standard Code for Information Interchange – стандартний код інформаційного обміну). Для запису одного символу в пам'яті виділяється один байт (8 біт).



В одному байті можна записати 256 різних кодів, які можна вважати цілими числами від 0 до 255. У системі ASCII закріплені 2 таблиці кодування: базова й розширена. Базова таблиця закріплює значення кодів від 0 до 127, розширена від 128 до 255.

**Таблиця 2 – Базова таблиця кодування ASCII**

32 пропуск	48 0	64 @	80 P	96 `	112 p
33 !	49 1	65 A	81 Q	97 a	113 q
34 "	50 2	66 B	82 R	98 b	114 r
35 #	51 3	67 C	83 S	99 c	115 s
36 \$	52 4	68 D	84 T	100 d	116 t
37 %	53 5	69 E	85 U	101 e	117 u
38 &	54 6	70 F	86 V	102 f	118 v
39 '	55 7	71 G	87 W	103 g	119 w
40 (	56 8	72 H	88 X	104 h	120 x
41 )	57 9	73 I	89 Y	105 i	121 y
42 *	58 :	74 J	90 Z	106 j	122 z
43 +	59 ;	75 K	91 [	107 k	123 {
44 ,	60 <	76 L	92 \	108 l	124
45 -	61 =	77 M	93 ]	109 m	125 }
46 .	62 >	78 N	94 ^	110 n	126 ~
47 /	63 ?	79 O	95 _	111 o	127 •

У перших 32 кодах (0–31) розміщуються так звані керівні коди, яким не відповідають жодні символи мов і, відповідно, ці коди не виводяться ні на екран, ні на пристрої друку. Серед них 13 – код кінця рядка, 32 – код прогалини, 36 – код кінця файлу.

Починаючи з коду 32 до коду 127 розміщені коди символів англійського алфавіту. Символи національного алфавіту розміщені в кодах від 128 до 255. Кодування Windows1251 стала стандартом в українському секторі Wold Wide Web.

**Таблиця 3 – Таблиця кодування Windows 1251**

128 Ъ	144 ђ	160	176°	192 А	208 Р	224 а	240 р
129 Ѓ	145 ‘	161 Ў	177 ±	193 Б	209 С	225 б	241 с
130,	146 ’	162 ў	178 І	194 В	210 Т	226 в	242 т
131 ґ	147 “	163 Ј	179 і	195 Г	211 У	227 г	243 у
132 „	148 „	164 џ	180 г	196 Д	212 Ф	228 д	244 ф
133.	149	165 Ѓ	181 μ	197 Е	213 Х	229 е	245 х
134 †	150 -	166 †	182 ¶	198 Ж	214 Ц	230 ж	246 ц
135 ‡	151 -	167 §	183 ·	199 З	215 Ч	231 з	247 ч
136?	152 □	168 Ѓ	184 ё	200 И	216 Ш	232 и	248 ш
137 ‰	153 ™	169 ©	185 №	201 Ё	217 Щ	233 й	249 щ
138 Љ	154 љ	170 €	186 €	202 К	218 Ъ	234 к	250 ъ
139 <	155 >	171 " "	187 " "	203 Л	219 Ы	235 л	251 ы
140 Њ	156 \	172 ¬	188 j	204 М	220 Ь	236 м	252 ь
141 Ќ	157:	173 -	189 S	205 Н	221 Э	237 н	253 э
142 Ћ	158 ħ	174 ®	190 s	206 О	222 Ю	238 о	254 ю
143 Џ	159 ц	175 Ī	191 ĩ	207 П	223 Я	239 п	255 я

**КОИ8** (код обміну інформацією) є стандартним кодуванням у повідомленнях електронної пошти й телеконференцій.

**ISO** (International Standard Organization) – міжнародний стандарт. Це кодування використовується рідко [2].

#### 4. Формати файлів текстових документів:

**RTF** (англійською Rich Text Format – «формат збагаченого тексту») – текст з форматуванням і вставленими об’єктами;

**DOC** – текст із форматуванням і вставленими об’єктами, основний формат Microsoft Word 97/2000/XP/2003;

**DOCX** – текст із форматуванням і вставленими об’єктами, основний формат Microsoft Word 2007/2010;

**ODT** – текст із форматуванням і вставленими об’єктами, основний формат LibreOfficeWriter;

**PDF** – формат для перегляду та друкування з незмінним вихідним форматуванням для різних платформ;

**TXT** – текст із розбиттям на абзаци без форматування;

HTML – (англ. HyperText Markup Language – мова розмітки гіпертекстових документів) – стандартна мова розмітки веб-сторінок в інтернеті і відповідне розширення файлу;

XML (англ. Extensible Markup Language – розширювана мова розмітки) – стандарт побудови мов розмітки ієрархічно структурованих даних для обміну між різними застосунками, зокрема, через інтернет і відповідне розширення файлу.

### **Контрольні питання**

1. Що таке семіотичний трикутник?
2. Що розуміють під інформацією в різних науках?
3. Які типи інформації існують згідно лінгвосеміотичній теорії А. О. Білецького?
4. Що таке інформаційна модель?
5. Як кодується текст?
6. Для чого використовують кодові таблиці ASCII, Windows1251?
8. Які існують формати файлів текстових документів?
9. Яка є різниця між неформатованим та форматуваним текстом?
10. Які формати використовують для перегляду готових текстових документів?

### **Список використаної літератури**

1. Карпіловська Є. А. Вступ до прикладної лінгвістики : комп'ютерна лінгвістика : підручник / Є. А. Карпіловська. – Донецьк : Юго-Восток, 2006. – 188 с.
2. Здобін Г. Г. Основи інформатики, комп'ютерної техніки і комп'ютерних технологій : підручник / Г. Г. Здобін. – Київ : Каравела, 2007. – 240 с.

## ЛЕКЦІЯ 3

### КОМП'ЮТЕРНА ЛІНГВІСТИКА

1. Комп'ютерна лінгвістика (КЛ) як прикладна лінгвістична дисципліна. Об'єкт, предмет і завдання КЛ.
2. Зв'язок КЛ з іншими науковими дисциплінами.
3. Мови програмування як штучні мови. Мова гіпертекстової розмітки HTML.
4. Поняття метамови. БНФ-нотації.

1. **Комп'ютерна лінгвістика (КЛ)** – самостійна лінгвістична дисципліна, яка розв'язує теоретичні й прикладні завдання мовознавства за допомогою комп'ютера і вивчає лінгвістичні основи інформатики та всі аспекти зв'язку мови і мислення, моделювання мови та мислення в комп'ютерному середовищі за допомогою комп'ютерних програм. Становлення комп'ютерної лінгвістики відбувалося в 50-ті рр. ХХ ст. на базі комп'ютерної науки (обґрунтування та створення перших кібернетичних систем Н. Вінером і Дж. Фон Нейманом у 40-ті рр.), теорії інформації (К. Шеннон, В. Вівер) і структурно-математичної лінгвістики.

Як окремий напрям прикладного мовознавства КЛ оформилася в 60-ті роки ХХ ст. Першими її лінгвістичними орієнтирами були розробки трансформаційного й дистрибутивного методів З. Харріса, генеративна концепція Н. Хомського, формальні граматики генеративного типу, спроби машинного перекладу та створення інформаційно-пошукових систем і мов.

**Об'єкт дослідження КЛ** становить мова в усіх трьох способах свого існування: мовна система; мовлення; мовна діяльність.

**Предмет дослідження КЛ** – ознаки будови, змісту та функціонування одиниць мовної системи, продуктів мовлення та мовної діяльності – звукових або письмових текстів, які

могли б служити для їхнього моделювання й використання в процесах комп'ютерного опрацювання мовної інформації.

Перед комп'ютерною лінгвістикою стоять такі **завдання**:

- аналіз основних результатів наукових досліджень у галузі розв'язання інтелектуальних лінгвістичних систем;
- проблеми розв'язання і стан машинного перекладу;
- експертні системи і роботизація;
- аналіз і синтез мовлення;
- моделювання вводу і виводу різних типів текстів.

2. КЛ тісно пов'язана з проблемами реєстрації та автоматичного пошуку документальної інформації на електронних носіях. Це актуальне завдання для сфери керування, промисловості, науки і техніки, де є проблема автоматичної переробки інформації у сфері довідково-бібліографічних, інформаційно-пошукових систем, технологій; юридичної, економічної, історичної, політичної тощо.

КЛ тісно пов'язана з педагогікою, зокрема з комп'ютерним навчанням. Впровадження програмного навчання у викладання мов із застосуванням комп'ютерних технологій вирішує проблему інтенсифікації навчання.

КЛ сприяє розробці проблем створення штучного інтелекту. Вирішення цієї проблеми залежить від можливостей комп'ютеризації та інформатизації. КЛ прагне будувати моделі, що відображають конкретні процеси мовленнєво-розумової діяльності людини в конкретній ситуації.

КЛ як складова науки про мову вивчає лінгвістичні основи інформатики і всі аспекти зв'язку мови, мислення як форми думки і моделювання мовної дійсності за допомогою комп'ютерних програм. КЛ досліджує проблеми мовних одиниць – менших за слово, які дорівнюють слову, словосполученню, реченню (висловлюванню), тексту загалом, а також проблеми моделювання мовленнєвих операцій вилучення змісту з тексту або перекладу текстів з однієї мови на іншу.

Можливості лінгвістичних автоматів розширюються і, як наслідок, стає більш широкою проблематика КЛ [1].

**3. Штучні мови** – це знакові системи, створювані для використання в тих галузях науки і техніки, де застосування природної мови обмежене, менш ефективне або неможливе [3]. Будь-яка штучна мова порівняно з природною – обмежена мова, яка слугує для розв’язання певних завдань.

Класифікація штучних мов:

1. Неспеціалізовані мови загального призначення (наприклад, есперанто, волапюк).

2. Спеціалізовані мови різного призначення (наприклад, символічні мови науки (математика XVI ст., логіка XVIII ст.)).

До 2-го класу входять мови людино-машинного (комп’ютерного) спілкування (мови програмування, мови операційних систем, мови інформаційних систем тощо). Лінгвіст повинен мати уявлення про те, які бувають мови, яка структурна організація мов і як створюються штучні мови. Цьому і присвячені подальші наші міркування.

**Мови програмування (МП)** – це клас штучних мов, призначених для обробки інформації за допомогою комп’ютера. Мов програмування у світі близько 1000.

Історично мови програмування виникли в 40-х роках ХХ ст. і якісно вдосконалювалися в бік спрощення свого опису, методів програмування і додатків для оброблення інформації людиною. Можна виділити такі якісні рівні розвитку МП (тобто так, як їх класифікують у програмуванні):

1. Низький рівень – Асемблери (40–50-ті рр.). Робота з машинними кодами.

2. Середній рівень – Фортран, Алгол (60-ті роки).

3. Високий рівень – Паскаль, С, С++, Prolog (з 80-х років ХХ ст.).

Наразі більшість веб-документів створюють за допомогою мови розмітки гіпертексту HTML (англійською Hyper Text Markup Language – мова розмітки гіпертексту).

Гіпертекст – це текст для перегляду на комп'ютері, який містить зв'язки з іншими документами («гіперзв'язки» чи «гіперпосилання»). Читач має змогу перейти до пов'язаних документів безпосередньо з вихідного (первинного) тексту, активувавши посилання.

HTML (мова гіпертекстової розмітки) – це код, який використовується для структурування і відображення веб-сторінки та її контенту. Наприклад, контент може бути розбитий на параграфи (абзаци), містити список, зображення чи таблицю.

У 1989 р. Тім Бернерс-Лі (британський спеціаліст з інформатики, творець HTML і автор інших розробок в галузі інформаційних технологій, засновник і голова консорціуму W3C, головний розробник Всесвітньої павутини, автор концепції семантичної павутини) запропонував упровадити гіпертекстову систему документів. Наприкінці 1991 року він опублікував у інтернеті перший загальнодоступний опис мови розмітки HTML – «HTML теги». Він же створив і першу програму для перегляду гіпертекстових документів – браузер.

Бернерс-Лі розглядав HTML як похідну мову від SGML, і в середині 1993 року Спеціальна Комісія інтернет-розробок (IETF) офіційно визначила її такою, опублікувавши першу специфікацію HTML. Призначення HTML – описати зовнішній вигляд документа за умови, що розміри екрана і масштаб подання тексту можуть змінюватися у процесі перегляду.

Розширення (тип) гіпертекстового документа, написаного мовою HTML, – htm або html.

Документ HTML можна розглядати як сукупність вказівок і даних як безпосередньо розміщених у документі, так і пов'язаних із ним посиланнями, які під час інтерпретації програмою-браузером відтворюють вигляд сторінок документа. Згідно з ухваленими в інформатиці визначеннями документ HTML вважають програмою, описаною мовою високого рівня. Вказівки HTML називають тегами.

Дескрипторами називають ті теги, які вказують на спосіб відтворення інформації програмою-браузером.

Теги-вказівки HTML – записують у «кутових» дужках, утворених символами «<» і «>» – «менше» й «більше».

Контейнери – це парні теги. Контейнери позначають початок і кінець області дії відповідної вказівки. Між цими тегами може міститися текст та інші теги. Тег, що закриває область дії, порівняно з тим, що відкриває, має додаткову косу риску одразу після «кутової» дужки.

Опис властивостей контейнера записують у першому дескрипторі контейнера через пропуск після назви контейнера або з нового рядка. Нечислові значення параметрів прийнято записувати у лапках.

HTML-файл – це програма подання тексту, написана мовою HTML, яку опрацьовує браузер і подає результат у своєму програмному вікні.

4. Якщо поглянути на мову програмування з лінгвістичного погляду, можна побачити, що МП так само, як і будь-яка мова, має рівневу структуру. Перший нижчий рівень – символний, його елементи – літери, спецсимволи (за аналогією з природною мовою (ПМ) – графічний рівень). Другий рівень – це рівень імен, зарезервованих слів, виразів (у ПМ – це лексичний рівень). Третій рівень – операційний (командний), синтаксичні конструкції мають владний характер (у МП – аналог синтаксичного рівня), і останній – рівень програми, завжди є синтаксично і семантично закінченою послідовністю приписів-команд. Програма – це текст, записаний за заданими правилами штучної мови, тобто МП.

Загальною ознакою опису спеціалізованих штучних мов є формальний метод їхнього опису та визначення за допомогою завдання алфавіту, словника і системи правил освіти та перетворення виразів – це граматики. Формальний метод служить для породження «правильних текстів» («правильних» – означає «записаних за певними правилами»).



Загалом під час написання «правильних текстів» широко використовуються формальні граматики.

Формальна граMATика – це абстрактний апарат, що дозволяє за допомогою однакової процедури отримувати правильні тексти заданої мови.

У 1959 році американський вчений-лінгвіст Н. Хомський запропонував першу класифікацію формальних мов за типом правил формальної граматики.

Існує безліч формальних граматик, як наприклад:

1. Генеративна (або породжувальна) граMATика Хомського.

2. Контекстно-вільна граMATика.

3. ГраMATика безпосередньо-складових.

4. Трансформаційна граMATика (граMATика дерев).

5. ГраMATика Монтегю.

Метамова, яка застосовується на практиці, – це Бекуса – Наурови форми (БНФ-форми або БНФ-нотації), спосіб запису правил контекстно-вільної граматики, тобто форма опису формальної мови. Саме її зазвичай використовують для запису правил мов програмування та протоколів комунікації.

Нотація БНФ є набором «продукцій», кожна з яких відповідає зразку:

**<символ> ::= <вираз, що містить символи>**,

де вираз, що містить символи – це послідовність символів або послідовності символів, розділених вертикальною рискою |, що повністю перелічують можливий вибір символу з лівої частини формули.

Далі,

- < — лівий обмежувач виразу
- > — правий обмежувач виразу
- ::= — визначене як
- | — або

Ці чотири символи є символами метамови, вони не визначені в мові, котру описують. Решта описаних символів належать до «абетки» описуваної мови.

До прикладу, БНФ-нотація для опису англо-російського словника (рис. 1).

Пример типовой странички	<словарь> ::= [<раздел>]
	<раздел> ::= <заглавная лат.буква>[<словарная статья>]
	<заглавная лат.буква> ::= A   B   ...   Z
<b>Р</b>	<словарная статья> ::= <термин> <транскрипция> <перевод>.
Pay [peɪ] 1. платить; 2. заработная плата; 3. расплата.	<термин> ::= [прописная лат.буква]
Pea [pi:] горох.	<прописная лат.буква> ::= a   b   ...   z
Peak [pi:k] остроконечная вершина.	<транскрипция> ::= [ [<звук> ] ]
	<звук> ::= a   ə   ...   z
	<перевод> ::= <определение 1>   <определение 2>
	<определение 1> ::= <слово>   <словосочетание>
	<слово> ::= [прописная русск.буква]
	<прописная русск.буква> ::= a   в   в....   я
	<словосочетание> ::= [ <слово> ]
	<определение 2> ::= 1.<определение 1>; 2.<определение 1>; ...

Рисунок 1 – БНФ-нотація для опису англо-російського словника

Для прикладної лінгвістики побудова БНФ цікава як спосіб зрозуміти структуру будь-якої мови, побачити можливості з моделювання штучних мов.

### Контрольні питання

1. Що є об'єктом і предметом дослідження КЛ?
2. Які основні завдання КЛ?

3. З якими галузями прикладного мовознавства пов'язана КЛ?
4. Що таке штучна мова? Наведіть класифікацію штучних мов.
6. Як класифікуються мови програмування за рівнями?
7. Які формальні граматики ви запам'ятали?
8. Яку метамову використовують для запису правил мов програмування та протоколів комунікації?

### **Список використаної літератури**

1. Волошин В. Г. Комп'ютерна лінгвістика : навчальний посібник / В. Г. Волошин. – Суми : «Університетська книга», 2004. – С. 11–22, 154–165, 286–325.
2. Карпіловська Є. А. Вступ до комп'ютерної лінгвістики : підручник / Є. А. Карпіловська ; відп. ред. А. Загнітко. – Донецьк : Юго-Восток, 2006. – 188 с.

## **ЛЕКЦІЯ 4**

### **ЛЕКСИКОГРАФІЯ ЯК ОДИН ІЗ ВАЖЛИВИХ НАПРЯМКІВ ПРИКЛАДНОЇ ЛІНГВІСТИКИ**

1. Основні напрямки розвитку сучасної лексикографії.
2. Традиційна і машинна лексикографія. Основні напрямки машинної лексикографії.
3. Комп'ютерна лексикографія. Комп'ютерний, або автоматичний словник і комп'ютерний варіант традиційного словника. Електронні словники.
4. Ідеографічні словники та тезауруси. Тезаурус Роже.

1. **Лексикографія** (від грец. Λεξικός «що належать до слова» і γραφω «пишу») може бути визначена як прикладна дисципліна, у центрі уваги якої – методи створення (складання)

словників. **Словник** – книга, що містить зібрання слів (чи морфем, словосполучень, ідіом тощо), розташованих за визначеним принципом, і яка надає відомості про їхні значення, вживання, походження, переклад на іншу мову або інформацію про поняття чи предмети, що ними позначаються.

Сучасна лексикографія розвивається за двома основними напрямками. Один – створення **спеціалізованих словників**, у яких би містилась інформація лише одного типу: наприклад, тільки про написання слова, тільки про його походження, тільки про способи його поєднання з іншими словами тощо.

Інший напрям – створення **комплексних словників**, які б уміщували якомога більше відомостей про слово: не лише давали б тлумачення його значень, граматичні характеристики, правила вимови та написання, але й описували б його смислові зв'язки з іншими словами, особливості його використання в різних стилях, його словотворчі можливості.

2. У лексикографії існує 2 науково-практичних напрямки: **традиційна і машинна лексикографія**. Традиційна лексикографія має глибоке історичне коріння і більшою мірою вивчає теорію й практику складання «традиційних» словників. На відміну від першої, машинна (комп'ютерна) лексикографія – молода наука, яка реалізує традиційні напрацювання в технічних середовищах і створює різноманітні автоматичні словники.

Машинна лексикографія вивчає автоматизацію підготовки словників і переводить словники в комп'ютерну форму. Сьогодні виділяють три основних напрямки машинної лексикографії:

1) автоматичне отримання з тексту за допомогою комп'ютера різних словників (наприклад, частотні словники, словники конкорданс, термінологічні);

2) розроблення теоретичних і практичних аспектів складання комп'ютерних словників для машинного перекладу, штучного інтелекту тощо. (наприклад, словники коренів для морфологічного аналізу);

3) створення словників, є машинними версіями традиційних словників.

3. Комп'ютерна лексикографія посідає чільне місце в комп'ютерній лінгвістиці як комп'ютерний лексикографічний інструментарій у мовознавстві. **Завдання комп'ютерної лексикографії** – розробити комп'ютерні алгоритми, програми, системи та технології для укладання та використання словників. Лексикографічні системи дають змогу формувати словникові статті; зберігати текстову, візуальну та звукову інформацію; здійснювати оброблення словникової інформації (аналіз, пошук, фільтрування, відтворення тощо).

**Комп'ютерним словником** у комп'ютерній лексикографії часто називають «словник, процедури укладання якого здійснює комп'ютер».

З нашого погляду, терміни «автоматичні та електронні словники» можна вважати загальнішими та рівнозначними, а термін «комп'ютерні словники» – одним із їхніх підвидів.

Отже, автоматичні або електронні словники за їх носієм і засобами відтворення поділяємо на:

1) комп'ютерні (словники, які відтворюються за допомогою персонального або портативного комп'ютера);

2) стаціонарні (встановлюють на жорсткому диску комп'ютера);

3) переносні (записують на компакт-дисках і відтворюються на комп'ютері тільки за наявності в дисководі);

4) інтернет-словники (розміщують на комп'ютері-сервері в мережі й доступні з Інтернету).

Можливі комбінації таких комп'ютерних словників:

1) кишенькові (записані на кишенькових електронних пристроях тощо та відтворюються їхніми засобами, наприклад кишенькові перекладачі);

2) мобільні (записані в мобільних телефонах і відтворюються на екрані телефону) тощо.

Серед основних українських комп'ютерних словників можна назвати:

1. Інтегрована лексикографічна система «Словники України».

2. Електронний тримовний російсько-українсько-англійський багатотематичний тлумачний словник MultiLock v.3.80 (27 000 термінів).

3. Галузевий нормативно-термінологічний центр нафтогазового комплексу.

4. Компакт-диск зі збіркою українських академічних словників 30-х років ХХ ст. у форматі pdf, зокрема «Російсько- український словник» (1924–1933 pp., гол. ред. А. Кримський, С. Єфремов), видавництво «К. І. С.», 2007 р.

**Автоматичні словники** – це складний комплекс комп'ютерних програм. Автоматичний словник – це словник у спеціальному машинному форматі, призначений для використання на ПЕОМ користувачем або комп'ютерною програмою оброблення тексту.

Відзначимо деякі особливості автоматичних словників: крім словникової бази даних (перелік слів за алфавітом) для роботи автоматичного словника необхідні спеціальні алгоритми, наприклад, алгоритм лематизації. **Лематизація** – це приведення різних форм слова до його канонічної (вихідної) форми (схоже на морфологічний аналіз слів); у машинних словниках наявні не тільки переліки окремих слів, але і до 50 % словосполучень (особливо в термінологічних словниках, які дуже важливі для перекладу технічних текстів); автоматичні словники набули сьогодні великого поширення для машинного перекладу.

Важливим типом АС є автоматичні словники конкорданс, що містять у машинному вигляді приклади на вживання всіх слів (словоформ), поданих у конкретному творі або в творчому доробку письменника.

Сьогодні увазі перекладачів пропонуються найрізноманітніші електронні словники. Найбільш поширені серед них Lingvo, «МультиЛекс», «Мультитран» «Контекст», Polyglossum.

До очевидних переваг словника Lingvo потрібно віднести, звичайно ж, пояснення і дефініції, якими, крім власне перекладних еквівалентів, забезпечені багато статей. Це надає перекладачам неоціненну допомогу в поповненні їхнього словникового запасу, яке йде паралельно із збагаченням запасу інформаційного.

Користувачі Lingvo мають можливість скласти власні словники та обмінюватися ними. Словники проходять відбір, і кращі з них розміщуються на сайті для загального користування. Отже, процес поповнення і розширення словникової бази – під постійним контролем професійних лексикографів.

Особливе місце тут посідає «**Multitran**»: не лише тому, що є, по суті, своїм онлайн-словником, а й тому, що поповнюється в режимі реального часу, причому не професійними лексикографами. Можна сказати, що це і не словник, а словниковий ресурс загального користування, що ставить собі за мету накопичення максимальної кількості варіантів перекладу слів і словосполучень. «Словником» його називають, ймовірно, для простоти і стислості. Водночас «Мультитран» частково копіює паперові словники, а також поповнюється в режимі реального часу рядовими користувачами.

Підіб'ємо підсумок основних переваг словника «Мультитран»:

- великий обсяг;
- різноманітність варіантів перекладу;
- постійна поповнюваність;
- можливість повідомлення про помилки;
- велика аудиторія, що ділиться своїм досвідом;
- загальнодоступність.

Переваги електронних словників перед паперовими очевидні. Електронні та онлайн-ресурси більш зручні в користуванні, здебільшого, завдяки можливостям пошуку.

4. **Ідеографічні словники** є все ще незвичним явищем для української лексикографії: для української мови досі не створено жодного загальномовного ідеографічного словника. Словники, з якими ми найчастіше стикаємось, – орфографічні, орфоепічні, тлумачні, перекладні, фразеологічні – побудовано за алфавітним принципом. Проте алфавітний принцип, за всієї його зручності й універсальності, не дає змоги продемонструвати системність лексики.

Ідеографічний словник має ономасіологічне спрямування від понять до слів. Його реєстровими одиницями є поняття, кожному з яких підпорядкована словникова стаття – множина слів, які виражають це поняття в мові. Ідеографічна лексикографія має багату традицію в зарубіжному мовознавстві – словники П. М. Роже, Е. Блана, П. Буасьєра, Х. Касареса (іспанська), Р. Галліга і В. фон Вартбурга (французька), Ф. Дорнзайфа (німецька).

Клас **ідеографічних словників** (предметні, тематичні), до яких належать і тезауруси, – це особливі словники, організовані, по-перше, за тематичним принципом, і, по-друге, за принципом «від сенсу до слова», тобто ідеографічні словники орієнтовані на семантику мови, і кожен такий словник – це деяка семантична модель лексики, побудована на ієрархічних відносинах типу «рід – вид», «частина – ціле», «синоніми» тощо.

Найвідоміший класичний ідеографічний словник – **словник Роже** (Roget's Thesaurus, оригінальна назва Thesaurus of English Words and Phrases – «Тезаурус англійських слів і фраз»), у якому 1000 основних тем. Структура Тезауруса Роже містить шість основних семантичних категорій, всередині яких побудовані деревоподібні ієрархії підкатегорій.

**Тезаурус** – це також ідеографічний словник, але він має чітку ієрархію відносин. Словник тезауруса – це перелік дескрипторів (дескриптор – однозначний стандартний термін). У словникових статтях тезауруса відображені істотні для цього терміна зв'язки з іншими поняттями, інакше кажучи, – це



маленька енциклопедія. Тезауруси в машинній лексикографії дуже точні інструменти, тому що вони призначені для інформаційного пошуку, а характеристики інформаційного пошуку (інформаційний шум, точність видачі та ін.) безпосередньо залежать від структури тезауруса, тому є стандарти для порівняння інформаційно-пошукових тезаурусів.

### **Контрольні питання**

1. Які основні напрямки розвитку сучасної лексикографії ви знаєте?
2. У чому полягає відмінність традиційної лексикографії від машинної?
3. Які завдання розв'язує комп'ютерна лексикографія?
4. Як класифікують автоматичні або електронні словники?
5. Які українські комп'ютерні словники ви знаєте?
6. Як можна визначити автоматичний словники?
7. У чому полягають переваги електронних словників?
8. Чим відрізняється ідеографічний словник і тезаурус?

### **Список використаної літератури**

1. Перебийніс В. С. Традиційна та комп'ютерна лексикографія : навч. посібник / В. С. Перебийніс, В. М. Сорокін. – Київ : Видавничий центр КНЛУ, 2009. – 218 с.
2. Орел М. А. Словарь переводчику – друг, товарищ и Брут. Перевод: Информационные технологии. – Москва : Всероссийский центр переводов науч.-техн. лит. и документации, 2009. – С. 79–106.

## ЛЕКЦІЯ 5

### ЕЛЕКТРОННІ КОРПУСИ ТЕКСТІВ І ПЕРЕКЛАД

1. Поняття лінгвістичного корпусу. Історія корпусної лінгвістики.

2. Основні завдання та напрямки корпусної лінгвістики. Предмет дослідження корпусної лінгвістики.

3. Усні та письмові, одномовні та багатомовні корпуси.

4. Лінгвістичні дослідження на базі корпусів.

5. Отримання інформації з корпусу. Програма-конкорданс AntConc.

1. Перш ніж говорити про корпусну лінгвістику, необхідно визначити саме поняття лінгвістичного корпусу. **Корпус** є зібранням уривків текстів в електронній формі, відібраних відповідно до зовнішніх критеріїв, щоб найбільш повно репрезентувати мову або варіацію мови. Функціонує як джерело даних для лінгвістичних досліджень.

Приклади корпусів:

- тексти конкретного письменника або письменників;
- тексти за конкретне десятиліття або сторіччя;
- сучасні тексти певної тематики;
- сучасні тексти адекватно репрезентують мову або суспільство.

Приклади того, що може нам видати пошук за корпусом:

- вживання обраного слова в безпосередньому контексті;
- варіації і послідовність у використанні лексики;
- слова, які найчастіше стоять поруч з обраним словом;
- найбільш важливі відмінності між двома наборами текстів;
- як той чи інший письменник використовує слова і фрази;

- інтертекстуальність: значення слова як сума його вживань;
- приховані (потенційні) моделі використання лексики;
- порівняння мов.

**Репрезентативність**, тобто, здатність відображати всі властивості проблемної зони, є найважливішою властивістю корпусу. Репрезентативність визначається фонетичними, морфологічними, синтаксичними і стильовими параметрами корпусу. Саме репрезентативність відрізняє корпус від простого набору текстів. Репрезентативність істотно залежить від **розміру** корпусу.

Основні віхи створення комп'ютерних корпусів:

1. 1960-ті: Браунський корпус, (США), був створений під керівництвом У. Френсіса. Цей корпус містить 500 текстів загальним обсягом 1 млн слововживань. У кожному з текстів 2000 слововживань. Крім однорідності за довжиною, подані в корпусі тексти однорідні також за 1) часом публікації (усі вперше опубліковані в 1961 р.) та 2) характером мовного оформлення (всі створені літературною англійською мовою). Вони охоплюють 15 жанрів американського варіанта англійської мови: від газетних статей і релігійної й фахової літератури до художньої та ділової прози.

2. 1970-ті: LOB корпус (Великобританія, Норвегія), 1 млн слів <http://khnt.hit.uib.no/icame/manuals/lobman/>.

3. 1980-ті: з'явилася індексована (анотована) версія Браунівського корпусу, яка становила покажчик усіх поданих у його текстах слів з їхніми адресами та граматичними характеристиками.

4. 1990-ті: British National Corpus (BNC), 100 млн слів, національного корпусу <http://www.natcorp.ox.ac.uk/>.

Над реалізацією цього проекту працював колектив видавців, лінгвістів і програмістів під керівництвом видавництва Оксфордського університету (Oxford University Press) протягом 1991–1994 рр. Корпус постійно оновлюють і на сьогодні він налічує понад 100 млн слововживань (100 106 008

слів) у широкому спектрі письмових та усних текстів: від газет, фахової періодики та популярної белетристики до листів, спогадів і творів учнів шкіл чи студентів коледжів та університетів, а також від записів офіційних ділових чи урядових зустрічей до записів розважальних радіопередач і телефонних розмов.

5. The Bank of English, Birmingham (Collins Cobuild), 600 млн слів. (25 % корпусу становить усне мовлення, 75 % – письмове) <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>

6. 2000-i: American National Corpus, 100 млн. слів <http://www.americannationalcorpus.org/>.

7. Gigaword corpora, 2 млрд слів. Моніторингові корпуси англійської, арабської, китайської та інших мов. Спонсоруються Європейським Союзом, створює їх компанія Linguistic Data Consortium. вже 1 мільярд слововживань. Переважно тексти взяті з публіцистики і новин <http://www ldc.upenn.edu>.

Для української мови завдання створення різноманітних корпусів текстів і текстозорієнтованих баз даних різного призначення з відповідними інтерфейсами та текстовими процесорами стоїть не менш гостро, ніж формування словникових баз даних і конструювання комп'ютерних словників і загалом комп'ютерних лексикографічних систем.

2. Головною **метою корпусної лінгвістики** є об'єктивний лінгвістичний опис мовної системи, причому до цього опису корпусна лінгвістика підходить від вивчення конкретної людської комунікації, від реальних текстів, які раніше розглядалися лише як прикра перешкода. Як вторинне завдання розглядається вироблення особливого способу відображення мовного матеріалу в корпусі текстів. Цей спосіб, зокрема, може використовуватися іншими лінгвістичними дисциплінами.

**Напрямки корпусної лінгвістики.** По-перше, це лексикографічні дослідження, створення словників. Практично всі сучасні словники англійської мови (Collins, Webster, MacMillan тощо) видаються на основі величезних корпусів, які

дозволяють зробити словник репрезентативним. Тобто словник може бути правильним або неправильним щодо цього корпусу.

По-друге, вивчення корпусів дозволяє отримувати точні дані про лексичний склад мов, відносні частоти вживання тих чи інших слів.

По-третє, корпусна лінгвістика вивчає зміни в лексичному складі мов, різні його варіації (наприклад, поява і зникнення неологізмів).

Четвертий напрямок корпусної лінгвістики – вивчення граматики природних мов, зокрема – сполучуваності тих чи інших граматичних явищ один з одним. Природно, що дані, отримані з живої мови, набагато більш актуальні, ніж умоглядні граматики традиційної лінгвістики. Крім того, виходить більш об'єктивне дослідження: граматика правильна лише щодо того або іншого корпусу текстів.

По-п'яте, не залишилось без уваги і вивчення текстів. Наприклад, використовуючи корпуси, ми можемо навчитися визначати функціональний стиль через статистичні характеристики тексту – середню довжину слова і пропозиції, характерні поєднання слів тощо. Такі методи вже існують і використовуються в автоматичному реферуванні та тематичному пошуку. Причому вивчати таким способом можна не тільки письмовий, але й усний дискурс.

По-шосте, корпусна лінгвістика активно використовується в лінгводидактиці, тобто в навчанні іноземним мовам. Корпусна лінгвістика дає можливість оновити набір прикладів, які використовуються у викладанні мови.

І нарешті, особливий інтерес для перекладознавців становлять багатомовні корпуси, «вирівняні» або «порівняні» (*aligned*). У «вирівняному корпусі» кожній фразі однією мовою відповідає її еквівалент іншою мовою або мовами. Такі корпуси використовуються під час підготовки перекладачів або створення двомовних словників. Дуже важливі вони для створення систем автоматичного машинного перекладу (якщо

така система спирається на корпус перекладів, зроблених перекладачами-людьми, її якість буде набагато вище). Крім того, такий корпус можна використовувати для досліджень, пов'язаних із порівнянням оригінальних і перекладних текстів.

Паралельні корпуси також можуть бути двох типів – вирівняні (*aligned*) і невирівняні (*not aligned*). «Вирівняність» означає, що в корпусі існує чіткий зв'язок між одиницями перекладу, які відповідають один одному. Тобто ми можемо швидко знайти, як те чи інше слово або речення перекладалося на іншу мову. Зазвичай такими одиницями перекладу слугують все-таки речення, оскільки часто складно вирівняти слова (адже зазвичай переводять недослівно). Такий корпус найбільш корисний для перекладача, оскільки являє собою ту саму «пам'ять перекладів» (*translation memory*) – безцінний ресурс, що дозволяє використовувати попередні переклади. Невирівняні корпуси ще називають «порівняльними».

**Предмет дослідження.** Як і вся наука про мову, корпусна лінгвістика вивчає здебільшого опис і пояснення суті, структури та використання мови, а так само більш приватні питання: вивчення мов, їхні зміни тощо. Однак корпусна лінгвістика стоїть у мовознавстві дещо окремо. Можна зазначити, що часто вона обмежується вивченням швидше лексики та лексичної граматики, ніж синтаксису.

**3. Усні та письмові корпуси.** Велика частина корпусів першого покоління були винятково письмовими. Письмові тексти набагато легше збирати. Існують три методи введення письмових текстів у комп'ютер:

- заново набирати тексти (це краще, ніж пробивати перфокарти, як було з Браунівським корпусом);
- використовувати тексти, які вже існують в електронній формі;
- сканувати надруковані тексти (але водночас потрібно виправляти багато помилок).

Великі сучасні корпуси зазвичай комбіновані з переважанням письмових текстів. Навіть у BNC лише 10 % текстів є усними.

Мова переважно існує саме в усній формі, письмова його форма вторинна. Тому так важливі усні корпуси або змішані (наприклад, *Corpus of Spoken American English* (1991), 1 млн слововживань, 80 годин звучання. *Map Task Corpus* (1991, Університет Глазго, Шотландія), 147 тисяч слововживань, 16 годин звучання).

Усні корпуси можуть містити як монологічне, так і діалогічне мовлення. Щоб зібрати матеріал, використовуються записи з радіо і телебачення, опитування.

**Одномовні та багатомовні.** Корпусних лінгвістів (особливо пов'язаних із перекладом) завжди цікавило завдання складання корпусів декількома мовами. Уже в першому поколінні почали з'являтися двомовні корпуси для таких мов, як англійська, фінська, французька, німецька, грецька, норвезька, іспанська, шведська, валлійська. Такі корпуси ще називаються *bitexts*.

**4. Лінгвістичні дослідження на базі корпусів.** Лінгвістичні корпуси складають, щоб надати основу для більш точного й адекватного опису структурних і функціональних параметрів мови.

Практично всі сучасні словники англійської мови побудовані на базі корпусів. Корпуси допомагають достовірно визначити набір словоформ (*types*) у мові, демонструють появу нових словоформ, використовуються для уточнення різних значень одного слова і їх відносних частот. Найвідоміший приклад – словники видавництва Collins, побудовані на базі корпусу *Cobuild Project*.

Існують дослідження, які описують, яка лексика специфічна для певних типів текстів і навряд чи з'явиться в інших. Так, для наукових текстів одним із таких слів є дієслово *to measure*, а для художніх – *to kiss*.

Інше застосування корпусів у лексичних дослідженнях – опис сполучуваності слів (collocation). Електронні корпуси дозволяють досліджувати проблему сполучуваності більш об'єктивно. Наприклад, ми можемо поррахувати статистичну значущість різниці між тим, як фактично поєднуються слова в корпусі, і тим, як вони теоретично мали б поєднуватися, на підставі їхніх індивідуальних частот. Наприклад, слово *Christmas* часто поєднується з такими рідкісними словами, як *day, eve, tree, cards i present*.

Вивчати синтактику в корпусі складніше, ніж лексику. Основна причина – складність автоматичного синтаксичного аналізу мови (парсингу).

Проте, під час переходу від рівня слів на рівень фраз і речень, корпусні методи не втрачають своєї значущості. Корпусна лінгвістика переконливо показала хибність думки про те, що синтаксичні конструкції сполучуються та варіюються абсолютно вільно. Навпаки – вони також залежать від жанру, як і лексика.

Корпусні дослідження можуть допомогти розв'язати складні граматичні колізії з приводу норми. Так, після слова *different* можуть йти слова *than, to* або *from*. Неясно, яке з них є більш правильним. Дослідження на матеріалі Браунівського корпусу і LLC показали, що *different from* використовується набагато частіше за інших, а *different than* швидше характерно для американської англійської.

Отже, ми побачили, що описи таких аспектів мови, як лексика, морфологія, синтактика і усне мовлення, на базі аналізу корпусів дали нам дуже багато нових знань про мову. Корпусна лінгвістика поки не здійснила системний опис всіх аспектів англійської мови (щобільше, української), але зроблено вже досить, щоб припустити, що корпусний опис мови буде давати все більше інформації про відносну частотність явищ мови в різних її варіаціях.

**5. Отримання інформації з корпусу.** Лінгвістичну інформацію з корпусу можна отримати за допомогою



спеціальних комп'ютерних програм. Найпростіший формат відображення інформації про корпус – це прості списки. Ці списки можуть бути різних типів – від простих глосаріїв до конкордансів.

Формат простого списку не дає можливості зняти полісемію і неоднозначність граматичного класу слова, оскільки це неможливо зробити без контексту. Щоб розібратися з цим питанням, нам потрібно буде перейти до поняття «Конкорданс» (concordance).

Конкорданс – це не просто список слів або словосполучень. Його цінність полягає у тому, що він дає контекст слова. Тобто, ми можемо запустити пошук і отримати всі появи цього конкретного слова в тексті. Результати пошуку демонструються в форматі, який називається KWIC (*key word in context*). Зазвичай під час натискання на рядок програма-конкордансер видає повний контекст.

Однією з найбільш поширених програм-конкордансерів є WordSmithTools <https://lexically.net/wordsmith/> Майка Скотта з Оксфордського університету, але вона платна. Проте, можна скачати демонстраційну версію з обмеженими можливостями. Практично нічим WordSmith не поступається безкоштовна програма AntConc, розроблена японськими вченими <http://www.antlab.sci.waseda.ac.jp/software.html>. У ній реалізовані всі необхідні функції – список слів, конкорданс, пошук колокацій.

### **Контрольні питання**

1. Що таке лінгвістичний корпус? Наведіть приклади корпусів.
2. Що є предметом дослідження корпусної лінгвістики? Яка головна мета корпусної лінгвістики?
3. Що є предметом пошуку в корпусі? Які властивості корпусу ви знаєте?
4. Які основні напрямки корпусної лінгвістики ви знаєте?

5. Як можна отримати інформацію з лінгвістичного корпусу?
6. Що таке конкорданс?

### **Список використаної літератури**

1. Демська-Кульчицька О. Основи національного корпусу української мови / О. Демська-Кульчицька. – Київ : Інститут української мови, 2005. – 220 с.

## **ЛЕКЦІЯ 6 ПРИРОДНИЙ ІНТЕЛЕКТ І ШТУЧНИЙ ІНТЕЛЕКТ**

**1. Моделювання в комп'ютерній лінгвістиці. Мови представлення знань як варіант штучних мов.**

**2. Природний інтелект (інтелект людини) і штучний інтелект (інтелект комп'ютера) як його модель. Моделі знань в штучному інтелекті.**

**3. Підходи до створення систем штучного інтелекту. Машина та тест Тюрінга.**

1. Метод моделювання мови та мовних процесів широко використовується лінгвістами, тому що дає можливість реалізувати теоретичні знання на практиці. Прикладна лінгвістика прагне будувати моделі, що відображають конкретні процеси мовленнєвої діяльності людини в тій чи іншій ситуації.

У лінгвістиці особливо важливі моделі мовленнєвої діяльності, тісно пов'язані з моделями мислення (моделі мислення досліджуються в логіці), тому що логічні і граматико-семантичні форми – це форми одного і того ж самого об'єкта.

Моделювання мовного і людського мислення – це завдання штучного інтелекту як одного з провідних наукових напрямків інформатики, що створює машини, здатні розуміти,

моделювати й аналізувати тексти, зберігати і переробляти інформацію, ухвалювати рішення. Моделювання мови необхідно для повноцінного спілкування в людино-комп'ютерному середовищі, а моделювання логічних міркувань – для розв'язання різних завдань **інтелектуальною системою (ІС)**. Система штучного інтелекту повинна розв'язувати творчі завдання і вести діалог із людиною на основі закладених у неї знань (процедурних і декларативних).

У штучному інтелекті важливо поняття «знання», а під час побудови інтелектуальних систем – мови подання знань. Знання – це досвід, навички та вміння. Розрізняють два основних види знань – *декларативні* («знання що») і *процедурні* («знання як»). Декларативні знання подаються зазвичай у вигляді сукупності пропозицій, тверджень про що-небудь. Декларативні знання піддаються процедурі верифікації в термінах «True – False». Процедурні знання подаються як послідовність (список) операцій, дій, які потрібно виконати. Це деяка загальна інструкція про дії в деякій ситуації. Характерний приклад процедурних знань – інструкції з користування побутовими приладами.

На відміну від декларативних знань, процедурні знання неможливо верифікувати як істинні або хибні. Їх можна оцінювати тільки за успішністю-неуспішністю алгоритму.

Для прикладної лінгвістики особливо важливі **мови подання знань (МПЗ)**, що використовуються під час опису знань. МПЗ – це штучні мови, побудовані за законами штучних мов. Найбільш відомі МПЗ – це логічні, мережеві, фреймові й продукційні.

Логічні МПЗ подають знання у вигляді синтаксично правильних формул будь-якої формальної логічної системи. Логічних систем дуже багато, але найчастіше використовують логіку предикатів 1-го порядку та її мову.

Мережеві МПЗ як моделі використовують семантичні мережі, де вузли мережі – це будь-які інформаційні одиниці – поняття, факти, процеси, імена, а дуги – відносини між ними.

Відносини можуть бути будь-якими (тимчасові, причинно-наслідкові, більше-менше тощо).

Фреймові МПЗ. Основні характеристики цих МПЗ – компактність і вкладеність рівнів. Фрейм – це шаблон типової ситуації або деяка структура, яка містить відомості про певний об'єкт, його характеристики або їхніх значення і є цілісною одиницею знань.

Продукційні МПЗ надають здебільшого процедурні знання, причому формула проста – «Якщо ..., то ...». «Якщо температура 40<sup>0</sup>, то викликай швидку допомогу за телефоном «103».

**2. Штучний інтелект (ШІ)** – це наукова дисципліна, що виникла в 1950-х роках на стику кібернетики, лінгвістики, психології, програмування. Від початку дослідження у сфері ШІ здійснювалися у двох напрямках. Перший (біологічний) – спроби змоделювати за допомогою штучних систем психологічну діяльність людського мозку з метою створення штучного розуму. Другий (прагматичний) – створення програм, що дозволяють із використанням комп'ютера відтворювати не саму розумову діяльність, а процеси роботи. Результати, отримані в цьому напрямку, мають важливу практичну цінність.

ШІ – це програмна система, що імітує на комп'ютері мислення людини. Звичайні комп'ютерні програми працюють у режимі суворих послідовних кроків, продиктованих алгоритмом розв'язання завдання [1].

Основною відмітною ознакою ШІ є робота зі знаннями. Якщо для звичайних програм подання даних алгоритму визначається на рівні опису мови програмування, то для ШІ подання знань виливається в проблему, пов'язану з багатьма питаннями: що таке знання, які знання зберігати в системі у формі бази знань, у якому вигляді і скільки, як їх використовувати, поповнювати тощо.

На відміну від даних, знання мають такі властивості:

1. Знання у вигляді бази подають як саму інформацію, так і інформаційні структури, які можна використовувати під час формування інших знань.

2. Структурність – формування складних об'єктів виконується з простих за допомогою установаження зв'язків між ними.

3. Зв'язність – відбиваються закономірності зв'язків фактів і процесів у вигляді причинно-наслідкових відношень між ними.

4. Активність – знання передбачають цілеспрямоване використання інформації, спроможність управляти інформаційними процесами за розв'язанням завдань.

Основними частинами системи ШІ, заснованими на знаннях, є: 1) база знань; 2) механізми висновку; 3) інтерфейс користувача.

Кожна з цих частин може бути побудована по-різному в різних системах, усі вони мають спільне характерне завдання – моделювання людських міркувань.

Знання, на які спирається людина, розв'язуючи те чи інше завдання, дуже різномірні і містять такі складові пізнання: понятійні знання (набір понять і їхні взаємозв'язки); конструктивні знання (знання про структуру і взаємодію частин різних об'єктів); процедурні знання (методи, алгоритми і програми розв'язання завдань); фактографічні знання (кількісні і якісні характеристики об'єктів, явищ і їх елементів).

**Штучний інтелект (artificial intelligence), або штучний розум (artificial mind)** визначають як певну комп'ютерну модель природного інтелекту, або інтелекту людини.

Уміння правильно ставити завдання розумової діяльності, добирати потрібну й всебічну інформацію про досліджувані об'єкти дійсності, структурувати певну предметну галузь – об'єкт розумової діяльності, а саме: віднаходити її релевантні ознаки і встановлювати зв'язки між ними. Проблематика комп'ютерної лінгвістики часто

пов'язується з моделюванням спілкування, зокрема, із забезпеченням спілкування людини з комп'ютером природною або обмеженою природною мовою.

3. Відомий британський вчений Алан М. Тюрінг запропонував один з можливих критеріїв оцінювання рівня «машинного мислення». Розроблена ним процедура відома як «тест Тюрінга». Ідея досить проста: якщо людина, спілкуючись з машиною, не зможе помітити, що її співрозмовник – автомат, то можна визнати, що така машина мислить як людина, тобто має інтелект, подібний до інтелекту людини. Водночас як умова «чистоти критерію» висувається низка обмежень, вимог до ретельності виконання тесту: передусім – контакт між людиною і машиною мусить бути не прямим, а опосередкованим, тобто їхнє спілкування здійснює певний пристрій. Крім того, це основне обмеження може мати низку додаткових, підсилювальних умов проведення експерименту, наприклад, окреслення для спілкування певної предметної галузі, мови спілкування, її складу тощо. Як відзначають фахівці-практики, «незважаючи на свою суб'єктивність, тест Тюрінга як критерій інтелектуальності має великі переваги. Він спирається на порівняння з людиною. Доки в нас немає загальної теорії мислення, визнати дещо мислячим ми можемо тільки, порівнявши це «дещо» з людиною – єдиною істотою, що має цю характеристику апіорі.

Ще однією з найбільш цікавих комп'ютерних моделей діалогу, що викликала жваві теоретичні дискусії, була програма Джозефа Вейценбаум «Еліза», перший варіант якої з'явився в 1966 р. Спочатку «Еліза» створювалася як іграшка та навчальний зразок програми-імітатора, метою якої є не моделювання мислення в точному сенсі, а моделювання мовної поведінки. Програма підтримувала розмову зі співрозмовником у реальному часі, однак під час її розроблення були використані обмежені програмні ресурси, лінгвістичний аналіз і синтез також були зведені до мінімуму.

З лінгвістичного погляду алгоритми програми «Еліза» містять мінімум лінгвістичної інформації. По-перше, це комплекс ключових слів, які актуалізуються в деякі стійкі комунікативні формули (шаблони), по-друге, здатність відносно нескладно трансформувати попереднє висловлювання. Наприклад, програма в змозі переробити попереднє твердження в загальне питання <https://www.cyberpsych.org/eliza/>.

Комп'ютерний експеримент з програмою, що моделює поведінку учасника комунікації, дозволяє експериментально підтвердити або спростувати багато положень теорії діалогу, розроблених як у лінгвістиці, так і в суміжних дисциплінах – у дискурс-аналізі, теорії комунікації, психології та соціології спілкування.

На сьогодні в галузі штучного інтелекту накреслилися такі напрями, у розбудові яких є певні здобутки й перспективи: доведення теорем; розпізнавання образів (слухових і зорових); теорія ігор; адаптивне, динамічне й евристичне програмування; ухвалення рішень; природна мова та її машинне розуміння або спілкування з комп'ютером природною мовою; системи із самоорганізацією та саморегулюванням, синергетика; роботика; створення комп'ютером музики; самонавчальні мережі; оброблення даних, поданих природною мовою; вербальне й концептуальне навчання.

### **Контрольні питання**

1. Що таке система штучного інтелекту?
2. Наведіть класифікацію МПЗ.
3. Напрямки дослідження у сфері ШІ.
4. Які властивості знань вам відомі?
5. Які проблеми виникли у зв'язку з функціонуванням комп'ютерних моделей спілкування?
6. У чому полягає суть тесту Тюрінга?
7. Які напрями ШІ мають певні здобутки й перспективи?

## Список використаної літератури

1. Карпіловська Є. А. Вступ до комп'ютерної лінгвістики: підручник / Є. А. Карпіловська / відп. ред. А. Загнітко. – Донецьк : Юго-Восток, 2006. – 188 с.
2. Федоров Е. Е. Введение в системы искусственного интеллекта : учебн. пос. / Е. Е. Федоров. – Донецк : Наука і освіта, 2006. – 84 с.

## ЛЕКЦІЯ 7 СИСТЕМИ МАШИННОГО ПЕРЕКЛАДУ

1. Етапи розвитку машинного перекладу (МП).
2. Сучасні стратегії створення систем МП. Типи систем МП.
3. Переваги та недоліки систем МП.

1. Завжди потрібно чітко розрізняти автоматичний переклад (Machine translation, MT) і переклад за допомогою комп'ютера (computer-aided translation, CAT). Автоматичний переклад – це будь-яка система, яка власне перекладає. Всі інші програми, що так чи інакше допомагають перекладачеві, – CAT.

Вперше думку про можливість машинного перекладу висловив Чарльз Бєббідж (179–871), який розробив у 1836–848 рр. проект цифрової аналітичної машини – механічного прототипу електронних цифрових обчислювальних машин, що з'явилися через 100 років. Відзначимо деякі історичні дати, пов'язані з розвитком машинного перекладу.

1946 р. – американські фахівці У. Уївер і А. Буг дійшли думки про можливості технічного здійснення машинного перекладу (спочатку використовувався менш привабливий, але, по суті, більш правильний термін – механічний переклад).

1947 р. – розроблена перша програма послівного перекладу (дослівник). У. Уївер звертає увагу Н. Вінера



(творець науки кібернетики) на важливість теми механічного перекладу і намагається залучити його до цієї роботи. Вінер відповів, що механізація мови є передчасною через семантичні проблеми. Час підтвердив його слова.

1948 р. – сформульовано завдання виконання лематизації (реконструювання основної форми змінюваних частин мови за її довільною формою).

15 липня 1949 року – опубліковано знамениту статтю У. Уївера «Переклад», більш відому середі лінгвістів як «Меморандум Уївера». У цій роботі викладені суть завдання, способи розв'язання, визначені методологічні аспекти меж застосування механічних систем перекладу.

1951 р. – обговорюється можливість механічного синтаксичного аналізу.

1952 р. – перша міжнародна конференція з машинного перекладу. Учасники пропонували принципи морфологічного і синтаксичного аналізу, розглядали багатозначність слів, обговорювали ідіоматику, говорили про цінність використання словника словоформ і урахування контексту.

1954 р. – для IBM-701 написана перша програма російсько-гліійського перекладу (60 фраз) у Джорджтаунському університеті. Першою перекладеною фразою була «Величина кута визначається відношенням дуги до радіуса».

На Заході створюються системи машинного перекладу силами корпорацій для своїх технічних потреб, наприклад, СИСТРАН (корпорація Ксерокс), ЛОГОС (США), МЕТЕО (Канада), ДЖЕНЕРАЛ МОТОРС (США), ТИТУС (Франція), КУЛЬТ (Китай). Крім приватних фірм, багато університетів світу створили наукові лабораторії з машинного перекладу. Найбільш широкої популярності набули роботи таких університетів: Гренобльського, Монреальського, Гонконгського, Единбурзького, Массачусетського технологічного інституту та ін.

Автоматичний переклад (МТ) був головною метою комп'ютерної лінгвістики в 50-х і 60-х роках ХХ ст. Однак скоро

стало ясно, що поки не розв'язане завдання створення штучного інтелекту, неможливо змусити комп'ютер «розуміти» текст і контекст так, як його «розуміє» жива людина.

Отже, самостійний автоматичний переклад мало реальний. Якщо результати перекладу потрібно десь публікувати, то необхідно редагування людиною.

Важливо розрізнати два принципово різні способи створення систем автоматичного перекладу, які є актуальними сьогодні: **на основі правил (rule-based)** і **статистичний (statistical)**. Перший спосіб передбачає створення складних моделей мов, на основі яких і відбувається переклад. Другий спосіб використовує вже наявні двомовні корпуси перекладених текстів.

Системи машинного перекладу засновані на правилах – загальний термін, який позначає системи машинного перекладу на основі лінгвістичної інформації про вихідну і мову перекладу. Вони складаються з двомовних словників і граматик, що охоплюють основні семантичні, морфологічні, синтаксичні закономірності кожної мови. Такий підхід до машинного перекладу ще називають **класичним**.

На основі цих даних вихідний текст послідовно, речення за реченням, перетворюється на текст перекладу. Часто такі системи протиставляють системам машинного перекладу, які засновані на прикладах. Принцип роботи таких систем – зв'язок структури вхідного і вихідного речення. Водночас переклад виходить не особливо гарної якості.

Ці системи діляться на три групи:

- системи послівного перекладу;
- трансфертні системи;
- інтерлінгвістичні.

### **Послівний переклад**

Такі системи використовуються зараз вкрай рідко через низьку якість перекладу. Слова вихідного тексту перетворюються на слова перекладного тексту. Це найпростіший метод машинного перекладу. Він

використовується для перекладу довгих списків слів (наприклад, каталогів). Так само він може бути використаний для складання підрядника для ТМ-систем.

*Переваги:*

- простота;
- висока швидкість роботи;
- не вимогливі до ресурсів.

*Недоліки:* низька якість перекладу.

### **Трансферні системи**

Трансферні системи працюють за дуже простим принципом: до вхідного тексту застосовуються правила, які ставлять у відповідність структури вихідної і мови перекладу. Робота будь-якої системи трансферного перекладу складається як мінімум у п'яти частин:

- морфологічний аналіз;
- лексична категоризація;
- лексичний трансфер;
- структурний трансфер;
- морфологічна генерація.

*Переваги:*

- висока якість перекладу (за наявності потрібних словників і правил);
- зазвичай є вибір тематики тексту, який підвищує якість перекладу;
- можливе уточнення перекладу завдяки внесенню змін до бази даних перекладача (у такий спосіб користувач отримує потенційно безліч термінів, з якими можна вільно оперувати, і можна досягти «нескінченної» якості перекладу).

*Недоліки:*

- висока вартість і час розроблення;
- для додавання нової мови доводиться переробляти систему заново;
- потрібна команда кваліфікованих лінгвістів для опису кожної вихідної мови і кожної мови перекладу;
- вимогливість до ресурсів на етапі складання бази.

## **Інтерлінгвістичний машинний переклад**

Інтерлінгвістичний машинний переклад – один із класичних підходів до машинного перекладу. Оригінальний текст трансформується в абстрактне уявлення, яке не залежить від мови (на відміну від трансферного перекладу). Перекладний текст створюється на основі цього подання. Можна довести математично, що в межах цього підходу створення кожного нового інтерпретатора мови для такої системи буде здешевлювати її порівняно, наприклад, із системою трансфертного перекладу. Крім того, у межах такого підходу можна;

- реалізувати «переказ тексту», перефразування вихідного тексту в межах однієї мови;

- відносно проста реалізація перекладу сильно мов, що сильно відрізняються, таких як, наприклад українська та арабська.

Вперше інтерлінгвістичний підхід був запропонований у 17 столітті Декартом і Лейбніцем, які запропонували універсальні словники, які використовують числові коди.

Для використання системи інтерлінгвістичного машинного перекладу необхідні:

- словники для аналізу і генерації текстів;
- опис граматик мов;
- база знань понять (для створення міжмовного уявлення);
- правила проекції понять для мов та їх подання.

*Передбачувані переваги:*

- висока якість перекладу, незалежно від вибору мови.
- виділення сенсу з вихідного тексту відбувається один раз і потім записується на будь-яку мову, зокрема вихідний;
- низька вартість трудовитрат на додавання нової мови в систему.

*Недоліки:*

- спірність потенційної можливості;
- висока складність розроблення.

## Статистичний машинний переклад

Він використовує порівняння великих обсягів мовних пар. Статистичний МП має властивість «самонавчання». Чим більше в розпорядженні є мовних пар і чим точніше вони відповідають один одному, тим краще результат статистичного машинного перекладу.

Під час виконання перекладу комп'ютер не оперує лінгвістичними алгоритмами, а обчислює ймовірність застосування того чи іншого слова або виразу. Слово або послідовність слів, що мають оптимальну ймовірність, вважаються найбільш відповідними перекладу вихідного тексту і підставляються комп'ютером в одержуваний у наслідок цього текст.

У статистичному машинному перекладі ставиться завдання не перекладу тексту, а завдання його розшифровки.

Найбільшу популярність цей метод набув завдяки компанії Google. Крім перекладача Google, існує ще низка систем і бібліотек, що використовують статистичний підхід: Giza++; Moses; Pharaoh; Rewrite; BLEU scoring tool.

### *Переваги:*

- висока якість перекладу;
- кваліфіковані лінгвісти не потрібні безпосередньо для побудови системи, потрібні тільки інженери;
- праця людини мінімізована для створення таких систем;
- не потрібно перебудовувати систему в разі додавання нової мови;
- можливе навчання системи під час її експлуатації.

### *Недоліки:*

- для навчання потрібні великі паралельні корпуси тексту;
- складний математичний апарат;
- якісний переклад можливий тільки для фраз, які цілком містяться в n-програмних моделях;

- переклад залежить від корпусів, які використовувалися під час навчання;
- під час додавання нової мови доводиться аналізувати велику кількість паралельних корпусів; тривалий час навчання.

### **Контрольні питання**

1. Хто створив механічний прототип електронних цифрових обчислювальних машин?
2. Що таке *інтерлінгва*?
3. Який вид машинного перекладу має властивість «самонавчання»?
4. Який спосіб створення систем автоматичного перекладу використовує вже наявні двомовні корпуси перекладних текстів?
5. Які системи і бібліотеки, крім перекладача Google, що використовують статистичний підхід, вам відомі?
6. У якій системі застосовується попереднє оброблення тексту?

### **Список використаної літератури**

1. Ткачук В. М., Чумак Г. В. Теорія і практика машинного перекладу: навч. посібник / В. М. Ткачук, Г. В. Чумак. – Тернопіль : Підручники і посібники, 2006. – 72 с.
2. Kenny D. Teaching Machine Translation & Translating Technology : A Contrastive Study [Electronic Resource] / D. Kenny, A. Way // Workshop on Teaching Machine Translation : VIII MT Summit, 2004 : proceedings. – Geneva, 2001. – Way of access : <http://www.dlsi.ua.es/tmt/docum/TMT2.pdf>.

## **ЛЕКЦІЯ 8**

### **СИСТЕМИ АВТОМАТИЗОВАНОГО ПЕРЕКЛАДУ**

1. САТ-програми як один із найважливіших аспектів інструментального компонента для перекладу.

## **2. Принцип дії ТМ-інструментів. Переваги використання CAT-програм на основі ТМ-інструментів.**

### **3. Особливості процесу локалізації.**

1. На початку 80-х років минулого століття керівник японської національної програми з машинного перекладу Макото Нагао розробив нову концепцію перекладу, що ґрунтується на прикладах (*example based translation*). Ця ідея доволі проста. Відомо, що в галузі наукової комунікації мови характеризуються одноманітністю синтаксичних конструкцій і значною мірою термінологізованим лексичним наповненням. Під час зміни лексики та термінології часто повторюються одні й ті самі конструкції. Це наштовхує на думку про те, що за накопичення достатньо великої збірки раніше перекладених фраз великою є вірогідність того, що більша частина наступних текстів і їхніх перекладів буде аналогічною до вже перекладеного вручну. Для «навчання» системи необхідно сформувати великий масив вихідних текстів і їхніх перекладів (такі пари називають білінгвами), який повинен бути завантажений у надпотужний багатопроцесорний комп'ютер. Під час перекладу нових текстів із масиву білінгвів мають обиратися аналоги фрагментів цих текстів, які можуть бути використані для формування вихідного тексту.

Історію розвитку машинного перекладу ідея Макото Нагао через усталеність ставлення до цієї проблеми змінити ніяк не могла. Однак на розвиток сучасної технології перекладу з використанням комп'ютера вона мала великий вплив за останні 20–30 років. На сьогодні існує багато систем, які так чи інакше реалізують запропоновану концепцію. Так, дуже великою популярністю у світі користуються CAT-програми (*Computer Aided/Assisted Translation*). Ці програми не мають нічого спільного із системами машинного перекладу, їх на призначення полягає у спрощенні процесу перекладу однотипних документів. А тому в цьому автоматизованому процесі обов'язково має бути присутнім перекладач, що

виконує повною мірою свої функції. Єдине, що робить машина, – це запам'ятовування перекладених фрагментів тексту та їхнє використання в наступних перекладах. Така технологія отримала назву *Translation Memory (TM)*.

Варто відзначити, що значення та використання таких програм (*Trados, Transit, Across, Wordfast* тощо) стає обов'язковою вимогою для участі у великих перекладацьких проєктах. Окрім цього, багато агентств і бюро перекладів, а інколи навіть безпосередні замовники все частіше стали висувати такі вимоги до перекладачів під час роботи навіть над невеликими документами та текстами. Пов'язано це з тим, що компанії, які тісно пов'язані з перекладацькою галуззю, прагнуть створити та постійно поповнювати власну базу *TM* для її використання в подальшому.

Використання *CAT*-програм дуже часто дозволяє значною мірою скоротити час самого перекладу. Наприклад, якщо перекладач часто працює над технічними текстами, однотипними та схожими за структурою документами, згадані програми значною мірою допоможуть під час перекладу. За підрахунками компанії *Prompt*, використання *CAT*-технологій дозволяє підвищити ефективність перекладів до 80 %. Однак, якщо перекладач перекладає художні тексти, малоймовірно, що такі програми виявляться корисними.

Крім вже згаданих програм, у групи *CAT* може входити низка програм, призначених безпосередньо для локалізації різноманітних програмних продуктів, веб-сайтів тощо.

Отже, розгляньмо детальніше *TM*-інструменти за А. Л. Семеновим. Основне їхнє призначення – збереження пар речень у перекладацькій базі даних. Кожна така пара складається з речення з тексту оригіналу та його перекладу іншою мовою. Можливим є розміщення фрагментів тексту та інших форматів – таких, що перевищують довжину речення або становлять лише його частину. Але в автоматичному режимі зберігаються саме речення – тому досить часто такі програми називають «пам'ять речень» (*Sentence Memory*).



**2. Принцип дії ТМ-інструментів** простий. Для вивчення цих програм не потрібно жодних додаткових знань, а навичка користування ними з'являється дуже швидко. Зазвичай вони інтегруються з такими поширеними офісними програмами, як, наприклад, *Word*. Деякі з них мають власні засоби редагування тексту, інтерфейси яких мало відрізняються від звичних для сучасного перекладача інтерфейсів текстових редакторів.

Сучасний ТМ-інструмент являє собою складну комп'ютерну систему, у яку входить багато програм, що мають різні функції. Ці функції можна розділити на дві групи – основні та сервісні. Саме так побудована найпоширеніша система *TRADOS*, основними програмами якої є *Translator's Workbench* та *MultiTerm*.

Програма *Translator's Workbench* автоматично вводить речення, що підлягає перекладу, у базу даних (пам'ять перекладача). Зазвичай довжина запису визначається крапками – від крапки до крапки. Після цього автоматично до вихідного речення в базі даних приєднується виконаний переклад. Після закінчення роботи перекладач отримує виконаний ним переклад, а в базу даних надходять два фрагментовані паралельні тексти. Наступний текст, що перекладатиметься, буде оброблений так само та надійде до бази даних.

Спочатку використання ТМ-інструменту ніяк не впливатиме на продуктивність праці перекладача. Але база даних розширюється доволі швидко – відповідно до швидкості роботи перекладача. На початку перекладу кожного нового речення ТМ-інструмент автоматично перевіряє в базі даних наявність такого самого або схожого речення серед перекладених раніше.

Загальною тенденцією в сучасному перекладацькому бізнесі стає жанрово-тематична спеціалізація. Це означає, що і бюро перекладів, і індивідуальні перекладачі зазвичай виконують однотипну роботу. Наприклад, на продовольчому ринку пропонується кілька десятків однотипних електричних

приладів, кожен із яких супроводжується інструкцією користувача. Тексти цих інструкцій однотипні. Після перекладу першої з них у другій перекладачеві трапляється не більше ніж половина нового тексту. Якщо першу інструкцію було перекладено із застосуванням *ТМ*-інструменту, у другій доведеться перекласти лише ті фрагменти, що відрізняються. Ті фрагменти, що збігаються, будуть надані перекладачеві *ТМ*-інструментом, залишаться лише або погодитися із отриманою пропозицією, або відредагувати його.

Кількість однотипних текстів настільки ж велика, наскільки великий споживчий ринок. Додатковим значним джерелом текстів, які частково збігаються, є постійне вдосконалення вже існуючих товарів. Як приклад можна навести численні «Керівництва користувачів» програмних продуктів. Невдовзі після виходу на ринок першої версії певної програми «Керівництво користувача» для якої було перекладене на українську мову, з'являється оновлена друга версія. Ймовірніше за все, «Керівництво користувача» для оновленої версії буде не набагато відрізнятися від попереднього, можливо, усього на 10–15 %. Але його також необхідно видати українською мовою. Якщо воно являє собою об'ємний текст зі складною структурою, то його перегляд, пошук розбіжностей і механічна корекція будуть дуже трудомісткими, а, можливо, навіть певні незначні розбіжності взагалі залишаться непоміченими. *ТМ*-інструмент повністю перегляне новий текст за фразами, запропонує варіанти речень, які повністю збігаються, зверне увагу на ті, які частково збігаються, виділивши в них фрагменти, що відрізняються, а також покаже можливості редагування речень, що не збігаються. Велика перевага *ТМ*-інструментів полягає також у тому, що під час перекладу вони повністю зберігають формат тексту оригіналу. Отже, перекладач звільняється від необхідності стежити за кеглем, що використовується, абзацом тощо.

Найголовніша перевага використання *ТМ*-інструментів – звільнення перекладача від монотонних операцій, що постійно повторюються. Можливо, ці програми не підвищують продуктивності праці окремої особи, але вони значно підвищують продуктивність перекладацького бізнесу.

Більшість *ТМ*-інструментів існують у двох версіях: для інсталяції на персональному комп'ютері та для мережевого використання. *ТМ*-інструмент, інстальований у локальній мережі певного перекладацького бюро, перетворює усю перекладацьку діяльність бюро на колективну. У створенні бази даних беруть участь усі співробітники, а наявність і використання загальної бази сприяє встановленню єдиного перекладацького стилю усього бюро і значно підвищує якість перекладів, оскільки рішення багатьох перекладацьких проблем також стає колективним. Під час термінового перекладу великих за обсягом текстів часто до роботи доводиться братися одразу декільком перекладачам. Без використання *ТМ*-інструменту сумістити фрагменти, виконані декількома особами, – завдання нелегке. Основний принцип перекладу – єдність термінології – буде порушено. *ТМ*-інструмент у такому разі є єдиним засобом виходу зі складної ситуації, оскільки він візьме на себе функції координатора дій і результатів окремих перекладачів. Підвищення якості роботи бюро автоматично відбувається ще й тому, що база даних *ТМ*-інструменту перебуває в постійній роботі, тобто поповнення бази здійснюється і в режимі редагування. А вдосконалення цього режиму, зі свого боку, відкриває можливості для виправлення помилок і недопущення їх у подальших перекладах.

Своєрідний розподіл праці може бути досягнутий і в умовах індивідуальної перекладацької діяльності на персональному комп'ютері. У більшості робочих *ТМ*-інструментів передбачене підключення до вже накопичених на інших комп'ютерах баз даних.

Отже, узагальнимо переваги використання *САТ*-програм на основі *ТМ*-інструментів:

- гарантія того, що документ перекладено повністю, – програма просто не дасть вам залишити порожні сегменти;
  - забезпечення одноманітності термінології, що особливо важливо під час роботи над великими проектами або ж у разі, коли над одним проектом працює декілька перекладачів і редакторів;
  - можливість роботи з документами різних форматів (залежно від функціональних можливостей програмної оболонки), наприклад, із файлами Adobe Reader (.pdf), Microsoft Excel (.xls), Microsoft Power Point (.ppt), AutoCAD (.dwg) тощо, водночас немає потреби встановлювати на комп'ютер ці програми (вони можуть бути дорогими та вимогливими щодо системних ресурсів);
  - підвищення продуктивності під час роботи над проектом;
  - зниження часових і трудовитрат, особливо під час роботи над документами, що містять текст, який часто повторюється, або під час перекладу нових редакцій документів;
  - відсутність необхідності повторного форматування перекладеного тексту (зазвичай форматування автоматично повторюється програмою), що дозволяє суттєво скоротити часові затрати;
  - можливість збереження бази даних перекладів для подальшого використання під час роботи над новими проектами;
  - наявність функції суміщення технології ТМ і машинного перекладу (у тих випадках, коли це доцільно);
  - можливість вибору програмної оболонки того чи іншого розробника, що найповніше відповідає вашим вимогам до функціональності та зручності роботи.
- Труднощі, що виникають під час використання технології ТМ та її обмеження:
- інколи перекладачеві може бути складно відстежити контекст сегмента тексту, що перекладається;

– якщо ТМ-технологія раніше не використовувалася, необхідно буде ретельно передивитися увесь процес перекладу, щоб найбільш органічно та ефективно «вбудувати» технологію у свою роботу;

– програмні оболонки для технології накопичувального перекладу підтримують далеко не всі можливі формати; точніше кажучи, кількість форматів, що підтримуються, доволі обмежена (хоча для деяких пакетів й існують спеціальні надбудови – фільтри, що дозволяють додавати підтримку формату, однак їхня кількість також обмежена, і вони випускаються розробником доволі рідко);

– програмні оболонки зазвичай несумісні одна з одною (єдиного стандарту для усіх САТ-програм поки що не існує);

– висока вартість програм, особливо мережових версій (які тим не менш мають свої обмеження, наприклад, на обсяг бази даних, що зберігається на сервері); всі програмні доповнення (наприклад, фільтри) також платні; існують безкоштовні або умовно безкоштовні САТ-програми, але особливого розповсюдження вони не отримали через свою обмежену функціональність);

– якщо сегмент тексту було перекладено неправильно, а перекладач або редактор його пропустили, то під час використання цієї бази даних у подальшому усі аналогічні сегменти будуть також перекладені неправильно;

– необхідно постійно відстежувати якість баз даних ТМ, оновлювати та редагувати їхній зміст, дуже часто вручну – інакше їхня цінність значно знижується;

– низька ефективність під час роботи з текстами, у яких мало сегментів, що повторюються, або ж де такі сегменти взагалі відсутні;

– можуть виникати складнощі під час сегментації через те, що в різних мовах використовується різний порядок слів.

Першу програму, що використовувала базу ПП, – Translation Manager – створила компанія IBM. Спочатку компанія використовувала її винятково з власною метою – для

**локалізації (адаптації до іноземних мов)** власних програмних продуктів. Потім програма була запропонована ринку і довго була монополістом у галузі. Проста у використанні, швидка, вона забезпечувала гарні можливості для індивідуальних налаштувань користувача.

До найрозповсюдженіших CAT-програм можна віднести програму *Déjà Vu*, яка користується в перекладачів Європи популярністю завдяки своїй гнучкості та адаптивності. Крім зручності, вона вирізняється співвідношенням ціна/якість. Спеціалісти відмічають, що від початку ця програма розроблялась в контакті з майбутніми юзерами. Відомо, що остання версія програми тестувалася також і російськими перекладачами. Їхні статті й відгуки були опубліковані в *MultiLingual Computing & Technology*.

Популярна в Європі також програма *STAR Transit*. За своєю ідеологією вона зовсім не схожа на продукти класу: на відміну від їхнього принципу «*Information at your fingertips*», початкові дані *STAR Transit* зберігає в текстовому форматі у вигляді файлів, котрі користувач розміщує в директорії. Потім перекладач вказує які директорії/файли йому потрібні, а програма формує мережу посилань, з якими буде йти робота. У Росії та Україні *STAR Transit* не набув широкого розповсюдження, а *PROMT*, офіційний дистриб'ютор *Transit*, переключився на розповсюдження *Translator's Workbench* фірми *Trados*.

*Translator's Workbench*, або просто *Trados*, швидко отримав визнання в Україні й Росії, спочатку серед перекладацьких фірм, а потім й серед індивідуальних перекладачів. Великою перевагою є те, що він інтегрується в такі відомі програмні продукти, як *Word* та інтернет-браузери і можна використовувати всі його інструменти. *Trados* є надзвичайно зручним для пересічного перекладача, що не знайомий із програмуванням, він зрозуміліший ніж інші програми типу. Усі елементи управління мають підказки. Цей

продукт має всі корисні інструменти, які в інших продуктах є лише в частковій наявності.

Наприкінці 2008 року з'явилася перша вітчизняна система ПП AnyMem, розроблена київською компанією Advanced International Translations.

- *OmegaT* (багатоплатформова, підтримує усі мови);
- *OpenLanguageTools* (багатоплатформова, підтримує усі мови);
- *Transolution* (багатоплатформова, підтримує усі мови): для роботи необхідно мати встановлену мову програмування *Python*;
- *AidTrans Studio Basic* (безкоштовна, працює на базі *Windows*, підтримує усі мови, потребує установки *NET*);
- *MemoQ 4Free* (безкоштовна, працює на базі *Windows*, не підтримує деякі мови (зокрема деякі ієрогліфічні мови), потребує установки *NET2.0*);
- *MetaTaxis* (безкоштовна, працює на базі *Windows*, підтримує усі мови, потребує установки *Microsoft Word*);
- *across* (для перекладачів-фрілансерів доступна безкоштовна версія з обмеженою функціональністю);
- *SDL Trados*;
- *Tr-AID*;
- *T-Remote Memory*;
- *Wordfast*.

Зазвичай у програмних пакетах, окрім власне інструменту для роботи з базами ТМ, є додаткові функції: наприклад, інструмент для вирівнювання (*alignment tool*), який дозволяє створювати базу даних ТМ за старим документом і його перекладом, інструменти для аналізу тексту (визначення кількості повторів, а також кількості збігів із базою ТМ та ступеня збігу сегментів), інструмент для попереднього перекладу (оброблення текстів, що дозволяє підставити сегменти, які збігаються з базою, у деяких випадках – виділити ті сегменти, що не мають збігів). Програми також дозволяють

гнучко налаштовувати та змінювати інтерфейс залежно від потреб користувача.

У пакетах деяких програм містяться також додаткові програми, що дозволяють створювати термінологічні бази та працювати з ними. У пакеті *Trados* це, наприклад, програма *Multiterm*. Ця програма може використовуватися разом із програмою для перекладу, виділяючи терміни та пропонуючи варіанти для перекладу. Вона дозволяє зробити переклад за допомогою програмного пакету максимально простим та ефективним, автоматично допомагаючи перекладачеві виконати важливе завдання – забезпечення одноманітності. Окрім цього, додаткові програми для роботи з термінологічними базами дозволяють реалізувати вимоги до специфічної термінології, що висуваються замовником. Такі вимоги трапляються доволі часто, оскільки великі компанії нерідко використовують в документації свою роками напрацьовану систему визначень, термінів, абревіатур, скорочень тощо.

*CAT*-програми мають ще одну важливу перевагу: деякі з них пристосовані для використання під час локалізації.

**3. Локалізація** – це адаптація програмного забезпечення щоб з ним могли працювати користувачі, які розмовляють іншими мовами, з урахуванням особливостей, що прийняті в їхніх країнах. Зазвичай локалізації підлягають програмний інтерфейс та документація до програмного забезпечення, крім того, може виникнути потреба змінити одиниці виміру, а також виконати супутню переробку програм. Процес перекладу є лише частиною локалізації. Особливістю процесу локалізації є те, що дуже часто перекладач має тісно взаємодіяти з програмістами. До найпоширеніших **завдань локалізації** належать: використання символів національних валют, зміна форматів дати та часу, переклад елементів інтерфейсу, під час якого треба враховувати те, що в різних мовах один і той самий за змістом текст може відрізнятися за довжиною (кількістю символів) тощо.



Умовно усе програмне забезпечення, яке необхідно локалізувати, можна розділити на дві групи:

- програмне забезпечення, що розроблялося з урахуванням багатомовності;
- програмне забезпечення, що створювалося для використання лише однією мовою.

Друга група програм найскладніша для локалізації та потребує певних підготовчих операцій, крім того, вона може потребувати істотної переробки програмного коду.

Програми першої групи, хоча й викликають певні труднощі під час локалізації, усе ж розроблялися з урахуванням того, що їхній інтерфейс буде перекладатися на декілька мов. Відповідно програмісти потурбувалися про це заздалегідь і винесли перекладні елементи інтерфейсу в окремі файли, аби їхній переклад здійснювався без втручання до програмного коду. Перекладач, який має працювати безпосередньо у програмному кодї, може стикнутися з багатьма проблемами через свою некомпетентність у питаннях програмування, саме тому може знадобитися взаємодія перекладача та програміста.

Процес локалізації складається з трьох етапів:

- підготовка програмного забезпечення до перекладу;
- безпосередньо переклад та адаптація програми до національних стандартів;
- тестування, відстеження та виправлення помилок.

Невід’ємною частиною процесу локалізації є управління проектом, а також контроль забезпечення якості. Істотну допомогу під час реалізації цих заходів можна отримати, використовуючи спеціальні програми, що полегшують процес локалізації, до їхнього складу входять і певні *SAT*-програми. Низка *SAT*-програм дозволяє працювати безпосередньо із системними файлами програмного забезпечення, які відповідають за інтерфейс та інші текстові елементи, що потребують перекладу, а також із файлами довідкової системи та іншої документації.

Отже, цілком очевидно, що за використання таких засобів автоматизації, як *CAT*-програми, можна підвищити ефективність процесу перекладу, скоротити витрати та час виконання різних видів робіт. У будь-якому разі доцільність застосування цих програм необхідно визначати індивідуально. Проте, як уже зазначалося вище, застосування програм із підтримкою технології накопичувальної пам'яті вже давно стало стандартом галузі, і необхідність використовувати *CAT*-програми у своїй роботі варто сприймати як звичну річ.

### **Контрольні питання**

1. Що таке *CAT*-програми і навіщо вони потрібні перекладачеві?
2. Який принцип функціонування *TM*-інструментів?
3. Які переваги та недоліки *TM*-інструментів?
4. Які найрозповсюдженіші *CAT*-програми вам відомі?
5. Що таке локалізація? Навіщо вона потрібна? Чому вона складає значні труднощі для перекладача?

### **Список використаної літератури**

1. Семенов А. Л. Современные информационные технологии : [учебное пособие для студ. перевод. факультетов высш. учеб. заведений] / А. Л. Семенов. – Москва : Академия, 2008. – 224 с.
2. Bowler L. Computer Aided Translation Technology : A Practical Introduction / L. Bowler. – Ottawa : University of Ottawa Press, 2002. – 185 с.

## LECTURE 9

### INTRODUCTION. BASIC COPYWRITING NOTIONS. BENEFITS OF THE COPYWRITING

There are probably many definitions of the term ‘copywriting’. **Copywriting** is simply a written form of communicating facts and emotions. It is a mental process.

In other words, **copywriting** is the act of writing text for the purpose of advertising or other forms of **marketing**. The definition of Tom Albrighton is – copywriting is the optimum use of language to promote or persuade.

Copywriting should not be confused with ‘**copyright**’. It means that an individual or company has the exclusive legal right to reproduce, publish, sell, or distribute someone’s work (such as books, music, artistic items). The purpose of a copyright is to protect that material and prevent illegal use of it by unauthorized agents. The owner designates the material is copyrighted with the symbol ©.

Copywriting should not be also confused with **rewriting**, because it means to write something such as a book or speech or law in a different way in order to improve it or change it. In Merriam-Webster Dictionary it is stated that **rewrite** means:

1. to write in reply;
2. to make a revision of (something, such as story): cause to be revised such as a. to put (contributed material) into form for publication; b. to alter (previously published material) for use in another publication; to revise something previously written.

The copywriting product is called copy. Copy is written content that aims to increase brand awareness and ultimately persuade a person or group to take a particular action.

There are 15 axioms of the copywriting.

*Axiom 1: copywriting is a mental process, the successful execution of which reflects the sum total of all your experiences, your specific knowledge and your ability to mentally process that information and transfer it onto a sheet of paper for the purpose of selling a product or service.*

According to Tom Albrighton copywriting is all about finding the **optimum** way to communicate. The professional copywriter is always looking for the right answer: the right length of copy, the right structure, the right tone, the right choice of words. Diligent copywriters are convinced that there is a single best solution, and they are driven to find it.

Next, **use** indicates that copywriting is a ‘useful art’: a creative activity with a practical purpose. In contrast to ‘pure’ creative writing – writing principally to entertain or provoke thought – copywriting is all about achieving a particular outcome in the real world. The value of copywriting is the extent to which it succeeds in its purpose.

**Language** is the raw material of the copywriter. Copywriting can include any carefully chosen language, including broadcast media or one-to-one communications like telephone scripts. It may also include visual language as well as verbal: the copywriter will often want to influence context and presentation (typography, design, imagery) to heighten the impact of their copy.

Most copywriting exists in order to **promote** something: products or services mainly, but also new ideas (as in rebranding exercises) or points of view (as in political marketing). The copywriter’s goal is to communicate the strengths, advantages or benefits of whatever they are promoting so their audience buys into them – whether literally or metaphorically.

**Persuasion** means getting people to think, feel or act in a certain way. Effective copywriting leads the audience by the hand across the steppingstones of reading, thinking, feeling and acting – in that order. It is all about using intangible tools – words and thoughts – to achieve an outcome in the real world. And this, ultimately, is the fascination of copywriting: making things happen with something as insubstantial as words on a page.

### **Benefits of the Copywriting**

Benefits are the key to all good copywriting. In a nutshell, copywriting that focuses on benefits is more persuasive, more compelling and sells better. We can define benefits very simply:

benefits are the good things that a product or service does (or promises to do) for its customers. Whatever you're asking readers to think, do or feel when they read your copy, it needs to offer them something good. All copywriting promises something of value or benefit to the reader.

### **Meeting a need**

The first and foremost benefit of a product or service is meeting a need. Do not underestimate the power of stating this simple truth to a reader. If your product solves a problem, make sure people know it. Your best customers are the ones who are looking for what you are selling, so make sure you cover the basics by confirming to them that you have got what they want. Making your copy too clever can sometimes obscure what you are actually offering, which is fatal.

This particularly important online, where people are impatiently searching, and you need to confirm that they have found the right thing as quickly as possible.

### **'Hard' benefits**

Then we come to 'hard', concrete benefits. These often boil down to one of three things: save time, save money or (for businesses) make money. They have tangible effects that can be measured – they are bigger, faster or cheaper. Hard benefits are powerful because they are so solid. They are based in facts and cannot be debated or contradicted. If you have got this kind of benefit, it is always worth considering whether it should lead your copy – either by forming the basis of your headline, or just by being mentioned very early on.

### **'Soft' benefits**

People are also interested in 'softer' emotional benefits such as convenience, fun, style, fashion or the sense of having made a sound buying choice. For example, when you buy jeans or trainers, you are looking for more than the optimum cost-benefit ratio – you want to buy into a brand that feels cool and appropriate for your age and style.

‘Quality’ could qualify as both a hard and a soft benefit, since its definition is so fluid. For example, it might apply to something as concrete as ‘build quality’ in engineering – the durability, tolerance and precision of the components used to make something. But in more subjective areas of judgement, such as graphic design, one person’s concept of ‘quality’ may be very far from another’s and affected by a range of personal or cultural factors.

‘Hard’ benefits are more important in business-to-business marketing, while ‘soft’ benefits appeal to the consumer. But even if you are marketing to a business, the buying decision will always be taken by a human and that human has emotions. So, if you know who they are you can appeal to those emotions.

### **Turning features into benefits**

All features of a product or service must be ‘turned outwards’ and expressed as benefits. Using the word ‘you’ is an excellent way to make a benefit feel directly relevant to the reader:

Brand/product	Feature	Benefit	Copy
L’Oréal	improve appearance of hair	feel attractive	‘because you’re worth it’
The Independent	politically neutral	be seen as discerning and intelligent	‘It is. Are you?’

### **Focusing on the customer**

One way to assess how well your copy is expressing benefits is to think about where it is predominantly focused: on the company, the product, or the customer. Imagine a conversation between the company and the customer. They are talking over a table, on which is the product being sold. It is a fairly one-sided conversation – the company is doing the talking, and the customer is listening. When the company has finished talking, the customer will decide whether or not to buy. This is essentially what happens when a customer

encounters your marketing copy. Unless they get bored or turned off, they ‘listen to what you have written as if it was a one-sided conversation. Now imagine a line stretching from the company through the product and on to the customer, as shown below:



We might call this line the ‘self-sell continuum’. The nearer the focus is to the business, the more selfish the copy will be, and the less it will sell. As it moves nearer to the customer, the more it will mean to the target audience, and the more it will sell.

### **Company-focused copy**

Purely selfish copy is all about the company: how long it has been trading, who runs it, where it is located, its principles and vision. Unless these points can be translated into benefits, they have got no place in marketing copy. This is the stuff that goes in ‘About us’ on websites, so people can easily avoid it. Admittedly, some company facts do constitute indirect reasons to buy – being a market leader, for example, is compelling – but most do not. Slightly less selfish is stuff on the boundary between the company and the product – how a product was developed, the thinking behind it and so on. This might add some value, but it’s background at best.

### **Product-focused copy**

Straightforward factual description will only sell to those who are already very clear about what they want and why. Lists of features are the kind of content that might appeal to technical staff rather than commercial managers. However, this practical content is good raw material – provided it can be reexpressed as customer benefits in order to sell harder. Copy about the interface between product and customer concerns how the product can be bought, how and when it is used, what it does and so on. This is where things start to get interesting for the reader, particularly if the text explains why the various attributes described can benefit them.

### **Customer-focused copy**

This content starts with customer concerns and goes on to explain how the product will help them, in words they will understand. Effective copywriting spends most of its time here – or, at the very least, it starts here before moving across to the other areas if and when it needs to.

### **How to achieve customer focus?**

As a newcomer and an outsider, the copywriter's job is to move the emphasis to the customer by (politely) asking questions such as:

- How does that help me as a customer?
- How does that affect my decision to buy, or not to buy?
- As a potential customer, why should I be interested?

Any points that are too company- or product-focused should be recast in terms of things the customer wants or failing that deleted. The result should be text that talks directly to the customer's own priorities, linking them clearly to the product. To confirm that this is so, compare the number of times you have said 'you' as opposed to 'we' or 'us'. There should be at least twice as many mentions of the customer as of the company. Marketing may be a one-way communication, but as with any other conversation, acknowledging the other person's point of view is more likely to get positive results.

## **REFERENCES**

1. Albrighton, Tom. The ABC of Copywriting [Electronic resource]. – Access mode : <https://www.amazon.co.uk/Tom-Albrighton/e/B07D8T8XMZ>.

2. Maslen A. The Copywriting sourcebook. How to write better copy, faster – for everything from ads to websites [Electronic resource]. – London : Marshall Cavendish Business, 2010. – 223 p. Access mode : <https://www.amazon.com/Copywriting-Sourcebook-better-everything-websites/dp/0462099741>.

3. Ng, Edmond. Copywriting Basics for Beginners. [Electronic resource]. – Access mode: <https://ezinearticles.com/?Copywriting-Basics-for-Beginners&id=5258540>.



4. Merriam-Webster Dictionary On-line [Electronic resource]. – Access mode: <https://www.merriam-webster.com/dictionary/rewrite>.

### **KEY TERMS**

Copywriting, marketing, copyright, rewriting, copy, optimum, use, language, promote, persuasion, ‘hard’ benefits, ‘soft’ benefits, customer, product, service.

### **SELF-ASSESSMENT QUESTIONS:**

1. Give the definitions of the terms copywriting, copyright, and rewrite.
2. How do we call written content that aims to increase brand awareness and ultimately persuade a person or group to take a particular action?
3. What does it mean a ‘useful art’ in terms of copywriting?
4. What is the goal of a good copywriter?
5. What term denotes ‘getting people to think, feel or act in a certain way’?
6. Name the major benefits of the copywriting?
7. Explain the difference between the ‘hard’ benefits and ‘soft’ benefits.
8. Why must all features of a product or service be ‘turned outwards’ and expressed as benefit?
9. What is the way of assessment how well your copy is expressing?
10. How do you understand the ‘self-sell continuum’ line? What are the components of it?

## LECTURE 10

### UNDERSTANDING THE PROCESS.

#### WORKING SPHERES OF A COPYWRITER. CLICHÉS

The copywriting process includes creating writing, then polishing it, adding the information, deleting sentences, changing the order of sentences or even paragraphs.

The preparation to become a copywriter involves knowledge. There are two types of it. The first is a very broad or general knowledge and the second is a very specific or targeted knowledge.

The best copywriters in the world are those who are curious about life, read a great deal, have many hobbies, like to travel, have a variety of interests, often master many skills and get bored with them and then look for other skills to master. They hunger for experience and knowledge and find other people interesting. They are very good listeners.

There is nothing really new in our life. The best copywriters have to make an experiment all the time because experiences create ideas. It is simply a matter of taking previous pieces of knowledge and putting them together in a unique and different format. Matter is not created and destroyed. Everything on earth that was here a billion years ago is pretty much here now. The only difference is that it has taken new forms.

A copywriter can write for **brand and marketing**. Brand naming is a tough call. Where a lot of the challenges of copywriting require a number of language crafts and creative skills, creating a single name, a word, is anyone's game. Having the eye to know which will work well and which will not, is not as easy as it might seem.

A copywriter can write for **advertising and direct marketing**. Advertising is a one-way communication that aims to raise awareness of a client's message to a specific audience, and it operates in every medium and in many different forms. Of all these forms, billboard campaigns are one of the most challenging: your

message has to be read from a distance, in an instant, and be both compelling and memorable.

Many large scale billboard campaigns are about creating awareness and are part of a wider media mix, but others have to pay their way – they have to generate an immediate positive response from their audience, the general public, and they have to work on their own.

A copywriter can write for **retailing and products**. The challenge in writing for retailing and products is in understanding the type of people you are writing for and what their mindset is – essentially, they are looking for something and you are making it easier for them to find it. Good copy is vital – get this right and you will be boosting the success of the product.

A copywriter can write for **company magazines, newsletters and internal communications**. A carefully thought through, well-written and well-designed company magazine is a very powerful brand communications tool that can play a leading role in ensuring that a client's employees and customers think favourably about their organization. As the writer, it is essential that you take the time to understand the context of your story fully, research the background and present your information credibly to an informed audience.

A copywriter can write for **catalogues**. There are arguments for and against long descriptive product copy in catalogues. There are also debates about the value of having fewer products per page, or including as many as possible, in terms of which creates the best overall sales results. In terms of the approach to copywriting, one school of thought says that a quality photograph will speak volumes about the product in question, and all the copy needs to do is support the image with a few key points about the product's features.

This approach works well if the client wishes to put a lot of items on the page and does not have a lot of room to play with. At the other extreme, detailed product descriptions, which explain the features and sell the benefits to the reader, are believed to work much better than simple product entries as they present compelling reasons

to purchase and work hard to close the sale. As a copywriter, you may be expected to advice on the best approach to take with catalogue copy. This exercise will help you develop a greater understanding of the challenges of long versus short copy. It will enable you to build your knowledge so that you can provide considered advice.

A copywriter can write for the **digital environment, websites**. As the copywriter for a website it is your role to draw in the visitors and to hold their limited attention span for as long as possible. How you do this depends on what your audience expect from the site, and what you feel will work best. You are providing the options to the visitor, and the way you create and present the links between sections and pages in your site is one of the most important elements of your copywriting. Visitors scan for information rather than read the text, and search for the material they want. When they find it, they are likely to want plenty of detail, often in a downloadable format.

It is better not to use clichés in your writing. **Clichés** are expressions that either have a general meaning or have ‘lost their meaning’ over time. Collins English Dictionary defines cliché as an idea or phrase which has been used so much that it is no longer interesting or effective or no longer has much meaning. Cliché came to mean a word or phrase that gets repeated often. These overused phrases do not provide a specific meaning or image. Clichés such as ‘leave no stone unturned’ have been used so much in writing that they have lost all their effectiveness. Why should not we use clichés?

Clichés make you seem boring. By using a cliché, we are telling our reader that we lack originality, making them want to yawn and stop reading your paper.

Clichés make your writing and argument interchangeable with anybody else’s. Make sure that your argument and writing are specific to you and your writing task.

Clichés are vague. It is best to use the most precise wording in order to present evidence and support your arguments as clearly

as possible. Specific details and explanations make better evidence than generalizations and trite phrases.

Clichés make you seem lazy. They are a hedge when you do not want to do creative work.

Clichés make you lose credibility. Your reader will not trust you as an authoritative source if you cannot come up with a better description than a cliché.

Clichés are poor substitutes for actual evidence. Because clichés are not specific, they do not offer strong enough commentary to prove your point. Make sure that every sentence of your paper is working toward a goal by eliminating meaningless phrases.

Clichés are often idioms. This makes clichés difficult to translate into other languages because their meaning will not always be understood by people from different cultures.

## **REFERENCES**

1. Albrighton, Tom. The ABC of Copywriting [Electronic resource]. – Access mode : <https://www.amazon.co.uk/Tom-Albrighton/e/B07D8T8XMZ>.

2. Maslen A. The Copywriting sourcebook. How to write better copy, faster – for everything from ads to websites [Electronic resource]. – London : Marshall Cavendish Business, 2010. – 223 p. Access mode: <https://www.amazon.com/Copywriting-Sourcebook-better-everything-websites/dp/0462099741>.

3. Ng, Edmond. Copywriting Basics for Beginners. [Electronic resource]. – Access mode : <https://ezinearticles.com/?Copywriting-Basics-for-Beginners&id=5258540>.

4. Collins English Dictionary On-line [Electronic resource]. – Access mode : <https://www.collinsdictionary.com/dictionary/english/cliche>.

## **KEY TERMS:**

Brand and marketing, experience, advertising and direct marketing, retailing and products, company magazines, newsletters

and internal communications, catalogues, digital environment, websites, clichés.

### **SELF-ASSESSMENT QUESTIONS:**

1. What does include the copywriting process?
2. Who are the best copywriters?
3. Name the types of knowledge.
4. Why is it important to make an experience all the time?
5. What is the goal of large scale billboard campaigns?
6. Explain the main challenge in writing for retailing and products.
7. When does the approach to copywriting concerning writing for catalogues work well?
8. What is your role as the copywriter for a website?
9. Give the definition of the term ‘cliché’.
10. Why shouldn’t a good copywriter use cliché in his/her own advertising?

## **LECTURE 11**

### **THE ELEMENTS OF THE ADVERTISEMENT (AD)**

1. **Headline:** gets your attention and draw you to the subheadline.
2. **Subheadline:** gives you more information and further explain the attention-getting headline.
3. **Photo or drawing:** gets your attention and illustrates the product more fully.
4. **Caption:** describes the photo or drawing. This is an important element and one that is often read.
5. **Copy:** conveys the main selling message for your product or service.
6. **Paragraph headings:** breaks up the copy into chunks, thereby making the copy look less imposing.
7. **Logo:** displays the name of the company selling the product.

8. Price: let the reader know what the product or service costs. The price could be in large type or could be buried in the copy.

9. Response device: gives the reader a way to respond to the ad, by using the coupon, toll-free number or ordering information, usually near the end of the ad.

10. Overall layout: provides the overall appearance for the ad, by using effective graphic design for the other elements.

One of the most important axioms is number two.

*Axiom 2: all the elements in an advertisement are primarily designed to do one thing and one thing only: get you to read the first sentence of the copy.*

### **Defining Headline:**

1. In print advertising, it is the headline and visual.
2. In an email, it is the From and Subject lines.
3. In radio or TV ad, it is the first few seconds.
4. In a direct mail-package, it is the outer envelope or first few sentences of the letter.
5. In a press release, it is the lead paragraph.
6. In a sales brochure or catalogue, it is the front cover.
7. In a sales presentation, it is the first few slides or flip charts.
8. On a website, it is the first screen of the homepage.

### **Headline Functions:**

1. Get attention;
2. Select the audience;
3. Deliver a complete message;
4. Draw the reader into the body copy.

### **Headlines Get Attention if:**

1. Appeal to the reader's self-interest.
  - *"Why Swelter Through Another Hot Summer?"* – GE air conditioners.
2. Give the reader news:
  - *"Introducing....."*
3. Offer freebies and use the word *"free"*:
4. Offer useful information:
  - *"How to Bake Beans"* – Van Camp's.

5. Proven headline buzzwords:

– *Free, how to, why, sale, quick, easy, bargain, last chance, guarantee, results, proven, save.*

### **Headlines Select the Audience**

Headline can qualify the buyer – an ad for a \$85,000 sports car should (somehow) say, *“This is for rich folks only!”*

Select the right audience, screen out others:

– *“We’re looking for People to Write Children’s Book!”*

– *“Is Your Electric Bill Too High?”*

### **8 Types of Headlines**

Direct headlines

Indirect headlines

News headlines

How-to headlines

Question headlines

Command headlines

Reason-why headlines

Testimonial headlines

### **4 U’s Formula for Writing Effective Headlines**

1. Urgent – gives the reader reason to act now instead of later.

– *“Make \$100,000 working from home this year”* (add *‘this year’*)

2. Unique – says something new or something familiar in a new way.

– *“Why Japanese women have beautiful skin?”*

3. Ultra– specific – ‘fascinations’ tease the reader into reading further.

– *“Best time to file for a tax refund”*

4. Useful – appeals to reader’s self-interest by offering a benefit.

– *“An Invitation to Ski and Save”*

### **The Copy**

You sell more merchandise when you write clear copy.  
Causes for confusing copy:



– ‘borrowed interest’, lengthy sentences, clichés, big words, not getting to the point, a lack of specifics, technical jargon, and poor organization.

### **11 Tips for Writing Clear Copy:**

1. Put the reader first, use ‘you’;
2. Carefully organize your selling points;
3. Break the writing into short sections (long paragraphs are intimidating);
4. Use short sentences – easier to read (best average length = 14 to 16 words);
5. Use simple words;
6. Avoid technical jargon;
7. Be concise – avoid wordy phrases, unnecessary adjectives, redundancies, etc.;
8. Be specific;
9. Go straight to the point – start selling in the first line.
10. Write in a conversational style.
11. Avoid sexist language.

### **Stylistic Tricks of Writing a Copywriting Product**

- End sentences with a preposition;
- Use sentence fragments;
- Begin sentences with conjunctions – easy transitions;
- Use one-sentence -paragraphs – lively change of pace;
- Use graphic techniques to emphasize words or phrases – underlining, capitalization, boldface, italics, arrows and notes, etc.;
- Use bullet points.

Having written a copy, you have to ask next question:

### **Copywriter’s Checklist:**

- Does the copy fulfil promise from the headline?
- Is the copy interesting?
- Is it easy to read?
- Is it believable?
- Is it persuasive?
- Is it specific?

- Is it concise?
- Is it relevant?
- Does it flow smoothly?
- Does it call for action?

Feature of the product or service is its descriptive fact; it is what the product/service is or has. So you have to write about benefits (what the product does). It is what the user gains as a result of the feature. Customers do not buy products or services. They buy what these products or services do for them.

### **The Motivating Sequence:**

5-step formula to writing copy that sells:

1. Get attention:
  - headline & visual should focus on single strongest benefit;
  - do not use clever phrases, puns, irrelevant information, while serving strongest benefit for a big finish inside the copy;
2. Show a need:
  - with most products, the need for the product may not be obvious;
  - demonstrate that need.
3. Satisfy the need and position your product as a solution:
  - after demonstrating need, quickly show that your product/service satisfies the need, answers questions, or solves the problem;
4. Prove your product/service can do what you say it can do:
  - testimonials, comparisons to competition, studies/research, etc.
5. Ask for action:
  - last step must be a call for action for the next step in buying process;
  - if possible, give an incentive for responding now: discounts, etc.

Practice is a great teacher – the more you write, the better you get.

First, you learned that you have, in your life, experienced general knowledge through your actions, circumstances and

personality. You have also been given the tools to obtain specific knowledge, such as the ability to inquire, read and research. Then you learned that practice is a great teacher – that the more you write, the better you get. And finally, you learned that copywriting is the mental process of transferring what is in your head onto a sheet of paper.

### **REFERENCES**

1. Albrighton, Tom. The ABC of Copywriting [Electronic resource]. – Access mode : <https://www.amazon.co.uk/Tom-Albrighton/e/B07D8T8XMZ>.
2. Maslen A. The Copywriting sourcebook. How to write better copy, faster – for everything from ads to websites [Electronic resource]. – London : Marshall Cavendish Business, 2010. – 223 p. Access mode: <https://www.amazon.com/Copywriting-Sourcebook-better-everything-websites/dp/0462099741>.
3. Ng, Edmond. Copywriting Basics for Beginners [Electronic resource]. – Access mode : <https://ezinearticles.com/?Copywriting-Basics-for-Beginners&id=5258540>.

### **KEY TERMS:**

Headline, subheadline, photo, drawing, caption, copy, paragraph heading, logo, price, response device, overall layout, testimonials, discounts, buzzwords, bullet points, visuals, flip charts.

### **SELF-ASSESSMENT QUESTIONS:**

1. What are the main elements of the advertisement?
2. Expand the essence of the second axiom.
3. Define the headline in all kinds of media.
4. Name the functions of the headline.
5. When does the headline get attention?
6. What types of headlines do you know?
7. Explain 4 U's formula for writing effective headlines.

8. What are the causes for confusing copies?
9. Represent the stylistic tricks of writing a copywriting product.
10. Describe the 5-step formula to writing copy that sells.

## **LECTURE 12**

### **FIRST SENTENCE. THE ENVIRONMENT OF THE COPYWRITER**

The headline must be so powerful and compelling that you must read the subheadline, and the subheadline must be so powerful that you are compelled to read the first sentence, and the first sentence must be so easy to read and so compelling that you must read the next sentence and so on, straight through the entire copy to the end.

What is the purpose of the first sentence? It is to convey a benefit or explain a feature, that would be impossible. How could a short first sentence do anything more than gets you to read it? The correct answer is: the purpose of the first sentence is to get you to read the second sentence. Nothing more, nothing less. What the purpose of the second sentence is: to get you to read the third sentence. The same situation is with the third, the fourth, the fifth and the next sentences.

*Axiom 3: the sole purpose of the first sentence in an advertisement is to get you to read the second sentence.*

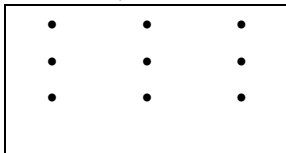
The copywriter has to control over the environment. The environment you choose is created in both the graphic elements and the copy, but especially the copy – by the way you phrase your words, the choice of words and the level of integrity you convey. And to create that environment, you attract the customer's attention (the headline, photos, logo, etc.) and then you have got to get the person to read the first sentence by making it so simple and so compelling that the reader cannot help but read it. And the next sentence and then the next.

*Axiom 4: your ad layout and the first few paragraphs of your ad must create the buying environment most conducive to the sale of your product or service.*

Creating the ideal buying environment comes from experience and the specific knowledge you get from studying your product and potential customer. It comes from understanding the nature of your product or service. The next step is to get the customer to nod in the affirmative and agree with you or at least you make truthful statements that the prospect knows are correct and would concur with. Make sure that the prospect does not disagree with something you are saying. But as long as the reader keeps saying yes or believes what you are saying is correct and continues to stay interested, you are going to be harmonizing with the reader and you and the reader will be walking down that path toward that beautiful room in the art gallery.

*Axiom 5: get the reader to say yes and harmonize with your accurate and truthful statements while reading your copy.*

One of the best ways to understand this is the following nine-point puzzle. This puzzle is as an example of an assumed constraint. The rules to solve this puzzle are very simple. You must draw four straight lines and connect all the points without taking your pen off the page. In short, all the lines have to be connected.



Most people who try to solve the puzzle make one fatal error and it is simply that they do not realize the constraint they subconsciously put upon themselves. As you can see, we were constrained by the box surrounding the points and did not go out of the box in order to find the solution. Often, to solve that very important problem, you have got to go out of the problem area itself to find the answer.

*The next Axiom 6: when trying to solve problems, do not assume constraints that are not really there.*

You never really know what will work or what will not. If you believe in your idea, do it. Step out of those assumed constraints. When thinking about a problem or looking for a solution, do not rule anything out. Sometimes that big idea will come to you if you step out of the traps that we very often fall into.

The following business examples of other historical assumed constraints:

1. *“There is no reason anyone would want a computer in their home.”*

This was said by Ken Olsen, president, chairman and founder of Digital Equipment Corporation, in 1977.

2. *“Computers in the future may weigh no more than 1.5 tons.”*

This was quoted by Popular Mechanics magazine forecasting the relentless march of science in 1949.

The first sentence is so important and in order to make it so compelling to read you have to make it short. Some typical first sentences might be:

*Losing weight is not easy.*

*It's you against a computer.*

*It's easy.*

*It had to happen.*

*Hats off to IBM.*

Each sentence is so short and easy to read that your reader starts to read your copy almost as if being sucked into it.

Many magazines use a variation of this technique in their articles. They start an article not with a very short sentence but maybe with very large type. Once they have you sucked into reading the copy and you turn the page to read the rest of the article, you notice that the typeface has gotten smaller. But that is okay. The purpose of the large type was to get you into the article and it worked. So your first sentence should be very compelling by virtue of its short length and ease of reading. No long multisyllabic words. Keep it short, sweet and almost incomplete so that the reader has to read the next sentence.

*Axiom 7: keep the copy interesting and the reader interested through the power of curiosity.*

It means that at the end of a paragraph, it is better to put a very short sentence that offers some reason for the reader to read the next paragraph. You can use such sentences as:

*But there's more.*

*So read on.*

*But I didn't stop there.*

*Let me explain.*

*Now here comes the good part.*

These seeds of curiosity cause you to subconsciously continue reading even though you might be at a point in the copy where the copy slows down. In print, the ideal situation is to create such interesting and compelling copy that you do not need the seeds of curiosity, but often that is very difficult. And using these seeds of curiosity enhances most copy. But like every good thing, do not overdo it.

Seeds of curiosity can be used at the beginning of an ad where you mention some benefit or payoff that you are going to reveal somewhere in your copy. In short, the reader has to read the entire ad to find it.

## REFERENCES

1. Albrighton, Tom. The ABC of Copywriting [Electronic resource]. – Access mode : <https://www.amazon.co.uk/Tom-Albrighton/e/B07D8T8XMZ>.

2. Maslen A. The Copywriting sourcebook. How to write better copy, faster – for everything from ads to websites [Electronic resource]. – London : Marshall Cavendish Business, 2010. – 223 p. Access mode: <https://www.amazon.com/Copywriting-Sourcebook-better-everything-websites/dp/0462099741>.

3. Ng, Edmond. Copywriting Basics for Beginners [Electronic resource]. – Access mode : <https://ezinearticles.com/?Copywriting-Basics-for-Beginners&id=5258540>.

## **KEY TERMS:**

First sentence, sole purpose, ad layout, buying environment, puzzle, assumed constraints, seeds of curiosity.

## **SELF-ASSESSMENT QUESTIONS:**

1. What is the purpose of the first sentence?
2. How could a first sentence do anything more than gets you to read it?
3. What is the purpose of the second sentence?
4. How to create the ideal buying environment?
5. How to solve the problems with assumed constraints?
6. Give the business examples of the historical assumed constraints.
7. What is the purpose of the large type?
8. What should be the length of the first sentence?
9. What is the meaning of the power of curiosity?
10. Where is it better to put the seeds of curiosity?

## **LECTURE 13 EMOTIONS IN ADVERTISING**

There are just three points to remember about the emotions in advertising.

Emotion principle 1: every word has an emotion associated with it and tells a story.

Emotion principle 2: every good ad is an emotional outpouring of words, feelings and impressions.

Emotion principle 3: you sell on emotion, but you justify a purchase with logic.

Let us take the last point first. Why do you think people buy the Mercedes-Benz automobile? The answer: we buy on emotion and justify with logic. When most people buy a Mercedes, they tell people that the reason they bought it was because of a series of technical features. But the real reason is they want to own a



prestigious car and belong to the crowd that drove a Mercedes. In reality such people lie to themselves using logic – something that they really believe.

In writing copy for an advertisement, often you get your reader in an emotional frame of mind as a result of the environment you have created, and logic becomes less important. For example, you can use the phrase near the end of your ads,

*“If you aren’t absolutely satisfied, return your product within 30 days for a prompt and courteous refund.”*

Who ever heard of a refund being courteous? It does not matter. The emotion or the feel of that phrase really says that we are a very respectful and understanding company that will return your money very promptly. Often, a phrase or sentence or even a premise does not have to be correct logically. As long as it conveys the message emotionally, it not only does the job, but does it more effectively than the logical message.

The next examples are:

*Example 1: The old woman in the motel.*

*Example 2: The little old lady in the cottage.*

The word cottage gives to the copy a better emotional feel. You should not materially change the facts of a situation to suit an emotional feeling, but you can add some emotional details. Sometimes changing a single word will increase response in an ad.

*Axiom 8: never sell a product or service. Always sell a concept.*

Other terms commonly used are Big Idea, or USP (unique selling proposition). Whatever it is called, it means basically the same thing. You sell the sizzle and not the steak – the concept and not the product. The only exception to this rule is when the product is so unique or new that the product itself becomes the concept. Finding the concept is often not easy. It takes all the skills of a conceptual thinker to come up with the right idea and the right position.

*Axiom 9: the incubation process is the power of your subconscious mind to use all your knowledge and experiences to*

*solve a specific problem, and its efficiency is dictated by time, creative orientation, environment and ego.*

Do plenty of thinking about what you want to write. Put all your thoughts down on paper. This is just preparation. Do something else. Forget about the project. Do something pleasurable – a stroll in the park, a walk down the street or lunch with a good friend. Whatever you do, let it be a total diversion from what you are currently working on, and do not even think of the copy project.

Whether you realize it or not, you are actually working on the ad constantly even though you have put it out of your mind. Your subconscious mind is actually processing everything you have learned – all of that data that you have accumulated in general and all of the information in particular. And your mind is then taking all of that data and running it through everything you know about copywriting and communications, mentally preparing the first version of your ad copy. You are just out having a good time while your brain is working like crazy. This entire subconscious activity is called the incubation process, and the time you are giving to it is called the incubation period.

*Axiom 10: copy should be long enough to cause the reader to take the action you request.*

There are two variants about the copy length: short or long. If the copy is interesting, the reader will read it all even if it is long. The main demand is to tell the full story of your product.

*Axiom 11: every communication should be a personal one, from the writer to the recipient, regardless of the medium used.*

Every advertisement should be a personal message from the advertiser to the prospect. An effective way to do this in print is to use a byline. Use your name or the name of somebody in your organization such as the president – like the news organizations do in a magazine or newspaper article. This allows you to use words like *I* and *me* and *we* and *you*. Let us look at the example of the ad for BluBlocker sunglasses.

***Headline: Vision Breakthrough***

**Subheadline:** *When I put on the pair of glasses what I saw I could not believe. Nor will you.*

**Byline:** *By Joseph Sugarman*

**Copy:** *I am about to tell you a true story. If you believe me, you will be well rewarded. If you don't believe me, I will make it worth your while to change your mind. Let me explain.*

Such words as *I* and *you* and *me* were used – all very personal words used in a one-to-one conversation. This shows how very personal you can get in copy and still convey a very powerful thought or develop the environment and slippery slide you need to cause your reader to continue reading and then respond.

As you start to write copy to reach and motivate an individual, think in terms of writing in the first person with a personal message.

### **The Copy Sequence**

INTEREST AND EXCITEMENT
DRAMA
WHY DIFFERENT
HOW TO PLAY
UNIQUE FEATURES
JUSTIFY PURCHASE
LASTING PLAY VALUE
SERVICE
ASK FOR ORDER

With the first paragraph it is better to create *interest and excitement* for the product/service.

With the second paragraph copywriters start to weave the *drama* of the product and the differences between advertising product/service and similar one.

In the next block of copy you explain why and how the product/service is *different*, *how to use it* and mention some of its *unique features*.

So the next block should *justify the purchase*. Present here all variants why this is a great purchase.

By now the customer is saying to himself, “*Okay, I want to get the unit and I can justify the purchase, but what if I use it*», in such a case it is better to go into the fact that it has *lasting play value*.

Copywriters then raise the *service* question in the ad copy and answer it. The point of each of these blocks of copy is that they are logically placed as if to anticipate the next question a prospect is going to ask – all in an environment that you have created and all flowing logically to the last part of the ad when you *ask for the order*.

*Axiom 12: the ideas presented in your copy should flow in a logical fashion, anticipating your prospect’s questions and answering them as if the questions were asked face-to-face.*

The next axiom concerning editing is:

*Axiom 13: in the editing process, you refine your copy to express exactly what you want to express with the fewest words.*

This may mean that you rearrange the words you wrote to make the thought more direct. Or it may mean cutting out words that have little contribution to the overall feel of the ad. It may mean substituting new words that express your thoughts better. And it may even mean adding words to clarify a thought. But the goal in writing ad copy is to express the thoughts you want to convey in the most powerful way but with the fewest words.

## REFERENCES

1. Albrighton, Tom. The ABC of Copywriting [Electronic resource]. – Access mode : <https://www.amazon.co.uk/Tom-Albrighton/e/B07D8T8XMZ>.

2. Maslen A. The Copywriting sourcebook. How to write better copy, faster – for everything from ads to websites [Electronic resource]. – London : Marshall Cavendish Business, 2010. – 223 p. Access mode : <https://www.amazon.com/Copywriting-Sourcebook-better-everything-websites/dp/0462099741>

3. Ng, Edmond. Copywriting Basics for Beginners. [Electronic resource]. – Access mode : <https://ezinearticles.com/?Copywriting-Basics-for-Beginners&id=5258540>.

### **KEY TERMS:**

Emotions, logical message, concept, unique selling proposition, incubation process, byline, interest and excitement, unique features, justify the purchase, lasting play value, editing process.

### **SELF-ASSESSMENT QUESTIONS:**

1. What are three points to remember about the emotions in advertising?
2. Why does logic become less important in writing copy for an advertisement?
3. What is it better to sell in advertising? A product/service or a concept?
4. Explain the meaning of the incubation process?
5. What is the role of the byline in an advertisement?
6. Is there any differences between incubation process and incubation period?
7. Describe the copy sequence.
8. What is the point of each of the blocks of copy?
9. What is the goal in writing ad copy?
10. What is the goal in editing ad copy?

## **LECTURE 14 SOME PRINCIPLES OF EDITING**

1. Look for any ‘that’ words. For example,  
“*And, if you’ve tried it, you know that...*” The words up to and including ‘*that*’ can very often be eliminated.
2. Edit for rhythm. Make sure that you vary the length of sentences so they do not sound monotonous.

3. Consider combining sentences. For example, “*A bathroom scale is like a report card. It’s a feedback mechanism that tells you how well you’ve done*”.

It can be condensed into “*Your bathroom scale is like a report card – a feedback mechanism that tells you how well you’ve done.*” It is saved only one word by doing this, but it made sense to combine the sentences and eliminating even one word is a good move.

4. Eliminate unnecessary words. Look at the word ‘*the*’ in the phrase ‘*and seeing the positive results.*’ The word ‘*the*’ can easily be eliminated without changing the meaning so that the final sentence will read, “*and seeing positive results.*”

5. Rearrange thoughts so they flow better. Note that in the first draft, the flow of the copy pointed out that the scale was a report card and the second thought was that part of the pleasure of a weight loss program was stepping on a scale and seeing the positive results. By reversing these two thoughts, advertisement becomes more emotional by focusing on the pleasure of using a scale when losing weight and then it is explained why. This sounds a lot better and more logical from a flow standpoint than the first draft.

One more example:

**Headline:** “*The more you learn about our DC-10, the more you know how great it really is.*”

**Byline:** *Pete Conrad, Former Astronaut, Division Vice President, McDonnell Douglas*

**Copy:** *I’ve watched airplanes and spacecraft take shape for much of my adult life. I’m certain that nothing made to fly has ever been designed or built to more exacting standards than our DC-10.*

*Eighteen million engineering man hours were invested in this plane’s development. That includes 14,000 hours of wind tunnel testing, as well as full– scale ‘fatigue testing’ for the equivalent of 40 years of airline service.*

*I’m convinced that the DC-10 is the most thoroughly-tested jetliner ever built. Along with U.S. Government certification, the DC-10 has passed structure tests just as demanding, in their own way, as*

*those required of U.S. Air Force fighter planes. The DC-10 fleet demonstrates its dependability flying more than a million miles a day and serving 170 cities in 90 countries around the globe.*

In this ad the objection was totally ignored. The resulting ad was hollow. The ad then ended with a place to write McDonnell Douglas to get more information.

The copy was good copy – very persuasive in terms of building confidence in the plane. And all this was presented by a former astronaut to add credibility. But it lacked an important opening that would have made the copy many times more effective.

This is an example how to improve this ad:

**Headline:** *DC-10's Big Secret*

**Subheadline:** *You've heard a lot of bad publicity about the DC-10. But here's something you may not have known.*

**Byline:** *By Pete Conrad*

**Copy:** *It was horrible. When American Airlines flight 191 crashed at O'Hare in Chicago last May 25th, hundreds of people lost their lives in what was considered one of the worst plane crashes in American history. The plane – a DC-10. But as the facts emerged, it was learned that a series of coincidences resulted in a hydraulic malfunction that in all likelihood would never happen again. But it did happen. And in the subsequent weeks, a series of fail-safe systems were installed that make the DC-10's hydraulic system among the safest of any jetliner.*

*In addition, two other recent crashes of DC-10s were determined to be totally not the fault of the airplane but of the pilots. But because of the negative publicity generated by the American Airlines crash and these other two crashes, we have been even more diligent. All airlines are required to give each DC-10 a complete inspection every 50 hours instead of the required 100 hours of flying...*

After reading the second version of the ad, people would leave with a positive feeling about both the company and the message. We would think that it was a sincere effort to dispel those

false impressions on the safety of the DC-10. The message is an emotional presentation that shows concern, integrity and leadership.

Keep in mind that the copy they had was very good. They just left off the entire front end of the ad and were avoiding the real issue – the events that actually happened that prompted their ad.

The lesson to be learned here is to realize the importance of raising an objection, regardless of how embarrassing or detrimental it may seem, and then doing your best to resolve it.

### **How to organize Product Brochure Copy**

Introduction – capsule description.

Benefits – why customers should buy.

Features – how it is different from competitor.

‘How it works’ – description of how it works.

Types of users (markets)

Applications

Product availability – models, sizes, etc.

Technical specifications

Questions and answers – FAQ

Company description

Support – info on delivery, installation, guarantees.

The next step – how to order or get more info

### **How to organize Service Brochure Copy**

Intro – outlines services, types of accounts handled, reasons why reader should be interested.

Service offered – detailed descriptions.

Benefits – describes what readers will gain and differentiates you from competitor.

Methodology – outline’s firm’s methods.

Client list – list of customers who endorse service.

Testimonials

Fees and terms.

Biographical info – key employees

The next step – instructions on what to do next.



## **Catalogues vs. Brochures**

Brochures tell in-depth story about a single product. Catalogues give short descriptions of many products. Catalogue description must be terse.

Brochure's mission is to provide enough info for reader to take next step in buying process. Catalogues are mail-order vehicles from which you order; salesperson not needed for catalogues.

### **Tips for Writing Catalogues**

1. Write snappy headlines – strong benefit or descriptive adjective.
2. Include a letter from the manufacturer.
3. Give all key product facts.
4. Devote the most space to best-sellers.
5. Use techniques that stimulate sales – toll-free phone numbers, credit cards accepted, free gifts.
6. Make the order form simple
7. Indicate discounted items – 25 % off! or 11,95\$ – 8?95\$.

### **Other types of Sales Literature**

Annual reports

Flyers

Broadsides – flyers folded for mailing

Invoice stuffers – small promos mailed with monthly bills/statements

Circulars – 4–8 pages, full colour, contain coupons for local products

Pamphlets – similar to brochures but more general

White papers – in-depths reports about product or industry under guise of unbiased information.

## **REFERENCES**

1. Albrighton, Tom. The ABC of Copywriting [Electronic resource]. – Access mode : <https://www.amazon.co.uk/Tom-Albrighton/e/B07D8T8XMZ>.
2. Maslen A. The Copywriting sourcebook. How to write better copy, faster – for everything from ads to websites [Electronic

resource]. – London : Marshall Cavendish Business, 2010. – 223 p.  
Access mode : <https://www.amazon.com/Copywriting-Sourcebook-better-everything-websites/dp/0462099741>.

3. Ng, Edmond. Copywriting Basics for Beginners. [Electronic resource]. – Access mode : <https://ezinearticles.com/?Copywriting-Basics-for-Beginners&id=5258540>.

### **KEY TERMS:**

Editing, rhythm, combine, eliminate, catalogue, brochure, sales literature, FAQ, benefits, pamphlet, annual reports, flyers, broadsides, invoice suffers, circulars, white paper.

### **SELF-ASSESSMENT QUESTIONS:**

1. Why do we need to look for any ‘that’ words in the editing process?
2. Is it necessary to consider combining sentences in the editing process and why?
3. Is it possible to eliminate unnecessary words without changing the meaning?
4. How to organize product brochure copy?
5. What information is it obligatory to represent in the component ‘support’ of the product brochure copy?
6. How to organize service brochure copy?
7. Explain the contradiction between catalogues and brochures.
8. What are the tips for writing catalogues?
9. What are the techniques that stimulate sales?
10. What types of sales literature do you know?

## **LECTURE 15**

### **CONCEPTS THAT YOU SHOULD REVIEW FOR EACH AD YOU WRITE**

1. Typeface: this element is really important. If you are a graphic designer, you know that each typeface has its own personality, emotion and legibility. And that is the point of this subject. You have got to determine the combination of personality and legibility that will make your ad easy to read and inviting. Serif type has short protruding lines whereas the other style of type, called ‘sans serif’, does not. Serif type produced greater comprehension than sans serif type and by a 5 to 1 margin. Another important factor is the legibility of all your typefaces in other parts of your ad such as the headline, subheadline and various headings. Fancy type might look elegant to the type designer, but if it can’t be read it has no value. The most important role a typeface has is to allow the greatest comprehension possible, and the second role, less important by far, is to convey the image of a company.

2. Second sentence: this sentence is almost as important as the first. You have got to maintain interest, so you must also create another sentence with a compelling reason to cause your reader to want to continue. You must continue this momentum throughout the first and second paragraphs. Put aside any facts, benefits or product features. Your only goal is to get the momentum going and create that buying environment.

3. Paragraph headings: they have little to do with selling or presenting your product. They simply make the copy look more inviting so your reader will start the reading process. Use paragraph headings to break up copy in the middle of columns but not at the very end or beginning. Avoid placing paragraph headings right next to each other in two adjoining columns.

4. Product explanation: sounds simple. Sounds basic. Many ads leave out the simple step of explaining what the product does. A rule of thumb here is to explain a complicated product in a very simple way and explain a simple product in a very complex way.

5. New features: highlight those features that make your product or service new, unique or novel. This might appear to be the same as the copy element ‘Product Explanation’, but it is different. Here you are revealing not just the features of the product, but the features that distinguish it from anything else on the market.

6. Anticipate objections: this is a very important element to consider when writing copy. If you feel that your prospect might raise some objection when you are describing a product, then raise the objection yourself. Remember, you are not in front of the consumer and you have to sense what the next question might be. If you sense that there might be an objection and you ignore it, it is like ignoring that consumer. You will not get away with it. The consumer is too sharp and will not buy.

7. Resolve objections: just as you have to recognize objections, it is your opportunity and duty to resolve the objections, too. You must be honest and provide alternative solutions or dispel the objections completely.

8. Gender: who is the consumer? Male, female or both genders? Are they female golf players, lady pilots or professional women? Make sure there are no sexist comments that would offend any group, and know your target audience so that you can communicate in their terms.

9. Clarity: You remember that your copy should be clear, simple, short and to the point. Avoid big words that confuse those who do not know them and which often establish the writer as a pompous snob – unless, of course, you’re trying to appeal to pompous snobs. Keep it simple. The clearer the copy and the more concise, the easier it will be for people to read and get on the slippery slide and stay there. The only exception to this rule is when you give a technical explanation.

10. Clichés: avoid the obvious ones: “*Here’s the product the world has been waiting for,*” or “*It’s too good to be true.*” If you feel inclined to use a cliché, do not. Clichés seem to be used when you have nothing really significant or good to say and must fill up space. How do you know if you are writing a cliché? If it sounds like you

are writing typical advertising copy a copywriter may have written 20 years ago.

11. Rhythm: it has no distinct pattern: a short sentence, then a long sentence followed by a medium sentence followed by a short sentence and then another short sentence and then one really long sentence. In short, a mixture of sentence lengths that, when read together, gives a sense of variety and rhythm. Think about how copy would sound if all the sentences were very short or very long or all had a distinct predictable pattern. Pretty boring. And that is the point of copy rhythm. Vary your sentences; vary their length to give your copy a rhythm.

12. Service: if you are selling an expensive product or one that is not easily returned for service, you must address the question of service and convey the ease of that service to the consumer. Often the mention of a brand name manufacturer is all that is needed to establish ease of service. But if there is a remote possibility that the consumer would still ask about service, then you must address this issue in your ad.

13. Physical facts: in copy you must mention all the physical facts about a product (weight, dimensions, size, limits, speed, and the like) or you risk reducing your response. Sometimes you might think that a certain dimension is not really important or the weight may not be necessary. But it is not true.

14. Trial period: you must offer a trial period for any product that the consumer cannot touch or feel at the time of purchase. The only time you can make an exception to this rule is when the value is so strong and the product so familiar that the consumer is willing to take the risk. Make sure your trial period is at least one month, or even better, two months. Tests have proven that the longer the trial period, the less chance the product will be returned and the more confidence the consumer will have in dealing with you and purchasing the item.

15. Price comparison: whenever possible, offering a price comparison to another product establishes value in the mind of the purchaser. This points out one of the really important considerations

that motivate consumers to buy – namely that they are getting real value.

16. Testimonials: a testimonial is a good way to add credibility if it is from a very credible person or organization. This approach can be used not only in the copy but in the headline or photo. See if your ad copy could use a celebrity for a testimonial, but make sure the testimonial makes sense for the product. Make sure that the celebrity matches with the product and adds credibility. The use of a celebrity who does not make sense for your product or does not add credibility could backfire, have the wrong effect and kill sales if the offer is not believable.

17. Price: Another important copy point to consider is the price. Should the price be obvious? Should it be set in large type? Small? If you are selling a product or service at a very good price, then set the price in larger type. After all, you want people to see that benefit very clearly.

If the product is expensive and it is not the price that will sell it, you want to underplay it. Do not hide it; just underplay it. By putting the price in a logical position in the copy – whether it be in the coupon, which is the ideal place, or in the copy highlighted with a bold typeface – you are answering the question that the reader asks without knowing when the reader is ready to ask the question. The reader simply scans the ad and if the price is in bold or in the coupon, the price will pop out and answer the question.

18. Offer summary: it is a really good idea to summarize what you are offering the consumer somewhere near the end of your ad. Many ads, by the way, miss this important point.

19. Avoid saying too much: this is probably the biggest mistake most copywriters make. But there one positive issue in saying too much. It is normal to say as much as you can about a subject and then refine the copy to a point where it flows smoothly. This usually means editing and reducing the copy length until it has rhythm and it flows. But this could take time and involves a few steps.

20. Ease of ordering: make it easy to order. Use a toll-free number, a coupon, a tear-off reply card or any vehicle that is easy to understand and use. It is better to use a coupon with dotted lines. It usually generates more response because the dotted lines clearly convey at a glance that you can order the product from the ad.

21. Ask for the order: always ask for the order near the end of your ad. This is often forgotten by many copywriters. At the end of an ad, state like the following or something similar: *“I urge you to buy this at no obligation, today.”* You have got to ask for the order, and if you are doing it right, it should be at the very end of the ad where you have finished selling your prospect, you have summarized the offer and your prospect is ready to buy.

## REFERENCES

1. Albrighton, Tom. The ABC of Copywriting [Electronic resource]. – Access mode : <https://www.amazon.co.uk/Tom-Albrighton/e/B07D8T8XMZ>

2. Maslen A. The Copywriting sourcebook. How to write better copy, faster – for everything from ads to websites [Electronic resource]. – London : Marshall Cavendish Business, 2010. – 223 p. Access mode: <https://www.amazon.com/Copywriting-Sourcebook-better-everything-websites/dp/0462099741>.

3. Ng, Edmond. Copywriting Basics for Beginners. [Electronic resource]. – Access mode : <https://ezinearticles.com/?Copywriting-Basics-for-Beginners&id=5258540>.

## KEY TERMS:

Concept, typeface, objection, testimonials, order, summary, trial period, cliché, gender, clarity, price, rhythm, service.

## **SELF-ASSESSMENT QUESTIONS:**

1. Explain the difference between serif and sans serif type. Which one is it better to use in your ad copy?
2. What is the most important role a typeface has?
3. If you feel that your prospect might raise some objection when you are describing a product, what should you do?
4. Should a copywriter mention all the physical facts about a product?
5. What should a copywriter offer when the consumer cannot touch or feel the product at the time of purchase?
6. What is a good way to add credibility to your copy product?
7. Where should a copywriter put the price if the product or service is expensive and it is not the price that will sell it?
8. What is the biggest mistake that most copywriters make?
9. Why should a copywriter use in his/her ad a toll-free number, a coupon, or a tear-off reply card?
10. What is it necessary to put at the very end of the ad?

## **LECTURE 16**

### **THE PSYCHOLOGICAL TRIGGERS. GRADE LEVEL. SEVEN STEPS TO WRITING GREAT COPY**

*Axiom 14: in the editing process, you refine your copy to express exactly what you want to express with the fewest words.*

*Axiom 15: selling a cure is a lot easier than selling a preventive, unless the preventive is perceived as a cure or the curative aspects of the preventive are emphasized.*

The psychology should be considered when writing the advertising message. The following are the 31 psychological triggers to remember or review when you are writing your ad copy:

1. Feeling of involvement or ownership;
2. Honesty;
3. Integrity;
4. Credibility;



5. Value and proof of value;
6. Justify the purchase;
7. Greed;
8. Establish authority;
9. Satisfaction conviction;
10. Nature of product;
11. Nature of prospect;
13. Current fads;
13. Timing;
14. Linking;
15. Consistency;
16. Harmonize;
17. Desire to belong;
18. Desire to Collect;
19. Curiosity;
20. Sense of urgency;
21. Fear;
22. Instant gratification;
23. Exclusivity, rarity or uniqueness;
24. Simplicity;
25. Human Relationships;
26. Storytelling;
27. Mental Engagement;
28. Guilt;
29. Specificity;
30. Familiarity;
31. Hope;

Best-selling books are written for the 8th to 10th grade level. The following are the steps for determining grade level of the copy you write:

1. Take a sample of your copy – start with 100 to 125 words from the very beginning of the ad.
2. Count the words in each sentence. Dates and numbers equal one word and independent clauses count as separate sentences (e.g., “*We studied, and we learned*” would be two sentences).

3. Divide the total number of words by the number of sentences to get the average sentence length.

4. Count the number of long words (those of three or more syllables), but:

– do not count short-word combinations like pawnbroker or yellowtail.

– do not include proper names.

– do not include verbs that have become three syllables by adding *-ed* or *-es*.

5. Divide the number of long words by the total number of words in the selection to get the percentage of long words.

6. Add the average sentence length to the percentage of long words.

7. Multiply this total by 0.4 to get the grade level.

Clarity is one of the most important factors in writing copy, and the grade level gives you an insight into how important short sentences and simple words can be. But do not be obsessed with achieving low grade level scores at the expense of common sense. You need to vary the length of sentences and use three-syllable words when you need them and realize that every audience is different.

There are seven steps to writing a clear copy, they are:

Step 1: become an expert on the product or service you are planning to sell. Discover everything you could about a product or service from all kinds of sources.

Step 2: know your prospect. You might become an expert on your product or service but if you do not know your customer, you are at a big disadvantage. What will motivate your prospect to become a customer? Who is your typical prospect? This will give you insight and also spur lots of good ideas.

Step 3: write your headline and subheadline. They must grab the reader and create enough curiosity to cause the reader to get to the first sentence. The best ones are short. “*Vision Breakthrough*,” “*Pocket CB*,” “*Pocket Yellow Pages*” – all are concise and yet create enough curiosity to get you to the subheadline. Subheadlines should

be around 16 words, and first sentences should be as short as possible.

Step 4: write the copy. Do not worry about sentence structure, grammar, punctuation – just start writing and keep writing.

Let all your ideas and thoughts flow into the computer and above all do not worry about making mistakes. The main idea is to take everything that is in your head about the subject and dump it into your hard drive. If you write your copy by hand or with a typewriter, write with the same abandon. However, you are a lot better off writing your ad with a computer, especially when it comes to editing.

Step 5: edit your copy. Go through it and correct the spelling, grammar, punctuation and sentence structure. Eliminate extra words not necessary to express your thoughts. Tighten things up.

Step 6: incubate. Stop editing, put the text aside and take a walk or do something pleasurable. You will be amazed at what getting away from your text will do for you. If you can come back to it the next day, better yet. The more time between what you have just edited and taking the next step – all the better.

Step 7: take a final look at your copy. You will be amazed at how much more you are going to catch and how much more refined you will be able to make the copy with this final look. Of course you could easily repeat steps 5 and 6 and continue to edit until you are happy with your results.

These seven steps summarize what you as a copywriter will go through to write copy.

## **REFERENCES**

1. Albrighton, Tom. The ABC of Copywriting [Electronic resource]. – Access mode : <https://www.amazon.co.uk/Tom-Albrighton/e/B07D8T8XMZ>.

2. Maslen A. The Copywriting sourcebook. How to write better copy, faster – for everything from ads to websites [Electronic resource]. – London : Marshall Cavendish Business, 2010. – 223 p.

Access mode : <https://www.amazon.com/Copywriting-Sourcebook-better-everything-websites/dp/0462099741>.

3. Ng, Edmond. Copywriting Basics for Beginners. [Electronic resource]. – Access mode : <https://ezinearticles.com/?Copywriting-Basics-for-Beginners&id=5258540>.

### **KEY TERMS:**

Refine, cure, preventive, psychological triggers, integrity, credibility, timing, grade level, urgency, simplicity, final look.

### **SELF-ASSESSMENT QUESTIONS:**

1. What is the contradiction between selling a cure and a preventive?
2. What psychological triggers do we need to remember or review when we are writing our ad copy?
3. What is the grade level of the best-selling books?
4. How to determine a grade level of the copy you write?
5. What gives you an insight into how important short sentences and simple words can be?
6. Describe seven steps to writing a clear copy.
7. What is a recommended number of words in a subheadline?
8. Do you need to worry about the sentence structure, grammar, and punctuation when you start writing a copy?
9. Is it necessary to repeat some of the seven steps to writing a clear copy?
10. Should a good copywriter follow all seven steps to writing a clear copy?

Навчальне видання

**Прокопенко** Антоніна Вадимівна,  
**Гарцунова** Лариса Іванівна

# **ОСНОВИ ІНФОРМАТИКИ ТА ПРИКЛАДНОЇ ЛІНГВІСТИКИ ЗІ ЗМІСТОВИМ МОДУЛЕМ: КОПРАЙТИНГ**

Конспект лекцій

для студентів спеціальності 035 «Філологія»  
денної та заочної форм навчання

Відповідальний за випуск І. К. Кобякова  
Редактор І. О. Кругляк  
Комп'ютерне верстання: Л. І. Гарцунова, А. В. Прокопенко

Підписано до друку 09.06.2020, поз. 91.  
Формат 60×84/16. Ум. друк. арк. 6,51. Обл.-вид. арк. 5,98. Тираж 5 пр. Зам. №

Видавець і виготовлювач  
Сумський державний університет,  
вул. Римського-Корсакова, 2, м. Суми, 40007  
Свідоцтво суб'єкта видавничої справи ДК № 3062 від 17.12.2007.