

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Уманський державний педагогічний університет імені Павла Тичини

Навчальний посібник
ПРИКЛАДНА ЛІНГВІСТИКА

Умань – 2019

УДК 811.161.2'1 (075.8)

П 75

*Затверджено до друку науково-методичною комісією
факультету іноземної філології, Уманського державного педагогічного університету
імені Павла Тичини
(протокол № 3 від 31. 11. 2019)*

Рецензенти:

Гриценко П.Ю., доктор філол. наук., професор, Інститут української мови НАН України

Бацевич Ф.С., доктор філол. наук., професор, Львівський національний університет імені Івана Франка.

П 75 Прикладна лінгвістика: навчальний посібник/ МОН України, Уманський держ. пед. ун.-т. імені Павла Тичини/; упорядник: Н. А. Цимбал. – Умань: Візаві, 2019. – 106 с

У навчальному посібнику розглядаються основні напрями сучасної прикладної лінгвістики, а саме корпусна, комп'ютерна лінгвістики, перекладознавство, термінознавство та нейролінгвістичне програмування як прикладні лінгвістичні напрями.

Для студентів, магістрантів, аспірантів, викладачів.

УДК 811.161.2'1 (075.8)

ЗМІСТ

Вступ	4
1. Лекція № 1 Прикладна лінгвістика як мовознавча дисципліна	5
2. Лекція № 2. Комп'ютерна лінгвістика	22
3. Лекція № 3. Перекладознавство	37
4. Лекція № 4. Корпусна лінгвістика	49
5. Лекція № 5. Термінознавство	58
6. Лекція № 6. Нейролінгвістичне програмування	66
Рекомендована література	75
Ресурси	77
Додатки	79

ВСТУП

Прикладна лінгвістика — розділ мовознавства, пов'язаний із практичним розв'язанням питань, які пов'язані з вивченням мови.

До традиційних проблем прикладної лінгвістики належать укладання словників, розробка алфавітів і систем письма, транскрипції усного мовлення та транслітерації іншомовних слів, лінгвістичне обґрунтування викладання рідної та іноземних мов, переклад з однієї мови на іншу, стандартизація й уніфікація науково-технічної термінології, укладання спеціальних лінгвістичних довідників, створення штучних мов, удосконалення орфографії і пунктуації, мовна культура тощо.

До нових проблем прикладної лінгвістики належать автоматичний (машинний) переклад, створення інформаційних мов, автоматичне анотування та індексування документів, лінгвістичне забезпечення роботи інформаційних систем, автоматична переробка текстової інформації, лінгвістичне забезпечення автоматичних систем управління (АСУ), автоматичний аналіз (розпізнавання) й автоматичний синтез тексту тощо. Прикладна лінгвістика тісно пов'язана з математикою, кібернетикою та інформатикою.

Найважливіші проблеми прикладного мовознавства — машинний переклад та інформаційна служба.

Однією з актуальних проблем прикладної лінгвістики є проблема інформаційної служби, точніше проблема лінгвістичного забезпечення інформаційно-пошукових систем машинного типу, метою яких є зберігання, пошук і видача інформації. З проблемою пошуку інформації пов'язані питання компресії тексту й автоматичного реферування. Тут велике значення має статистика. «Читаючи» текст, машина відбирає найчастотніші слова і, комбінуючи їх за певними правилами, складає анотацію.

Незважаючи на нерозв'язані питання і недосконалість сучасних програм, машина вже зараз полегшує працю лінгвіста. Вона зберігає в пам'яті текстові ілюстрації значень слів, може розташувати слова за алфавітом, дати оточення кожного слова, підрахувати частоту вживання слова тощо.

В Україні створено Український мовно-інформаційний фонд Національної академії наук, завданням якого є організація автоматизованої системи для укладання одномовних українських та двомовних (перекладних українсько-іншомовних й іншомовно-українських) словників. У цьому фонді готується фундаментальна академічна електронна лексична картотека, яку невдовзі зможуть використовувати мовознавці для своїх теоретичних досліджень.

Лекція № 1 ПРИКЛАДНА ЛІНГВІСТИКА ЯК МОВОЗНАВЧА ДИСЦИПЛІНА

1. Прикладна лінгвістика серед дисциплін мовознавчого циклу.
2. Напрями прикладної лінгвістики.

Сучасна система лінгвістичних наук є нерівнорядною, охоплюючи класичну теоретичну лінгвістику, де основні

напрями вивчення пов'язані з розглядом фонології, морфології, синтаксису, семантики лексичної, до складу яких сучасні науково-лінгвістичні парадигми додали когнітивну лінгвістику, функційну лінгвістику і генеративну лінгвістику; дескриптивну лінгвістику, що містить антропологічну лінгвістику, еволюційну лінгвістику (порівняльно-історичне мовознавство, етимологія), фонетику і соціолінгвістику, а прикладна лінгвістика – це комп'ютерна лінгвістика, математична лінгвістика, судова лінгвістика, юридична лінгвістика, методики викладання і вивчення мови (Language assessment / Language development), прескриптивна лінгвістика, нейролінгвістика, психолінгвістика, стилістика тощо. Для прикладної лінгвістики, що безпосередньо зорієнтована і пов'язана з вивченням мови, властиві такі напрями: лексикографія, лінгводидактика, термінознавство, перекладознавство; а основними напрямками прикладної лінгвістики (англ. applied linguistics), пов'язаної з практичним застосуванням, постають: 1) комп'ютерна лінгвістика (англ. computational / computerized linguistics) з її основними завданнями машинного перекладу, автоматичного розпізнавання символів (англ. OCR), автоматичне розпізнавання мовлення (англ. ASR), автоматичне віднаходження даних (англ. Data Mining), автоматичне реферування, побудова систем управління знаннями, створення електронних словників, тезаурусів, онтологій, корпусна лінгвістика; 2) лінгвістична експертиза (наприклад, у судовій практиці); 3) наука про упорядкування і стандартизацію науково-технічної термінології.

У мовознавстві склалося два розуміння прикладної лінгвістики – широке і вузьке. Для першого прикладна лінгвістика – це будь-яке застосування наукових знань про мову в нелінгвістичних науках. У західноєвропейській традиції аналоги цього терміна (англ. applied linguistics, нім. angewandte

Linguistik) застосовувані насамперед на позначення теорії і практики викладання іноземних мов, охоплюючи методику, особливості опису граматики для навчальних потреб тощо. Так, інститут у Дубліні з назвою School of Applied Languages (Школа прикладної лінгвістики) є за своєю суттю інститутом іноземних мов. У вітчизняних студіях від часу появи терміна (п'ятдесяті роки ХХ століття) цей термін був зорієнтований на автоматичне опрацювання текстової інформації (машинного перекладу, автоматичного реферування і под.), чим і мотивоване використання в цьому розрізі двох термінів прикладна лінгвістика і комп'ютерна лінгвістика, інколи автоматична лінгвістика, обчислювальна лінгвістика, інженерна лінгвістика, що не можна вважати мотивованим, тому що кожна з цих дисциплін у рамках прикладної лінгвістики має свої методи, завдання, предмет та об'єкт роботи.

Найприйнятнішим для прикладної лінгвістики, що безпосередньо зорієнтована і пов'язана з вивченням мови, у вишівській практиці як спеціальності є напрям комп'ютерної лінгвістики (за своєю суттю до нього прилягають: математична лінгвістика, обчислювальна лінгвістика (англ. computational linguistics) – зорієнтований на штучний інтелект, що передбачає або ставить основним завданням використання математичних моделей для опису природних мов. У цьому розрізі комп'ютерна лінгвістика перетинається з опрацюванням природних мов, але в останній час у центрі уваги перебувають прикладні методи опису та опрацювання природної мови для комп'ютерних систем, а для комп'ютерної лінгвістики основним є утворення і використання абстрактних моделей. У цьому розрізі основне поле діяльності – це опрацювання алгоритмів і прикладних програм для опрацювання мовної інформації (цьому завданню відповідає «лінгвіст-програміст»).

Прикладні проблеми поставали перед мовознавством від самого початку його зародження. Ці проблеми спрямовані на оптимізацію функцій мови – від формування і підтримування традиції читання та розуміння сакральних текстів (екзегетика і герменевтика) до забезпечення й активізації міжнародних контактів. У теоретико-філософських та лінгво-теоретичних пошуках здебільшого розрізняють комунікативну, епістемічну (гносеологічну) та когнітивну функції мови, кожна з яких об'єднує частковіші функції. Так, комунікативна функція охоплює фатичну (контактновстановлювальну) та інформаційну (у вузькому сенсі) функції, функції впливу, соціальну функцію (виявляється у мовній політиці). В епістемічному (гносеологічному) плані мовна система постає як спосіб збереження та передавання знань, як вияв специфічно особливого національного погляду на світ. Когнітивна функція належить до тих сфер життя мови, що пов'язані з мисленням та пізнанням людиною дійсності.

У функційному аспекті прикладна лінгвістика постає як академічна дисципліна, що охоплює вивчення і дослідження способів оптимізації різних сфер функціонування мовної системи. Мовні функції є опертям класифікації нерівнорядних сфер застосування лінгвістичних знань. На оптимізацію комунікативної функції зорієнтовані насамперед такі наукові дисципліни: теорія перекладу, машинний переклад, теорія і практика викладання рідної і нерідної мови, теорія і практика інформаційно-пошукових систем, створення інформаційних (ширше – штучних) мов, теорія кодування. З оптимізацією соціальної функції мови (частина комунікативної) співвіднесені дослідження із соціолінгвістики (зокрема, вивчення та обґрунтування мовної політики), орфографії та орфоєпії, теорії впливу, політичної лінгвістики. На оптимізацію епістемічної

(гносеологічної) функції зорієнтовані лексикографія (зокрема комп'ютерна), дослідження з термінології і термінографії, корпусна і польова лінгвістика. Оптимізація когнітивної функції мови прямо корелює зі студіями з комп'ютерної лінгвістики, психолінгвістики й афазіології, квантитативної лінгвістики, «лінгвістичної кримінології».

Прикладну лінгвістику можна розглядати у сфері засобів оптимізації функціонування мови як засобу передавання інформації. Тут на перше місце постають перекладознавчі виміри. У межах комп'ютерної лінгвістики (для зручності слід користуватися поняттям прикладна комп'ютерна лінгвістика) особливу вагу має машинний переклад, що є процесом перекладу текстів (письмових, а в загальноцивілізаційному вимірі також і усних) однієї природної мови іншою за допомогою спеціальної комп'ютерної програми, тому таким важливим є напрацювання навичок створення алгоритмів перекладу та побудови подібних систем. Наука про переклад (англ. translation studies; нім. *Übersetzungswissenschaft* Translationswissenschaft) охоплює кілька важливих напрямів: теорія перекладу, аналіз перекладу, методика навчання перекладу. Особливий статус належить машинному перекладу – наукова і технологічна дисципліна, пов'язана з наукою про переклад, з комп'ютерною лінгвістикою. Наука про переклад – міждисциплінарна, значною мірою активізована зв'язками з літературознавством, когнітивною наукою і культурною антропологією. За активністю перекладацької діяльності розмежовують усний і письмовий переклади, різниця між якими постає особливою в психологічному аспекті. Прикладний аспект теорії перекладу пов'язаний насамперед з тим, що необхідно послідовно враховувати особливості невідповідності між граматичними і лексичними системами різних мов, різницю в

прагматичному функціонуванні мовної системи. Сюди можна віднести труднощі в категоризації дійсності, зумовлені граматичними особливостями різних мов (наприклад, розбіжності в граматичних категоріях тощо), особливості лексичного значення і под. Найважливішою лінгвістичною проблемою науки про переклад є адекватне відтворення так званих «прихованих категорій» – категорії невідчужуваної належності, означеності-неозначеності тощо. Виступаючи в одних мовах граматичними, в інших – набуваючи нерегулярного вираження, вони складають надзвичайну трудність у перекладі. Машинний переклад своїми витокami та постановням мотивований практичними потребами, що особливо актуальними постали в п'ятдесяті роки двадцятого століття, коли зросли обсяги і ємність науково-технічної інформації.

У 1949 році фахівець з дешифрування Уоррен Уївер (США) склала меморандум з обґрунтуванням принципової можливості створення систем машинного перекладу. Ця ідея ґрунтувалася на тому, що структурну подібність між мовами можна описати формально, що є необхідною умовою опрацювання алгоритмів для ЕОМ. Спочатку прогнозувалось, що системи машинного перекладу можуть легко забезпечувати переклад будь-яких науково-технічних текстів. Це сприятиме інтенсифікації перекладних студій, значною мірою вивільнивши процедуру перекладу від людських ресурсів. Уже перші спроби створення систем машинного перекладу засвідчили, що в цьому разі необхідні суттєві і нетривіальні інформації різного типу про функціонування мовних систем, їхнє облаштування, закономірності внутрішньої будови, що не була адекватно висвітлена в традиційно-класичних граматичних описах. Водночас комп'ютерне моделювання людської здібності

295перекладу вимагало врахування і суто лінгвістичних, і психологічних, і соціальних, і гендерних та інших чинників. Перші спроби опрацювання програм автоматичного перекладу дали незадовільний результат: переклад ЕОМ вимагав прискіпливого і глибокого редагування, а вартість машинного перекладу часто була вищою від оплати звичайного перекладача. Сьогодні розвиток і напрацювання систем машинного перекладу має кілька напрямів: 1) опрацювання людино-машинних систем, що враховує і прогнозує пряму участь людини в процесі перекладу на різних етапах; 2) проблемна сфера обмежується конкретною підмовою, або підстилем, наприклад текстами космічної галузі або будівельної індустрії; 3) системи машинного перекладу розглядаються як різновид підсистем / систем штучного інтелекту (останнє уможливорює використання комп'ютерних технологій, опрацьовуваних у комп'ютерному моделюванні людського мислення).

Набуті в теоретичній лінгвістиці знання про структуру і систему, особливості функціонування мови знаходять застосування в теорії і методиці викладання мови. Напрацювання в теорії, зміна теоретичних передумов ставали опертям динаміки методичних підходів у викладанні мови – рідної і нерідної. «Грамаічно-перекладний» метод мав опертям уявлення про ідеальність будови грамаічної системи мови, що підтверджувалося студіюванням такого зразка логічності і несуперечності – грамаічного ладу латинської мови. Тому вивчення мови мало постулатом засвоєння певних грамаічних правил з обов'язковим засвоєнням усіх винятків дії певного грамаічного правила. Методика ґрунтувалася на визнанні двомовності навчання: усі завдання та пояснення викладаються і пишуться рідною мовою, а визначальною метою є читання

текстів та їхній переклад рідною мовою. Наприкінці дев'ятнадцятого століття цей метод (граматично-перекладний) піддають сумніву в силу того, що він не забезпечував швидке оволодіння мовою, на початку двадцятого століття виник «прямий метод» навчання, що полягав в опрацюванні комплексу вправ, орієнтованих на моделювання оволодіння дитиною мови. Граматична система в цьому випадку вторинна, а оволодіння мовою первинне. Особливістю цього методу є одномовність, його використання передбачає опрацювання досконалої системи способів пояснення значення слова через синоніми, антоніми, дефініцію, опис, рисунок, картинку тощо. Прямий переклад у такому методі не застосовується. Активне використання в прямому методі технічних засобів навчання зумовило створення «аудіолінгвального» і «аудіовізуального» методів навчання іноземним мовам. Активізація прагмалінгвістичних досліджень зумовила виникнення «комунікативно-орієнтованого підходу», основу якого складає навчання спілкуванню іноземною мовою, максимальне наближення до того, як користуються певною мовою її носії. Це зумовило послідовне використання понять комунікативної ситуації і комунікативного наміру як визначальних, внаслідок чого мовні форми розглядають у зв'язку з конкретною ситуацією і намірами мовця. До того ж постулат наближення до реальності вимагає використання не тільки мовних висловів, але й застосування інших знакових систем, притаманних для відповідної ситуації спілкування, – дорожніх знаків, піктограм, розкладу руху потягів, плану міст, розташування вулиць і майданів тощо.

Вивчення мови перебуває в активній взаємодії з культурою народу (пор. завантаження навчальних дисциплін «Лінгвокультурологія», «Лінгвокраїнознавство» тощо). Усе це мотивувало кваліфікацію

такої концепції як «інтегроване країнознавство». Опертям прикладної комп'ютерної (комп'ютерно-прикладної) лінгвістики є опрацювання математичного апарату, методології та методики інформатики з розрізненням їхніх аспектів та природна мова – рідна або іноземна, що уможлиблює застосування тих чи інших алгоритмів на практиці. Вивчення мови як основного фаху дозволяє також належним чином напрацьовувати програми аналізу і синтезу природного мовлення, інтелектуальний аналіз даних (англ. Data Mining) – вияв прихованих закономірностей або взаємозв'язків між змінними у великих масивах неопрацьованих даних. У цьому вимірі вони диференційовані на задачі: класифікації, моделювання, прогнозування, структурування і под. Англomовний термін Data Mining перекладають то як добування даних, то як віднаходження даних, то як віднайдення даних, то як інформаційний прохід, то як витягнення даних/інформації, що зумовлює активне його використання мовою-оригіналом. Значного поширення набув термін «інтелектуальний аналіз даних» (ІАД), що вважається найбільш прямим відповідником. ІАД охоплює методи і моделі статистичного аналізу, машинного навчання, дистанціюючись від останніх в напрямі автоматичного аналізу даних. Інструментарій ІАД уможлиблює проведення аналізу даних фахівцями (аналітиками), які не оперують відповідними математичними знаннями.

Створення електронних словників також належить до актуальних завдань комп'ютерно-прикладної лінгвістики, де електронний словник витлумачувано як відповідну комп'ютерну базу даних, що містить особливим способом закодовані словникові статті. Це дозволяє зреалізувати швидкий пошук необхідних слів, здебільшого з урахуванням морфологічних

форм та з можливістю пошуку сполук слів (прикладів вжитку), а також з можливістю зміни напрямку перекладу (наприклад, англо-український або українсько-англійський, російсько-український або українсько-російський і под.), що дозволяє поповнювати наявні на сьогодні словники і програми.

Оптимізація епістемічної (гносеологічної) мовної функції прямо пов'язана зі створенням словників та їхнім функційним призначенням. Центром уваги будь-якої лексикографії загалом, а прикладово-комп'ютерної зокрема є способи організації словникової статті, структурування словника і технологія його створення. Створення словників безпосередньо корелює з особливостями мовно-державної політики, постає однією зі сфер прикладної лінгвістики. Сучасна лексикографія суттєво поліпшила інструментарій комп'ютерними технологіями створення та експлуатації словників. Взаємодія власне-словникового підходу з комп'ютерними технологіями постала результатом створення напрямку прикладної лінгвістики – комп'ютерної лексикографії, де активно застосовними є спеціальні програми – бази даних, комп'ютерні картотеки, програми опрацювання текстів. Такі програми уможливають в автоматичному режимі формування словникових статей, збереження словникової інформації та 298 опрацювання її. Сукупність комп'ютерних лексикографічних програм містить два компоненти: 1) програми підтримки лексикографічних праць; 2) автоматичні словники різних типів, у тому числі і лексикографічні бази даних.

З оптимізацією епістемічної функції пов'язана також і термінографічна робота, предметом якої є термінологія, що співвідноситься в широкому вимірі з усіма термінами природної мови, а у вузькому – з термінами певної конкретної наукової дисципліни або спеціальної сфери практичної діяльності

людини. У межах тієї чи іншої наукової сфери, окремої наукової теорії терміни утворюють терміносистему.

Терміни – це слова (словосполучення) метамови науки і застосувань наукових дисциплін, та слова на позначення специфічних реалій конкретних сфер людської практичної діяльності. Поняття терміна здебільшого визначається через реалізовані в терміносистемі його властивості. Використання термінів є не інтуїтивним, а мотивоване відповідними дефініціями. Результат термінографічної роботи є термінологічні словники, що наближають цей напрям прикладної лінгвістики до лексикографії. Про опрацювання природної мови (Natural Language Processing, NLP) як один із загальновизнаних напрямів штучного інтелекту і математичної лінгвістики слід окремо говорити в розрізі комп'ютерно-прикладної лінгвістики, тому що цей різновид спрямований на вивчення проблеми комп'ютерного аналізу (у штучному інтелекті – розуміння мови) і синтезу (у штучному інтелекті – генерація грамотного тексту) природних мов. Активізація цього напрямку сприяє оптимальному вирішенню проблеми взаємодії й активного діалогу комп'ютера і людини. Найбільшою проблемою і водночас магістральною виступає сприйняття комп'ютером слова, його розпізнавання, а поряд з нею – розуміння слова з усім розмаїттям його смислових площин, розпізнавання неологізмів, okazіоналізмів з диференціюванням і розпізнаванням таких похідних від звичайних технічних, коректорських та інших помилок, розмежування омонімів (особливо омофонів тощо).

Прикладний напрям з опрацювання природної мови (переклад англійського терміна Natural Language Processing) виник наприкінці 1960-х років і розвивався в рамках науково-

технічної дисципліни, що називається штучний інтелект. За своєю внутрішньою формою словосполучення «опрацювання природної мови» повинно охоплювати усі сфери, де комп'ютери використовуються для опрацювання мовних даних.

Реальним постанало вузьке закріплення і мотивоване використання терміна – опрацювання методів, технологій і конкретних систем, що забезпечують спілкування людини з ЕОМ природною або обмеженою природною мовою. Особливо активно розвивався цей напрям в 1970-ті роки, що розглядався в прикладній лінгвістиці як частина проблематики комп'ютерної лінгвістики. Це мотивувалось стрімким збільшенням кількості користувачів ЕОМ. У цьому разі постала проблема: якщо не можна і недоцільно навчати і навчити усіх користувачів ЕОМ відповідним мовам і технологіям програмування, то необхідні напрацювання взаємодії граничних користувачів з комп'ютерними програмами. Розв'язання цієї комунікативної проблеми передбачало або мало два основних шляхи: 1) постала спроба адаптації мов програмування та операційних систем до кінцевого користувача (мови низького рівня типу Ассемблера було замінено мовами вищого рівня типу Visual Basic, відбувався розвиток операційних систем у напрямі мобільнішого зображення та наочності; для створення комунікативного комфорту та мобільнішого розуміння в операційних системах почали використовувати метафори, що ґрунтуються на зрозумілих людиною поняттях – наприклад, робочий стіл, бібліотека, картотека тощо.

Інший напрям – опрацювання систем, що уможлилювали б взаємодію з ЕОМ у конкретній сфері природною мовою або відповідному її обмеженому варіанті. У сфері опрацювання природної мови

активними поставали встановлення особливостей спілкування, закономірностей структурування діалогу, типів і виявів комунікації в гуманітарних науках. Основна відмінність комп'ютерного моделювання комунікації в системах взаємодії з ЕОМ полягає в тому, що за обмеження проблемної сфери 300 розробникам доводиться враховувати усі аспекти, всі рівні реального спілкування, що складає досить вагому проблему теоретичного і прикладного аспектів.

Архітектура систем опрацювання природної мови в загальному вимірі охоплює блок аналізу мовленнєвого спілкування користувача, блок інтерпретації повідомлення, блок породження смислу відповіді і блок синтезу поверхневої структури висловлення. Особливу частину системи охоплює діалогічний компонент, у якому фіксовані стратегії ведення діалогу, його модифікації, умови застосування цих стратегій, способи подолання можливих комунікативних невдач. У цьому розрізі здебільшого вирізняють системи питання-відповідь (пор. систему ПОЕТ, створену колективом дослідників, очолювану Е.В. Поповим), діалогічні системи вирішення завдань (наприклад, система SNUKA, що вирішує в інтерактивному режимі завдання планування військових операцій), системи опрацювання зв'язних текстів (такими є різноманітні системи опрацювання патентних документів, що підтримують і розуміння тексту, і відповіді на питання). В сучасну епоху особливо актуальним і значущим є уміння правильно організації редагування з окресленням тих чи інших аспектів і вимірів цієї проблеми, зокрема автореферування. Основним завданням в цьому розрізі постає напрацювання вільного осмислення інформації та її структурування з послідовним відбором найбільш значущої та актуальної, що може досягатися розробкою різноманітних

методик стиснення текстів, простеження базових структурних компонентів, наростання їхньої спектральної заангажованості.

Управління знаннями (англ. knowledge management) – це особлива методологія в комп'ютерно-прикладній лінгвістиці, спрямована на підвищення рівня конкурентоспроможності і захисту компанії за рахунок максимального й оптимального використання повного інструментарію охорони, управління та економії матеріальних активів компанії. До цього слід віднести стратегії, спрямовані на надання вчасно необхідного обсягу і 301ємності знань тим членам спільноти, яким ці знання (очевидні та неочевидні – типу ноу-хау різні, секрети майстерності, інтуїція тощо) необхідні для підвищення ефективності дієвості співтовариства (від 2000 року курси з управління знаннями наявні на кафедрах менеджменту, їхня змонтованість у комп'ютерно-прикладну лінгвістику очевидна і незаперечна). У комп'ютерно-прикладній лінгвістиці реальним є розгляд особливостей типології спіральності знань (Ікуджіо Нонака), внутрішньої структурованості і зовнішньої структурованості знань, створення мап (карт) знань і вивчення особливостей корпоративної культури. Управління знаннями як напрям комп'ютерно-прикладних студіювань постав унаслідок активної дії трьох причин: 1) соціальної (інтелект, пам'ять, досвід, ініціатива тощо постають безпосередніми учасниками виробництва); 2) економічна (статус інформаційних технологій, рівень їхнього впровадження та запитуваність останніх працівниками, співвідношення таких технологій з оперативним інтелектуальним фондом, формування на цьому тлі прибутку); 3) технологічна (еволюційний процес розвитку інформаційних технологій – комунікація, підтримка мисленнєвої діяльності та віднімання). Усе це мотивує зв'язок теорії управління знаннями з теорією інтелектуального капіталу щодо поняттєвого поля

організаційного, соціального (або структурного), і насамперед – людського [капіталу].

Як складник комп'ютерно-прикладної лінгвістики корпусна лінгвістика спрямована на оптимізацію епістемічної (гносеологічної) функції мови і своїм виміром проблемативна опрацюванням, створенням та використанням текстових (лінгвістичних) корпусів (термін з'явився в 60-ті роки ХХ століття (перший комп'ютерний корпус – Браунівський в Університеті Брауна з 500 фрагментів текстів по дві тисячі слів у кожному з ємністю один мільйон слововживань; модель цього корпусу була використана при створенні корпусів інших національних мов (російської у 1980-ті роки в Університеті Уппсали (Швеція), активізація ж корпусно-лінгвістичних досліджень постала у вісімдесяті роки двадцятого століття у зв'язку з розвитком обчислювальної техніки). Для комп'ютерно-прикладної лінгвістики характерним є опрацювання загальних принципів побудови лінгвістичних корпусів даних з використанням сучасних комп'ютерних технологій, розуміння лінгвістичного корпусу як сукупності текстів, зібраних відповідно до визначених принципів, проіндексованих за визначеним стандартом і забезпечених спеціальною пошуковою системою. Крім власне-конструювання корпусів даних, корпусна лінгвістика створює комп'ютерний інструментарій для опрацювання текстів. Частковим виявом корпусу постає корпус текстів. Одиницями останнього виступають тексти або їхні достатні фрагменти. Ілюстративні корпуси створюються після проведення наукового дослідження, де метою постає не вияв нових фактів, а підтвердження отриманих результатів, їхнє обґрунтування. Це не ідеальні корпуси національної мови, навіть окремого функційного стилю, а тільки вияв окремого проблемного явища. У корпусній лінгвістиці значущим постає

розмежування динамічних (моніторних) і статичних корпусів. Спочатку корпуси текстів створювалися як статичні утворення, що виявляють відповідний часовий зріз мовної системи. Типовим виявом такого різновиду корпусів є авторські корпуси – колекції текстів письменників, наприклад, Корпус художніх текстів Павла Загребельного (кафедра української мови та прикладної лінгвістики Донецького національного університету), Корпус художніх текстів Ф.М. Достоєвського (Інститут російської мови РАН). Значна частина суто лінгвістичних та екстралінгвістичних завдань вимагає вияв і простеження історичних, еволюційно-динамічних процесів та змін у функціонуванні мовних явищ, пор., наприклад, зміни значення слів, частотність використання тих чи інших синтаксичних конструкцій, видозміни в їхньому структуруванні, співвідношення первинних і вторинних виявів функцій морфологічних форм тощо. На реалізацію таких завдань і спрямована опрацьована спеціальна програмна технологія побудови та експлуатації динамічного корпусу текстів. Такі корпуси кваліфікують як монітарні, особливість монтування таких корпусів полягає в тому, що вони не передбачають назавжди усталеного набору текстів. Протягом завчасно зфіксованого часового проміжку наявне оновлення і/або доповнення множинності текстового корпусу. До таких монітарних корпусів належать, насамперед, Бірмінгемський корпус, створений завдяки творчій ініціативі та керівництву Дж. Сінклера. За концептуальним баченням користувача, такі корпуси повинні бути репрезентативні (представницькі), повні та економічні. Інколи диференціюють корпуси, вирізняючи корпуси першого порядку – просте зібрання текстів, об'єднаних якою-небудь спільною ознакою (мовою, жанром, автором, періодом створення текстів, епохою, темою і под.). Доцільність

створення текстових корпусів мотивована: 1) поданням лінгвістичних даних у реальному контексті; 2) досить розмаїтим поданням даних (особливо у великому за обсягом корпусі); 3) можливістю багаторазового використання разово створеного корпусу для розв'язання різних лінгвістичних завдань. Сьогодні репрезентативні корпуси існують (опрацьовуються) для німецької, польської, сербської, словенської, новогрецької, вірменської, чеської, китайської, японської, угорської та інших мов, хоча важливими є дотримання авторських прав, зведення усіх текстів у єдино-цілісну систему і форму, тематичне і жанрове структурування корпусу тощо, що вимагає значної затрати сил та енергії. Поряд з репрезентативними корпусами (охоплення жанрових і функційно-стильових різновидів) наявні й опортуністичні колекції текстів – газетні, новинарні (Рейтер), колекції художньої літератури (Проект Гутенберг або Бібліотека Мошкова) та ін. З-поміж корпусів вирізняються дослідницькі корпуси та ілюстративні корпуси. Перші призначені для розгляду та аналізу особливостей функціонування мовної системи в різних аспектах, а останні – ставлять за мету не тільки і не стільки вияв нових фактів, скільки підтвердження й аргументація уже встановлених або здобутих результатів.

Рекомендована література:

1. Баранов А. Н. Введение в прикладную лингвистику. 3-е изд., стереот. Москва : Эдиториал УРСС, 2007. 360 с.
2. Дарчук Н. П. Комп'ютерна лінгвістика. Київ : ВПЦ «Київський університет», 2008. 351 с.
3. Денисюк В. В. Інформаційні технології в лінгвістиці. Умань : ФОП Жовтий О. О., 2015. 135 с.
4. Загнітко Анатолій Статус прикладної лінгвістики в системі лінгвістичних наук URL <http://litmisto.org.ua/?p=19771>

5. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика : підручник. Донецьк : ТОВ «Юго-Восток, Лтд», 2006. 188 с.
6. Ситар Г. В. Теоретичні та прикладні проблеми перекладу : навч.-метод. посібник. Донецьк : ДонНУ, 2009. 88 с.

Лекція № 2. КОМП'ЮТЕРНА ЛІНГВІСТИКА

1. Комп'ютерна лінгвістика як прикладна лінгвістична дисципліна
2. Деякі напрями комп'ютерної лінгвістики
3. Гіпертекстові технології подання тексту

Комп'ютерна лінгвістика як прикладна лінгвістична дисципліна. Під терміном "комп'ютерна лінгвістика" (computational linguistics) зазвичай розуміється широка область використання комп'ютерних інструментів - програм, комп'ютерних технологій організації та обробки даних - для моделювання функціонування мови в тих чи тих умовах, ситуаціях, проблемних областях, а також сфера застосування комп'ютерних моделей мови не тільки в лінгвістиці, а й у суміжних з нею дисциплінах. Власне, тільки в останньому випадку мова йде про прикладної лінгвістики в строгому сенсі, оскільки комп'ютерне моделювання мови може розглядатися і як сфера застосування теорії програмування (computer science) в галузі лінгвістики. Проте загальна практика така, що сфера комп'ютерної лінгвістики охоплює практично все, що пов'язано з використанням комп'ютерів у мовознавстві: "Термін" комп'ютерна лінгвістика "задає загальну орієнтацію на використання комп'ютерів для вирішення різноманітних наукових і практичних завдань, пов'язаних з мовою, ніяк не обмежуючи способи вирішення цих завдань".

Інституційний аспект комп'ютерної лінгвістики. Як особливу науковий напрямок комп'ютерна лінгвістика оформилася в 60-і рр.. Потік публікацій в цій області дуже великий. Крім тематичних збірок, в США щоквартально виходить журнал "Комп'ютерна лінгвістика". Велику організаційну і наукову роботу проводить Асоціація з комп'ютерної

лінгвістики, яка має регіональні структури по всьому світу (зокрема, європейське відділення). Кожні два роки проходять міжнародні конференції з комп'ютерної лінгвістики - КОЛІНГ. Відповідна проблематика широко представлена також на міжнародних конференціях з штучного інтелекту різних рівнів.

Деякі напрями комп'ютерної лінгвістики. Звернімося до тих галузей комп'ютерної лінгвістики, які безпосередньо пов'язані з оптимізацією когнітивної функції мови. Нижче як приклад розглядаються три сфери комп'ютерного моделювання, в яких використовуються знання про функціонування мовної системи: моделювання спілкування, моделювання структури сюжету і гіпертекстові технології представлення тексту.

Моделювання спілкування. У вузькому сенсі проблематика комп'ютерної лінгвістики часто пов'язується з моделюванням спілкування, зокрема, із забезпеченням спілкування людини з ЕОМ на природній або обмеженою природною мовою. Це відноситься до оптимізації мови як засобу спілкування. Втім, комп'ютерні моделі спілкування часто використовуються для вивчення самого процесу спілкування. Зупинимось докладніше на досвіді створення та використання саме таких моделей.

Вивчення вже нагромадилося досвіду експлуатації комп'ютерних систем, які вимагали забезпечення взаємодії з ЕОМ природною мовою, дозволило дослідникам по-новому поглянути на функції і структуру природної комунікації. У центр уваги потрапили питання, які раніше були на периферії теорії діалогу, дискурс-аналізу та теорії комунікації. Що забезпечує природність спілкування? Які умови зв'язності бесіди? Коли спілкування виявляється успішним? У яких випадках виникають комунікативні невдачі і чи можна їх уникнути? Які стратегії спілкування використовують учасники комунікативної взаємодії при досягненні своїх комунікативних цілей? Це далеко не вичерпний список теоретичних проблем, котрі виникли у зв'язку з функціонуванням комп'ютерних моделей спілкування.

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, як навчальний зразок програми-імітатора, метою якої є не моделювання мислення в точному сенсі, а моделювання мовної поведінки. Програма підтримувала розмову зі співрозмовником у реальному масштабі часу, однак при її розробці були використані обмежені програмістські ресурси, лінгвістичний аналіз і синтез також були зведені до мінімуму. Проте програма функціонувала настільки успішно, що фактично спростувала відомий тест Тьюринга на створення штучного інтелекту. Як відомо, Тьюринг замість софістичованого обговорення філософського питання про те, чи може машина мислити, запропонував ігрову завдання наступного типу. Нехай є три учасники: чоловік Л, жінка В і запитувач С. запитують не знає, хто чоловік, а хто - жінка. Ставлячи питання учасникам гри, С повинен спробувати визначити, хто є чоловіком, а хто - жінкою, при цьому учасник намагається містифікувати запитувача, видаючи йому не помилкову, але спотворену інформацію, а учасник В - навпаки, прагне допомогти С. Зрозуміло, що спілкування відбувається не безпосередньо, а через телетайп або за допомогою записок, віддрукованих на друкарській машинці. Що станеться, якщо в якості А виступатиме система ШІ? Чи буде запитувач помилятися так само часто? Простіший варіант цього тесту зводиться до того, що кілька учасників розмовляють з деяким іншим учасником Х. Проблема побудови штучного інтелекту вирішена, якщо більшість учасників не зможе встановити, з ким вони розмовляють - з людиною або машиною.

Програма "Еліза" була використана групою дослідників на чолі з М. Макгайром для вивчення структури діалогу та особливостей природномовної комунікації. У проводився експерименті з "Елізою" розмовляли протягом години 24 випробовуваних. Спілкування відбувалося за допомогою телетайпу. За час бесіди кожен учасник ввів від 10 до 65 реплік і отримав на них відповіді. Після закінчення 15 учасників (62%) були впевнені, що їм відповів чоловік, 5 піддослідних (21%) виявили певні коливання і лише четверо учасників (17%) були абсолютно впевнені, що спілкувалися з ЕОМ. З лінгвістичної точки зору алгоритми програми "Еліза" включають мінімум лінгвістичної інформації. По-

перше, це комплекс ключових слів, які актуалізації деякі стійкі комунікативні формули (шаблони), по-друге, здатність відносно нескладно трансформувати попереднє висловлювання.

Цікаво, що істотна тематична обмеженість комунікації і значна кількість помилок і неточностей у відповіді (близько 19% неточних або випадають з контексту реплік "Елізи" у згадуваному експерименті М. Макгайра), не завадили випробуванню визнати партнера по комунікації людиною. Справа тут зовсім не в патологічній дурості піддослідних. Це прояв найважливішою особливості комунікації на природній мові: природномовної дискурс дуже терпимо по відношенню до збоїв і помилок - він надмірний і помехоустойчив. Репліки "Елізи", випадали з нормального спілкування, випробовувані легко пояснювали звичайними збоями в розумінні своєї попередньої репліки, не цілком нормальними умовами спілкування, жартівливим настроєм партнера. Стійкість природного дискурсу пояснюється також здібностями людини до інтерпретації мовних дій: людина, що приймає роль учасника діалогу, веде себе відповідним чином. Маючи установку на спілкування, він повинен намагатися включити в комунікацію все те, що за формою нагадує мовленнєвий акт, репліку. Іншими словами, він схильний наділяти сенсом те, що часто сенсу не має. У цьому випадку випробовувані самі породжують сенс діалогу, самі забезпечують його зв'язність, самі приписують партнеру комунікативні інтенції.

Другий важливий висновок експерименту: випробувані досить швидко ухвалювали рішення про те, хто перед ними - комп'ютер або людина. 22 учасника з 24 усвідомили для себе ситуацію не більше, ніж за п'ять обмінів репліками, і далі не змінювали свого рішення. Визначення ролей у комунікації відноситься до метарівні спілкування, оскільки це становить одну з передумов успішної комунікації, що оберігає спілкування від численних комунікативних невдач.

Зрозуміло, що визначення ролей учасників багато в чому визначає вибір стратегії комунікативної поведінки. Дійсно, краще відразу визначити, з ким ми розмовляємо по телефону - з давнім другом або чиновником податкової інспекції. З'ясування того, ким є співрозмовник - машиною або людиною, також

відноситься до метарівні спілкування, і випробувані намагалися встановити рольові характеристики партнера як можна раніше.

Це властивість природномовної комунікації можна назвати принципом пріоритету метакомунікативних параметрів ситуації спілкування.

Третє важливе слідство з експерименту М. Макгайра пов'язано з існуванням різних типів комунікативної взаємодії між людьми. Успішне взаємодія між людиною і програмою типу "Еліза" можливо тільки в ситуації, коли відбувається так зване "асоціативне спілкування", при якому репліки діалогу пов'язані не стільки логічними відносинами типу "причина-наслідок", "посилка-висновок", а асоціаціями. Асоціативне спілкування не має конкретної спрямованості; саме підтримання розмови може служити її виправданням. Співрозмовники не переслідують мети вирішити якусь проблему або виробити єдину точку зору на якесь питання. У класифікації Р. Якобсона для комунікації такого типу запропонований термін "фатіческое спілкування". Зауважимо, що бесіда лікаря-психіатра з пацієнтом по формі також має вигляд фатической спілкування, хоча і переслідує цілком певну мету збору даних про захворювання пацієнта і наступному вербальному і невербальному впливі на його психіку для досягнення лікувального ефекту. "Еліза" не змогла б успішно імітувати спілкування в комунікативній ситуації, названої М. Макгайром "рішення завдань", оскільки вона не здатна зрозуміти проблемну ситуацію, тобто побудувати модель світу дискурсу, визначити альтернативи виходу з проблеми, вибрати одну з альтернатив і т. д. Одна з типових стратегій "відходу від нерозуміння", реалізована в програмі "Еліза" - зміна теми бесіди. Очевидно, що така стратегія ведення бесіди навряд чи приведе до успіху при спільному пошуку вирішення проблеми.

Нарешті, четвертий висновок можна сформулювати як неуніверсальність правил комунікативної взаємодії. Він стосується самих закономірностей спілкування на природній мові. Кожен тип комунікації обслуговується своїм набором відносно простих правил, що забезпечують зв'язність дискурсу, його осмисленість для учасників. Типологія видів спілкування задається відповідними наборами правил. З експериментів М. Макгайра з програмою

"Еліза" впливає, що крім асоціативного (= фатической) способу спілкування, виділяється ще "рішення задач", "задавання питань" і "уточнення розуміння". З лінгвістичної точки зору ці типи, швидше за все, неоднорідні, перетинаються і навіть знаходяться на різних рівнях дискурсу. Так, "уточнення розуміння" відноситься до метарівні комунікації, "задавання питань" може бути частиною стратегії "рішення задач" і "уточнення розуміння" і т.д. Істотно, що комп'ютерний експеримент з програмою, що моделює поведінку учасника комунікації, дозволяє експериментально підтвердити або спростувати багато положень теорії діалогу, розроблені як у лінгвістиці, так і в суміжних дисциплінах - в дискурс-аналізі, теорії комунікації, психології та соціології спілкування.

Моделювання структури сюжету. Вивчення структури сюжету відноситься до проблематики структурного літературознавства (у широкому сенсі), психології творчості та культурології. Наявні комп'ютерні програми моделювання сюжету ґрунтуються на трьох базових формалізму подання сюжету - морфологічному та синтаксичному напрямках подання сюжету, а також на когнітивному підході.

"Морфологія" сюжету. Ідеї про морфологічному пристрої структури сюжету сходять до відомих робіт В.Я. Проппа про російську чарівній казці. Пропп зауважив, що при великій кількості персонажів і подій чарівної казки кількість функцій персонажів обмежено: "Постійними, стійкими елементами казки служать функції дійових осіб, незалежно від того, ким і як вони виконуються. Вони утворюють основні складові частини казки". До числа базових відносяться, наприклад, такі функції:

- відлучення персонажа казки з будинку;
- заборона герою на дію;
- порушення заборони;
- отримання шкідником інформації про жертву;
- обман жертви шкідником;
- мимовільне пособництво жертви вредителю тощо.

Ідеї Проппа лягли в основу комп'ютерної програми TALE, що моделює породження сюжету казки. В основу алгоритму програми TALE покладена послідовність функцій персонажів казки. Фактично функції Проппа задавали безліч типізованих ситуацій, впорядкованих на основі аналізу емпіричного матеріалу. Можливості зчеплення різних ситуацій в правилах породження визначалися типовою послідовністю функцій - у тому вигляді, в якому це вдається встановити з текстів казок. У програмі типові послідовності функцій описувалися як типові сценарії зустрічей персонажів.

Надалі система була ускладнена за рахунок введення моделі світу казки, географія якого складається із звичайного світу, проміжного (середнього) світу та іншого світу. Кожен світ складається з локусів, пов'язаних між собою певними відносинами. Відносини пов'язують не тільки локуси всередині кожного світу, але і локуси різних світів. Звичайний світ складається з наступних локусів: місце проживання героя (локус 1), місце отримання завдання (локус Г), місце дарування чарівних предметів, що допомагають виконати завдання. Перший локус і локус штрих часто збігаються (пор. казки про падчерки і злий Мачусі). До звичайного світу відносяться також локуси 3 (їх може бути багато), в яких долаються перешкоди за допомогою чарівних предметів. Кількість перешкод, як правило, збігається з кількістю чарівних предметів. Після подолання перешкод герой опиняється в проміжному світі, вартовим якого є Баба-Яга. Середній світ відокремлює світ героїв від світу антигероїв. Функції Баби-Яги різняться - вона може виступати як дарувальниця інформації або чергового чарівного засобу, а може виступати на боці антигероїв (наприклад, при акценті на людоджерській поведінці Баби-Яги). Інший світ включає місце проживання антигероя (локус 5), місце битви між героєм і антигероєм (локус 6) і, нарешті, локус 7 - місце нагороди чи мети, якої домагається герой. Локуси пов'язані відносинами переходу, які представляють можливі послідовності розгортання сюжету.

Модифікована версія програми TALE має наступну блок-схему:

Блок-схема програми TALE

Робота програми починається з першого блоку, в якому вибирається тип сюжету казки та її персонажі. Тут же формується експозиція казки (setting). У другому блоці зберігаються описи, пов'язані з персонажами, а в четвертому - постійні характеристики персонажів. Описи даються під фреймоподібних структурах представлення знань. За допомогою другого і третього блоків формуються мотиви і вчинки персонажів. Третій блок задає послідовність руху персонажів за локусами. В останньому (шостому) блоці відбувається збірка породжених фрагментів казки.

Блок-схема модифікованого варіанту програми TALE показує, що чисто "морфологічного" підходу до структури сюжету казки явно недостатньо. "Морфеми" казкового сюжету повинні не тільки певним чином поєднуватися між собою, а й мати специфічні обмеження на сполучуваність. Фіксація одного типового порядку проходження функцій персонажів чарівної казки істотно обмежує наявні можливості сполучуваності. Більш адекватне вирішення цієї проблеми дає синтаксичний підхід до структури сюжету.

"Синтаксис" сюжету. Теоретичну основу синтаксичного підходу до сюжету тексту склали "сюжетні граматики" (story grammars). Сюжетні граматики з'явилися в середині 70-х рр. в результаті переносу ідей генеративної граматики Н. Хомського на опис макроструктури тексту. Якщо найважливішими складовими синтаксичної структури в породжує граматиці були дієслівні та іменні групи, то в більшості сюжетних грамастик в якості базових виділялися експозиція (setting), подія і епізод. У теорії сюжетних грамастик широко обговорювалися умови мінімальності: обмеження, що визначали статус послідовності з елементів сюжету як нормальний сюжет. Виявилось, однак, що суто лінгвістичними методами це зробити неможливо. Багато обмеження носять соціокультурний характер. Сюжетні граматики, істотно розрізняючись набором категорій в дереві породження, допускали вельми обмежений набір правил модифікації наративної структури. У переважній більшості випадків ці правила запозичені з тієї ж граматики. Потенціал варіювання структури сюжету забезпечується в першу чергу трансформаціями пересування та опущення. Наприклад, текст зізнання

злочинця, що фіксує реальну послідовність розгортання подій у злочині, можна за допомогою перестановок і опущень перетворити в детективний сюжет: {злочинець → задум → знаряддя вбивства → місце → вбивство → виявлення трупа → пошуки злочинця} → {виявлення трупа → виявлення знаряддя вбивства → пошуки злочинця}.

Використання сюжетних граматики в комп'ютерному моделюванні виявилось не зовсім вдалим. Синтактична компонент сюжету, описуваний граматами, відображає чисто зовнішні особливості тексту. Не вдається виявити операціональні критерії виділення різних складових сюжету.

Основний висновок дискусії про недоліки сюжетних граматики звівся до необхідності опису сюжету в рамках структури доцільної діяльності, тобто із залученням категорій "мета", "проблема", "план" і т.д. Іншими словами, метамови, що враховує тільки зовнішні особливості сюжету, явно недостатньо. Необхідно звернення до когнітивних станів персонажів.

Когнітивний підхід до сюжету. На початку 80-х рр.. однієї з учениць Р. Шенка - В. Ленерт - у рамках робіт зі створення комп'ютерного генератора сюжетів був запропонований оригінальний формалізм афективних сюжетних одиниць (АСЕ - Affective Plot Units), що виявився потужним засобом представлення структури сюжету. При тому, що він був спочатку розроблений для системи ШІ, цей формалізм використовувався в чисто теоретичних дослідженнях. Сутність підходу Ленерт полягала в тому, що сюжет описувався як послідовна зміна когнітивно-емоційних (афективних) станів персонажів. Тим самим у центрі уваги формалізму Ленерт стоять не зовнішні компоненти сюжету - експозиція, подія, епізод, мораль - а його змістовні характеристики. У цьому відношенні формалізм Ленерт почасти виявляється поверненням до ідей Проппа.

Кожна афективна сюжетна одиниця являє собою бінарне відношення, що зв'язує деякі події, оцінювані персонажами позитивно (+) або негативно (-), і когнітивно-емоційні стани персонажів (у різних комбінаціях - подія & стан; подія & подія і т.д.). Бінарне відношення не є однорідним. Усього виділяється п'ять типів бінарних відносин, специфічних в кожній афективній сюжетній

одиниці. Бінарне відношення може бути мотивацією (позначення - т), актуалізацією (а), припиненням однієї дії іншим (t), еквівалентністю (е), а також афективної каузальною зв'язком між персонажами. Кожна афективна сюжетна одиниця отримує назву, наприклад, УСПІХ, НЕВДАЧА, завзятість, ПРОБЛЕМА і т.д.

Гіпертекстові технології подання тексту. Феномен гіпертексту можна обговорювати з кількох точок зору. З одного боку, це особливий спосіб представлення, організації тексту, з іншого - новий вид тексту, протиставлений за багатьма своїми властивостями звичайного тексту, сформованому в гутенберговської традиції друкарства. І, нарешті, це новий спосіб, інструмент та нова технологія розуміння тексту.

Теоретичні підстави гіпертексту. Багато дослідників розглядають створення гіпертексту як початок нової інформаційної епохи, протиставила ері друкарства. Лінійність листи, зовні відбиває лінійність мови, виявляється фундаментальною категорією, що обмежує мислення людини та розуміння тексту. Світ сенсу нелінійний, тому стиснення змістової інформації в лінійному мовному відрізку вимагає використання спеціальних "комунікативних упаковок" - членування на тему і рему, поділ плану змісту висловлювання на експліцитні (затвердження, пропозиція, фокус) та імпліцитні (пресупозиція, наслідок, імплікатура дискурсу) шари. Відмова від лінійності тексту і в процесі його уявлення читачеві (читання і розуміння), і в процесі синтезу, на думку теоретиків, сприяв би "звільнення" мислення і навіть виникнення його нових форм.

Прототиповий текст - це монолог. Тим часом багато лінгвістів (і серед них М. Бахтін і Л. Якубинский) вказували на вторинність монологу в порівнянні з діалогом. Гіпертекст з цієї точки зору дозволяє усунути штучну монологичність тексту.

Звичайний текст, як правило, має автора. Гіпертекст автора в традиційному розумінні не має - у нього безліч авторів, причому для постійно змінюється гіпертексту авторський колектив також постійно змінюється. Зміна

статусу автора змінює і статус читача: в гіпертекстовій системі розуміння часто супроводжується зміною компонентів гіпертексту або, як мінімум, вибором шляху перегляду, що знову-таки акт творчий, авторський.

Множинність авторства має і ще один наслідок: у гіпертексті представлено багато точок зору на проблему, а в звичайному тексті - тільки одна. Саме тому гіпертекст більш об'єктивний і більше толерантний до читача, ніж класичний текст.

Гіпертекстові технології дозволяють легко поєднувати різні види інформації - звичайний текст, малюнок, графік, таблицю, схему, звук і зображення, що рухається. Як традиційний текст, так і гіпертекст - феномени, породжені новими технологіями. У першому випадку технологія дозволила легко тиражувати і розповсюджувати знання самих різних типів, а в другому - комп'ютерні технології дали можливість змінити сам зовнішній вигляд тексту і його структуру. Різноманітність гіпертексту - це перше технологічне властивість гіпертексту, технологічне в тому сенсі, що воно безпосередньо впливає з використовуваної комп'ютерної технології. Друге технологічне властивість гіпертексту - його нелінійність. Гіпертекст не має стандартної, звичайної послідовності читання. Інші властивості гіпертексту в тій чи іншій мірі є наслідками з цих двох технологічних властивостей.

Підсумувати відмінності тексту і гіпертексту можна наступним чином:

- кінцівку, закінченість традиційного тексту vs. нескінченність, незакінченість, відкритість гіпертексту;
- лінійність тексту vs. нелінійність гіпертексту;
- точне авторство тексту vs. відсутність авторства (у традиційному розумінні) у гіпертексту;
- зняття протиставлення між автором і читачем;
- суб'єктивність, однобічність звичайного тексту vs. об'єктивність, багатосторонність гіпертексту;
- однорідність звичайного тексту vs. неоднорідність гіпертексту.

Компоненти **гіпертексту**. Структурно гіпертекст може бути представлений як граф, у вузлах якого знаходяться традиційні тексти або їх

фрагменти, зображення, таблиці, відеоролики і т.д. Вузли пов'язані різноманітними відносинами, типи яких задаються розробниками програмного забезпечення гіпертексту або самим читачем. Відносини задають потенційні можливості пересування або **навігації** по гіпертексту. Стосунки можуть бути односпрямованим або двонаправленими. Відповідно, двонаправлені стрілки дозволяють рухатися користувачеві в обидві сторони, а односпрямовані - тільки в одну. Ланцюжок вузлів, через які проходить читач при перегляді компонентів тексту, утворює шлях або маршрут.

Елементи типології гіпертексту. Перше протиставлення належить до структури гіпертексту. Гіпертекст може бути ієрархічним або мережевим. Ієрархічне - деревоподібна - будова гіпертексту істотно обмежує можливості переходу між його компонентами. У такому гіпертексті відносини між компонентами нагадують структуру тезауруса, заснованого на родо-видових зв'язках. Ієрархічний гіпертекст не реалізує всіх можливостей технології гіпертексту. У середовищі розробників гіпертекстових систем він не користується популярністю (хоча і досить часто реалізується у працюючих системах).

Друге протиставлення характеризує не саму структуру гіпертексту, а можливості програмного забезпечення. Тут розрізняються **прості і складні** гіпертексти. Прикладом простого програмного забезпечення гіпертексту може служити електронне зміст документа, що дозволяє перейти до будь-якої частини змісту, минаючи етап перегляду всього тексту. До простого гіпертексту відноситься і система, яка дає можливість переглядати відсилання до літератури, що містяться в тексті, не звертаючись безпосередньо до списку літератури. Складні гіпертексти володіють багатою системою переходів між компонентами гіпертексту, в них відсутнє уявлення про базове тексті, з яким пов'язані другорядні за значимістю тексти. У певному сенсі нормальний, звичайний гіпертекст і є складним гіпертекстом.

За способом існування гіпертексту виділяються статичні і динамічні гіпертексти. Статичний гіпертекст не змінюється в процесі експлуатації; в ньому користувач може фіксувати свої коментарі, однак вони не змінюють суть

справи. Для динамічного гіпертексту зміна є нормальною формою існування. Зазвичай динамічні гіпертексти функціонують там, де необхідно постійно аналізувати потік інформації, тобто в інформаційних службах різного роду. Гіпертекстовою є, наприклад, Аризонська інформаційна система (AAIS), яка щомісяця поповнюється на 300-500 рефератів на місяць.

Відносини між елементами гіпертексту можуть спочатку фіксуватися творцями, а можуть породжуватися щоразу, коли відбувається звернення користувача до гіпертексту. У першому випадку мова йде про гіпертексту жорсткої структури, а в другому - про гіпертексту м'якої структури. Жорстка структура технологічно цілком зрозуміла. Технологія організації м'якої структури повинна ґрунтуватися на семантичному аналізі близькості документів (або інших джерел інформації) один до одного. Це нетривіальне завдання комп'ютерної лінгвістики. В даний час широко поширене використання технологій м'якої структури на ключових словах. Перехід від одного вузла до іншого в мережі гіпертексту здійснюється в результаті пошуку ключових слів. Оскільки набір ключових слів кожен раз може різнитися, щоразу змінюється і структура гіпертексту. Жорсткість чи м'якість архітектури гіпертексту залежить і від кількості інформації, яка в ньому міститься. Якщо вузлів у мережі гіпертексту порядку однієї-трьох тисяч, то найчастіше використовується жорстка архітектура, якщо ж кількість вузлів досягає декількох десятків тисяч або навіть мільйонів одиниць, то м'яка структура виявляється більш кращою, оскільки кодування жорстких зв'язків забирає надто багато часу. Зауважимо, що структура Інтернету часто функціонує як гіпертекст м'якої архітектури.

Технологія побудови гіпертекстових систем не робить різниці між текстовою та нетекстовою інформацією. Тим часом включення візуальної та звукової інформації (відеороликів, картин, фотографій, звукозаписів тощо) вимагає істотної зміни інтерфейсу з користувачем і більш потужної програмної та комп'ютерної підтримки. Такі системи отримали назву гіпермедіа або мультимедіа. Наочність мультимедійних систем визначила їх широке використання у навчанні, у створенні комп'ютерних варіантів енциклопедій.

Відомі, наприклад, чудово виконані CD-роми з мультимедійними системами по дитячих енциклопедій видавництва "Дорлінг Кіндерслі".

Деякі гіпертекстові системи. Технологічно в основі гіпертексту лежать комп'ютерні програми, які підтримують такі базові функції:

забезпечення швидкого перегляду інформаційного масиву (браузинг);

обробка посилальних відносин (звернення і виклик фрагмента тексту або іншої інформації, на яку проводиться відсилання);

навігація по гіпертексту, запам'ятовування маршруту руху; подання шляху руху в легко сприймається формі;

можливість формування звичайного лінійного тексту як результату руху по гіпертексту;

додаток гіпертексту новою інформацією;

введення нових відносин у структуру гіпертексту (для систем з жорсткою структурою).

Програмні оболонки гіпертексту, як правило, універсальні. Вони можуть використовуватися в різних областях для створення тематично різних гіпертекстів. Такі, наприклад, оболонка ZOG і розроблена на її основі промислова гіпертекстова система KMS (університет Карнегі-Меллон, США). Сфери застосування цих гіпертекстових систем надзвичайно різноманітні - від роботи з документацією та підтримки електронної пошти до гіпертекстів, призначених для експертів, що працюють над бюджетом. Є й спеціалізовані системи. Так, система NoteCards (продукт компанії "Xerox PARC") призначена для аналітичної роботи, а система WE, моделююча особливості отримання нового знання - для допомоги в авторській роботі. Найбільш популярні в даний час програмні пакети HyperCard компанії "Apple". Вони відносно прості у використанні. Гіпертекст в оболонці HyperCard представляється у вигляді каталожних карток. Користувач за допомогою досить простого інтерфейсу організує структуру картки і встановлює зв'язки між картками. Пакети HyperCard дозволяють поєднувати різні типи інформації, зокрема картки можуть включати графічну, звукову та ін інформацію. Слід зазначити, що сучасні бази даних також включають поля для візуальної та звукової форми

даних (пор., наприклад, базу даних ACCESS 7, що працює в середовищі Windows). Близька до HyperCard за своїми властивостями і програма SuperCard фірми "Silicon Beach". Деякі системи гіпертексту містять спеціальні засоби орієнтації користувача в гіперпросторі - карти або закладки, що відзначають найбільш відвідувані вузли гіпертексту. Комплексом засобів орієнтації має система Hypergate Writer фірми "Eastgate Systems Inc".

Рекомендована література:

1. Баранов А. Н. Введение в прикладную лингвистику. 3-е изд., стереот. Москва : Эдиториал УРСС, 2007. 360 с.
2. Дарчук Н. П. Комп'ютерна лінгвістика. Київ : ВПЦ «Київський університет», 2008. 351 с.
3. Денисюк В. В. Інформаційні технології в лінгвістиці. Умань : ФОП Жовтий О. О., 2015. 135 с.
4. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика : підручник. Донецьк : ТОВ «Юго-Восток, Лтд», 2006. 188 с.