

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний університет імені Тараса Шевченка
Інститут філології
Кафедра української мови та прикладної лінгвістики

Аналіз структурних особливостей мовлення людей із психічними розладами

Кваліфікаційна робота магістра
за спеціальністю 035 «Філологія»,
спеціалізацією 035.10 «Прикладна
лінгвістика»,
галузі знань 03 «гуманітарні науки»
ОП "Прикладна лінгвістика
(редакторсько-перекладацька
та експертна діяльність)"
Марини БОРОДИ

наукові керівники:
д.філол.н. Наталія ДАРЧУК
асистент Валентина РОБЕЙКО

рецензент:
к.філол.н.,доц. Оксана ЗУБАНЬ

Київ-2021

ЗМІСТ

ВСТУП	3
РОЗДІЛ 1. ЛІНГВІСТИЧНІ ПОНЯТТЯ ТА ПРИНЦИПИ	8
1.1. Головні лінгвістичні принципи	8
1.1.1. Семантика виразу та його імплементація в межах дослідження	9
1.1.2 Семантика дискурсу та його імплементація в межах дослідження	13
1.2 Дистрибутивна семантика та її використання в межах дослідження	15
Висновки до першого розділу	17
РОЗДІЛ 2. ПСИХІАТРИЧНІ ПОНЯТТЯ В КОНТЕКСТІ ДОСЛІДЖЕННЯ МОВЛЕННЯ ЛЮДЕЙ ІЗ ПСИХІЧНИМИ ВАДАМИ	18
2. 1 Поняття психічної структури мислення.....	18
2.2 Об'єкт дослідження та синдроми	21
2.3. <i>Мовлення</i> людей із психічними вадами: характерні особливості (у контексті психотерапевтичної та психіатричної практик)	24
2.4 Лінгвістичні дослідження орієнтовані на мовлення людей із психічними розладами.....	28
Висновки до другого розділу	29
РОЗДІЛ 3. ПОПЕРЕДНЯ ОБРОБКА ТА ПІДГОТОВКА МОВЛЕННЄВОГО МАТЕРІАЛУ	31
3.1. Створення корпусу для подальшої обробки даних	31
3.1.1. Збір та обробка мовленнєвого матеріалу	31
3.1.2 Формування корпусів.....	33
3.2 Підготовка корпусів для подальшої обробки	36
Висновки до третього розділу	41
РОЗДІЛ 4. АНАЛІЗ ЗІБРАНОГО МОВЛЕННЄВОГО МАТЕРІАЛУ	43
4.1 Ключові поняття та їх місце у структуризації мовлення	43
4.2 Візуалізація результатів	46
Висновки до четвертого розділу	48
ВИСНОВКИ	50
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	53
ДОДАТКИ	59

ВСТУП

У всесвітньому науковому доробку лише збільшується кількість досліджень, що стосуються ментальних структур людини та особливостей їх функціонування. Ці роботи здійснюються не лише у межах дискурсу психіатрії, а і як роботи у галузі мовознавства, комп'ютерних наук, тощо. Адже саме завдяки тому інструменту, із яким працюють лінгвісти сьогодні (наприклад, корпусна лінгвістика та стилістика зокрема, машинне навчання, тощо), можливо дослідити мовленнєві структури та закономірності мовлення за допомогою методів обробки природної мови.

Існує чимала кількість наукових статей та досліджень написаних, як при європейських, американських, так і вітчизняних інститутах, або ж лабораторіях та дослідницьких центрах. Наприклад, погляд на лінгвістичний аспект при роботі із мовленнєвим матеріалом при шизофренії у роботі «Лінгвістичний аспект при вивченні шизофренії: розумова стурбованість як лінгвістична патологія через позитивні симптоми» (англ. «The linguistics of schizophrenia: thought disturbance as language pathology across positive symptoms. *Frontiers in psychology*») [33] від Вольфама Хінцена та Джона Россело, де розглядається аспект саме лінгвістичної (не) організації мислення пацієнта при шизофренії.

Цікавою роботою також може бути книга від Януша Вробеля «Мова та шизофренія» [33], де розглядаються ключові аспекти функціонування саме мовлення людей із психічними вадами. Наприклад, розгляд мовлення в контексті мовленнєвого акту (даний аспект згадується у першому розділі більш детально).

Враховуючи той інструмент, яким користуються сучасні спеціалісти (корпус та робота із ним), зріст досліджень, які мають на меті встановлення та обробку залежностей у мові та мовленні ми вважаємо закономірним та логічним. Тож, ми вважаємо, що наше дослідження та спроба описати закономірності (або ж їх відсутність) у мовленні двох груп мовців (люди із психічними вадами та

люди без психічних вад) є актуальним та як таким, що розвивається у межах сучасних тенденцій.

Головною *метою дослідження* є виокремлення та визначення мовленнєвих особливостей та структурних закономірностей у мовленні людей із психічними вадами (мовлення хворих, яким діагностували шизофренію та шизоподібний розлад). Для увиразнення результатів ми маємо на меті порівняти приклади мовленнєвого матеріалу даної групи людей із матеріалом людей без психічних вад. Дані групи представлені двома корпусами – основний корпус та референтний корпус (більш детально про корпуси та їх збір у другому розділі).

На початку нашого дослідження, перед нами постали питання: а чи можемо ми розглядати мовлення підгрупи пацієнтів хворих на шизофенію та шизоподібні розлади, як би ми опрацьовували мовлення людей без названих відхилень? Яким чином такі поняття як семіотичний трикутник, мовленнєвий акт та комунікативна інтенція функціонують у даному контексті? І чи має сенс обробка мовленнєвого матеріалу у ході нашого дослідження зокрема? *На ці та інші питання, ми намагаємося дати відповіді у першому розділі.*

Одним із головних питань, на які нам потрібно дати відповідь є – які структурні аспекти мовлення ми можемо сформулювати у ході даного дослідження? На це питання, ми спробували дати відповідь у четвертому розділі.

Наукова проблема: Мовлення кожного окремого суб'єкта різниться значною мірою. Кожен із нас здатен відчувати не лише певні експлікативні особливості мовлення окремої людини (словниковий запас та здатність формулювати структуроване повідомлення, тощо), а і імплікативні (просодія, певна внутрішньо закладена інформація).

Аспект функціонування та структурних особливостей мовлення людей із психічними вадами завжди залишаться темою для роздумів. Саме даний аспект нас і цікавить значною мірою – які структурні особливості мовлення людей із

психічними вадами ми можемо охарактеризувати за допомогою лінгвістичних методів?

Мета нашого дослідження передбачає наступні *завдання дослідження*, які ми могли б умовно розділити на два головних блоки – попередня обробка та безпосередньо дослідження.

Попередня обробка передбачає наступні етапи виконання роботи:

1. Збір матеріалу та його оформлення за допомогою програми Praat, де ми здійснюємо анотацію записів.
2. Оформлення корпусу за допомогою мови програмування Python у середовищі PyCharm Community Edition.
3. Видалення непотрібних та зайвих елементів використовуючи бібліотеку regex (код описано третьому розділі).
4. Проведення стематизації та лематизації для подальшого обрахунку та роботи із мовленнєвим матеріалом.
5. Видалення стоп-слів, які могли б завадити при обрахунку подальших даних.

Другий етап передбачає безпосередньо обробку даних та складається із двох основних завдань – обрахунок еквівалентних класів та дослідження категорії аномальності тексту.

Для обрахунку еквівалентних класів, ми здійснили:

1. Латентний семантичний аналіз за допомогою якого обрахували ключові слова та поняття наявні у текстах, які складають корпуси.
2. Увиразнили ключові слова у мовленні.
3. Візуалізували отримані результати (для подальшого порівняння) за допомогою бібліотеки matplotlib мови програмування Python.

Об'єкт дослідження: мовлення категорій мовців, які були представленні вище.

Предметом дослідження є залежності та закономірності, які притаманні мовленню людей із вказаними психічними проблемами.

Методики і методи дослідження використані у ході даного дослідження передбачають використання експертно-фонетичного методу, який використаний при розробці і роботі із корпусом. За допомогою методу емпіричного аналізу, ми і намагаємося відповісти на поставлене питання в ході нашого дослідження за допомогою зібраних даних та їхнього ретельного аналізу. Саме для цього, ми користуємося синтезом квалітативного та квантитавного підходів до аналізу мовлення. Також, одним із методів є метод зіставлення, адже задля візуалізації та увиразнення наших результатів ми збрали референтний корпус із мовленням мовців, яким притаманна гарно структурована та виразна манера ведення монологу.

Характеристика новизни і практичного значення: ми сподіваємося, що у ході нашого дослідження, нам вдасться визначити закономірності притаманні мовленню людей із психічними вадами. Власне, виявлені закономірності та (ймовірна) відповідь на питання поставлені вище і мають складати новизну даного дослідження.

Наше дослідження може мати теоретичне значення, як робота покликана дослідити закономірності та аспекти функціонування мовлення людей із увиразненими психічними вадами (про симптоми, із якими працюємо більш детально вказано у першому розділі). Вона може мати не лише цінний вклад в розвиток вивчення мовлення, а і дискурсу та принципів його побудови.

Ми вважаємо, що практичний застосунок результатів, або ж певних аспектів даної роботи здатний не лише бути корисними для розвитку дискурс аналізу, а і спонукати інших зацікавлених (дослідників) до дискусії та роботою над увиразненням та удосконалені підходів стосовно аналізу аспектів мовлення людей із психічними вадами.

Структура: дипломна робота поділяється на чотири головних розділи.

У першому розділі ми представляємо ті лінгвістичні поняття, які нас цікавлять та в межах яких ми і будемо намагатися опрацювати мовленнєвий матеріал.

У другому розділі ми представимо та опишемо ключові аспекти для подальшої роботи із дослідження характерних особливостей мовлення людей із психічними вадами.

Третій розділ нашої роботи складається із опису процесу підготовки (збір корпусів та їх обробка) мовленнєвого матеріалу.

Четверний розділ складається із опису процесу аналізу та візуалізації результатів.

РОЗДІЛ 1. ЛІНГВІСТИЧНІ ПОНЯТТЯ ТА ПРИНЦИПИ

Що таке мовлення людини? На нашу думку, мовлення кожної людини є унікальним та як таким, що формується під впливом різних чинників (соціальних та психічних). На деякі із них, ми можемо впливати та змінювати (словниковий запас, використання тих чи інших виразів, мови); натомість, деякі із них можуть бути зумовлені складними психічними процесами (істерія, невроз, відчуття тривоги викликане внутрішньо психологічними тригерами, тощо). Проте, безсумнівним є те, що мовленню соціального суб'єкта притаманні «характерні особливості», які формуються як під впливом соціального контексту, так внутрішньо-психологічних закономірностей. Власне, згадані «особливості» і цікавлять нас та становлять основу наукового питання.

Яким чином ми могли б класифікувати та простежити згадані особливості? Який інструментарій доступний нам як лінгвістам для виокремлення даних залежностей, і чи існує він узагалі? Більш детально про дискурс в межах якого ми будемо проводити дослідження мовлення та його систематичних залежностей, ми будемо розповідати у даному розділі (пункт 1.1).

Даний теоретичний розділ, на нашу думку, є одним із найважливіших, адже саме тут ми намагатимемося представити ключові поняття, які і мають увиразнити подальшу методику стосовно обробки та аналізу даних.

1.1. Головні лінгвістичні принципи

У даному підрозділі, ми представимо ті лінгвістичні поняття, якими будемо оперувати в ході дослідження. Головні аспекти, які представимо – це сприйняття та актуалізація питання семантики виразу та семантики дискурсу для нашого дослідження. Саме в межах даних понять ми розвиваємо наш дискурс.

Уважаємо, що вказані аспекти не лише описують та ідентифікують головний аспект дискурсу в межах якого ми і розвиваємо наше дослідження, а і допомагають підійти контекстуально до розгляду головної методики, яка буде здійснюватися в межах дистрибутивної семантики та аналізу.

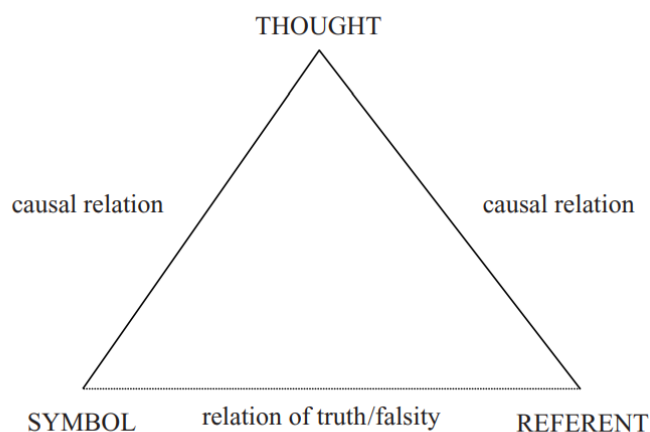
1.1.1. Семантика виразу та його імплементація в межах дослідження

Перш ніж продовжувати та представляти аспекти та головуючі напрями аналізу дискурсу, ми маємо ідентифікувати та увиразнити наступний аспект – що таке слово та яким чином воно набуває свого значення? Проте, найголовніше питання наступне – яка специфіка набуття словом свого значення у мовленні психічно-хворої людини? Чи є різниця та що нам варто згадати, щоб упевнитися напевне, що дане дослідження має місце бути в межах лінгвістичного дискурсу?

Перш за все, розглянемо тут найважливіші аспекти семіотичної теорії. Адже саме завдяки даній теорії ми здатні описати процес набуття словом *значення*. Семіотика включає у себе такі ключові розділи лінгвістичної теорії, як синтаксис, семантика та прагматика, які є взаємозалежними (більшою, або ж меншою мірою). [3]

Нижче представлений семіотичний трикутник, який власне і зображує ті залежності, які є основними у процесі набуття словом його значення. Тож, головними аспектами, які представлені є «символ», «референт» та «думка». Символ – це власне образ (комунікативний) того чи іншого об'єкта\ поняття\ суб'єкта. Натомість, референт – це один із варіантів використання даного поняття, тоді як думка – радше ментальна репрезентація поняття. [25]

Рис 1:



Даний трикутник, а певніші ті залежності, які він ідентифікує, і є ключовими аспектами набуття сенсу певною семіотичною одиницею. Проте, якщо у актуалізації даного поняття в межах мовлення людей без психічних хвороб сумніватися не доводиться, чи можемо ми поставити під сумнів актуальність та доцільність даного трикутника в межах мовлення людей, які є психічно нездоровими?

Тут перш за все пригадаємо про різновид діагнозів та специфіку процесу перебігу психічного захворювання як такого (при шизофренії та шизоподібному розладі в тому числі). Дані аспекти можуть значною мірою різнитися, і перш ніж нам дати остаточну відповідь – зазначимо кого саме ми маємо на увазі.

Нижче ми більш детально аргументуємо, що маємо на меті розглядати та працювати із структурою мовлення хворих, яким притаманне маячіння (при шизофренії та шизоподібному розладі). Адже саме структурні залежності маячіння є основним предметом нашого дослідження.

Ми вважаємо, що попри те, що структурні особливості психічного апарату (відсутність кореляції із навколишнім світом, тощо) людей із психічними вадами значною мірою різняться (від звичного сприйняття реальності) та є складними для розуміння, це не дає нам жодних сумнівів у тому, що сам принцип набуття семіотичних залежностей у мовленні, може різнитися від обумовленого вище (принаймні розглядаючи саме ті випадки, які представлені у нашому першому корпусі).

Для підтвердження даного судження, розглянемо прагматичний аспект мовленнєвого акту психічно-хворих (у яких простежується марення). Перш за все зазначимо, що ключовим аспектом прагматики є саме дослідження, які стосуються ситуативного погляду на мовлення, яке активізується в соціальному, ситуативному, політичному тощо контекстах. [40] Тобто, третій аспект семіотичної теорії є саме взаємодія мовця із іншими учасниками (ситуативного) дискурсу. Ми вважаємо, що саме приналежність та наявність даного аспекту

мовленню людей із психічними вадами і може виправдати актуальність головного питання даної роботи.

Яким чином прагматична складова мовлення реалізується у мовленні? І найголовніше – чи реалізується вона? Якщо так, то ми не бачимо підстав не розглядати мовлення категорії мовців згаданої вище та його особливостей у межах даного дослідження.

Мовці із психічними хворобами (на різних стадіях розвитку захворювань) не були помічені, як ті, що не є включеними у контекстуальний дискурс. Тобто, звісно, що внутрішньо-систематичні залежності зумовлюють мовлення (у даному випадку) бути «незрозумілим» та «незв'язним», проте, це не означає, що мовець не здатен реагувати на комунікативний акт на нього спрямований іншим мовцем (хоча реакція і може бути незрозумілою). Адже, сам факт відповіді на питання, яке є поставленим іншим учасником дискурсу, вже свідчить про стосунок мовця до діалогу\полілогу. Саме тому, ми не бачимо підстав вважати мовлення людей із психічними вадами як таке, що не доцільно було б досліджувати.

Досліджуючи семантику виразу та працюючи із аналізом структурних особливостей мовлення тієї чи іншої групи людей, виникає питання – а що таке значення виразу (його семантика, у даному випадку) та що саме має бути включеним в мовленнєвий акт, щоб ми могли із впевненістю розглядати та працювати із його значеннєвою організацією (семантикою)?

Наступним кроком після розгляду концепту, що стосується питання набуття словами свого значення, ми хотіли б розглянути концепт «мовленнєвого акту», який є важливим для роботи із дискурсом. [34] Адже, так само як із існуванням семіотичних залежностей та утворенням значень слів, варто аргументувати специфіку утворення (наші припущення) стосовно мовленнєвого акту у контексті даної роботи.

Тут варто пригадати роботу та вклад ідей філософської школи Філософія буденної мови, який набув чималої популярності в другій половині 20-го століття. [61] Евристика та головні підходи даної течії були зумовлені ідеями про важливість мовленнєвого акту як такого – що те, як влаштований даний акт є набагато важливішим, аніж абстрактне значення його. Попри те, що даний рух не зберіг «своїх позицій», як актуальний розвиток науки про мову та мовлення, вони змогли «дати лінгвістиці» поняття мовленнєвого акту, яким ми будемо користуватися в межах даної роботи.

Теорія мовленнєвих актів та розподілу їх на три головних типи – іллокутивний, локутивний, перлокутивний – розробив та вивів філософ та лінгвіст Дж. Остін, який належав до вищезгаданої філософської школи (Філософія буденної мови). Специфіка даних понять полягає у тому, що мовленнєвий акт вже не лише пояснює, або ж описує реальність, а радше – сам стає певною дією, частиною реальності.

Згодом дану теорію розвинув філософ Джон Серль. Він змінив та удосконалив певною мірою розподіл іллокутивних актів (зокрема, охарактеризував такі аспекти іллокутивного акту, як асертиви, директиви, декларативи, тощо). Проте, ми не вважаємо, що у межах даного дослідження було б доцільним згадувати про це, тому даний аспект описувати не будемо.

Розглядаючи аспект мовленнєвого акту, ми б також хотіли згадати про поняття *комунікативної інтенції* (за Граймсом, Grice's M-intentions), а саме – «якщо у виразі відсутня комунікативна інтенція, набір символів ми вважаємо несеміотичним». [29] Теорія комунікативної інтенції інша класична теорія стосовно мовленнєвого акту, яку Пол Грайс почав розробляти в другій половині ХХ століття. Згідно із нею будь-який іллокутивний акт має на меті «певною мірою інтенцію».

Грайс пропонує наступну залежність:

«U має на увазі щось кажучи R» є правдивим для певної групи слухачів (слухача) лише у тому випадку, якщо U сказав R розраховуючи, що (1) слухач (реципієнт) зможе надати певну реакцію на сказане, (2) реципієнт визнає, що U бажає (1); (3) та нарешті – реципієнт задовольнить бажання U (1), на базисі (2).» [29]

Згідно із даними залежностями зазначеними у теорії Грайса «Значення» [29], значення семантичних одиниць (наприклад, речення, або ж окремі слова) варто аналізувати згідно із поняттям комунікативної інтенції (англ. ‘communicative intention’, ‘meaning intention’, ‘m-intention’).

Із лінгвістичної точки зору, ми можемо охарактеризувати (у загальних рисах та у дискурсі лінгвістичних теорій представлених вище) побудову повідомлення мовцем свого повідомлення. Проте, яким чином поняття комунікативної інтенції та мовленнєвого акту як такого актуалізуються у мовленні людей із психічними вадами?

Ми можемо лише зробити певні припущення стосовно комунікативної інтенції у дискурсі психічно хворих. Так само, лише припустити ми можемо – яким чином подібні аспекти функціонування комунікації будуть активізуватися. Проте, сам факт їхньої реалізації у мовленні є безсумнівним.

За допомогою прикладу вище, ми мали на меті представити, яким чином можна було б інтерпретувати аспект специфіки комунікації в межах двох розрізнених дискурсів – мовлення людей із психічними вадами та мовлення людей без увиразнених психічних вад.

1.1.2 Семантика дискурсу та його імплементація в межах дослідження

Структура та аналіз дискурсу мовлення посідає ледь не ключове місце, в таких науках як лінгвістика, комп’ютерна наука (computer science), тощо. Дійсно на даний момент сформувалось безліч теорій та аспектів аналізу дискурсу (стосовно лінгвістичного погляду на специфіку функціонування дискурсу) від квантитативного аналізу тексту закінчуючи семантичним аналізом, тощо.

Для кожного із методів аналізу та обробки тексту характерні ті чи інші особливості підходу та аналізу. Проте, для більшості, якщо не для всіх характерним буде саме розподіл та розбиття тексту на структурні одиниці та дослідження специфіки взаємодії однієї із іншою (наприклад, СxG, або ж аналіз тексту методом дистрибуції, тощо).

Загалом вирізняють різні підходи та методики до вивчення тексту як дискурсу. Однією із методик є саме підхід, який зумовлює вивчення мови (тексту) як системи. Наступним способом аналізу дискурсу та досліджень мови є дослідження, які орієнтовані на дослідження мови у використанні безпосередньо. У даному випадку, дослідник працює радше зв'язок та залежності між різними елементами дискурсу. Інша методика аналізує залежності та характерні особливості, які зумовлені певною темою, контекстом. І останньою методикою є дослідження, які стосуються контексту соціуму, культури. [56]

Розрізняють 2 способи дискурс аналізу – квалітативний та квантитативний. Квалітативний спосіб передбачає дослідження на обмеженій кількості даних, хоча способи та підходи є більш різноманітними, аніж у квантитативному, який орієнтований на дослідження на великій кількості даних.

Сучасні мовознавці стверджують, що мова відображає ситуацію [26] (мовленнєву). Хоча ми б і хотіли наголосити, що погоджуємося лише частково – на нашу думку мовлення учасників дискурсу не відображає ситуацію, а – створює її.

Все ж є деякі аспекти із сучасної лінгвістичної теорії, які хотіли б наголосити та згадати у даній роботі. Перш за все, пригадаємо про семіотичний аспект, який стосується аналізу дискурсу. Тут йде мова про знакову систему кожної мови (та дискурсу) та складові даної системи, саме семіотична структура мови є визначальним та диференціальним аспектом, який відрізняє мовленнєві структури, як такі.

Наступними аспектами дискурс аналізу є питання мовлення та акту говоріння, як аспекту активного говоріння, у тому чи іншому соціальному контексті. Тут варто також і додати про соціокультурний та політичні аспекти, як ті що ідентифікують та визначають мовленнєву ситуацію разом із аспектом активного говоріння.

Також, важливим для дискурс аналізу є матеріальний аспект. Тобто тут ми наголошуємо на фізичних складових аспекту говоріння, а саме – місце, час, учасники мовленнєвого акту (мовці).

Дані аспекти і формують лінгвістичну систему. Зокрема, (деякі із них) ми будемо використовувати для опису та попереднього аналізу, ідентифікації мовлення людей із психічними хворобами (психотична психологічна структура, шизофренія) та людей без увиразнених психічних проблем.

Підсумовуючи згадане вище: дискурс аналіз – це дослідження комунікативної (та структурної) складової мови\комунікації в залежності із певною теорією, або ж лінгвістичним формалізмом. [55] Проте, є безсумнівним, що дослідження стосовно дискурсу та його внутрішніх закономірностей ми отримуємо саме завдяки дослідженню його окремих частин та взаємодії між ними.

Варто також зазначити, що усі згадані вище аспекти аналізу дискурсу знаходять своє втілення в межах нашого дослідження, яке стосується власне мовлення людей із яскраво вираженими психічними вадами. Тож, без жодного сумніву в актуальності та доцільності даної роботи, ми б хотіли надалі розглянути той лінгвістичний концепт в межах якого будемо аналізувати мовленнєвий матеріал.

1.2 Дистрибутивна семантика та її використання в межах дослідження

На нашу думку, одним із найвдаліших способів дослідити та представити внутрішні структурні залежності у реченні, та у дискурсі є використання саме підходу охарактеризованому у роботі Зелінга Харіса «Дискурс аналіз». [32]

Ключовим аспектом, згідно із статтею, є твердження, що дистрибутивний аналіз в межах певного дискурсу здатен задовільнити і аспект пов'язаний із обмеженістю досліджень за допомогою дескриптивної лінгвістики (дескриптивна граматики працює виключно із окремим реченням), і досліджень стосовно взаємодії соціальної ситуації, в якій мовець реалізує мовленнєвий акт, із мовою. Обидва дані підходи чи не ідеально підходять нам для досліджень мовлення двох різних категорій, які були окреслені у розділі 1. [2]

Дистрибутивна семантика представлена теорією, що слова, які зустрічаються в однакових контекстах є однаковими або ж схожими за значенням. Дана теорія була популяризована британським лінгвістом Джоном Фірсом. Тож, ми могли б припустити, що схожим чином можна було б охарактеризувати і загальну структуру дискурсу (за допомогою залежностей між словами, контекстами, які їх представляють, та суміжними контекстами).

Що характеризує дистрибутивну структуру (за Зелінгом Харісоном, 30)? Дистрибуція того чи іншого елемента характеризується як певна сума мовленнєвого оточення. За роботою Зелінга Харісона, ми можемо увиразнити існування дистрибутивної структури згідно із наступними аспектами: (1) синтаксичні залежності існування певного мовленнєвого елемента (ми будемо фрази та дискурс із одиниць відповідно до їх класової приналежності); (2) обмежувальні властивості утворення семантичних одиниць; (3) ми можемо охарактеризувати використання одного елемента у залежності до будь-якого іншого елемента, тож дистрибутивні судження здатні описати увесь мовний матеріал; (4) обмеження та використання кожного елемента можуть бути описані за допомогою внутрішніх залежностей.

Деякі дослідження стверджують, що повтор фраз та залежностей у тексті є чудовим способом дослідити специфіку використання та увиразнити лінгвістичні властивості лінгвістичної одиниці. [20] У даному випадку, мовці користуються не лише внутрішніми залежностями та впорядкованістю мовленнєвих структур, а і їх невідповідностями у мовленні.

Існує безліч методів, які власне і були виведені на базисі ідей увіражнених в межах дистрибутивної семантики. Зокрема, в даному дослідженні ми скористаємося латентно-семантичним аналізом (LSA) для знаходження ключових слів та понять у текстовій та моделі засновані на базисі синтаксичних, генеративних залежностей, а саме – визначення аномалій у тексті. [39]

Висновки до першого розділу

У першому розділі ми описали та детально дослідили аспект семантики виразу та семантики дискурсу. Вказали на ключові аспекти функціонування мовленнєвого акту та наголосили на доцільності їх вивчення не лише в контексті мовлення людей без психічних вад, а і людей із психічними вадами (такими як шизофренія). Даний аспект ми представили та аргументували вказавши, що усі ключові аспекти мовленнєвого акту активізуються (попри те, що даний процес складно описати) у межах дискурсу людей із психічними вадами. Функціонування даних аспектів у межах мовлення людей без психічних проблем (вказаних у підрозділах 1.2) ми вважаємо як само собою зрозумілим.

Також, у першому розділі ми представили основний підхід для обробки вище згаданої мовленнєвих даних. Подальше дослідження буде здійснюватися в межах концепту дистрибутивної семантики.

Дистрибутивна семантика та способи дослідження відповідно до її парадигм досліджують дистрибуцію того чи іншого елемента як певну суму мовленнєвого оточення. Ми вважаємо, що саме даний підхід дасть змогу нам якісно дослідити мовленнєвий матеріал та візуально представити ті, чи інші закономірності.

Саме в межах даного підходу та користуючися пов'язаними із ним методів дослідження та аналізу природної мови, а саме – методом латентного семантичного аналізу, ми збираємося проводити обрахунок та дослідження задля тих чи інших закономірностей у мовленні людей із психічними вадами.

РОЗДІЛ 2. ПСИХІАТРИЧНІ ПОНЯТТЯ В КОНТЕКСТІ ДОСЛІДЖЕННЯ МОВЛЕННЯ ЛЮДЕЙ ІЗ ПСИХІЧНИМИ ВАДАМИ

У даному підрозділі, ми більш детально представимо та опишемо специфіку мовлення людей із психічними вадами. Також спробуємо увиразнити предмет нашого дослідження за допомогою представлення його в межах природнього для нього дискурсу.

У підпунктах 2.1 та 2.2, ми надаємо ключову інформацію для подальшої роботи із мовленням пацієнтів із психотичним розладом мислення. Тут ми спробуємо охарактеризувати ключові аспекти дискурсу, понятійним апаратом якого будемо користуватися (психотичний тип психічної структури людини, МХБ-10, тощо).

Також, тут ми згадаємо, характерні аспекти мовлення людей із психічними вадами, які могли б стосуватися дослідження мовлення в межах дискурс аналізу (специфіка маячіння, асоціативний процес пацієнтів, тощо).

У підпункті 2.3, ми згадаємо дослідження та їх методи, які власне і стосуються мовлення людей із психічними вадами.

2. 1 Поняття психічної структури мислення

Завдяки роботам психіатра Еміля Крепеліна, Зигмунд Фрейд сформулював та вивів різні рівні функціонування психічної структури людини, а саме – невротична, психотична та погранична. [65] Даний момент є принципово важливим для нашого дослідження, тож розглянемо його більш детально та опосередковано.

На базисі даного твердження і ґрунтується наша робота, адже саме структуру мовлення (а також їхню відмінність, або ж схожість) людей, яким притаманна психотична структура мислення та людей, яким вона не притаманна, ми і будемо намагатися аналізувати. Попередньо, наша мета: виокремити та ідентифікувати структурні закономірності кожної, їхню схожість і відмінність.

Характерні особливості психотичної структури людини

Кажучи про психотичну структуру психічного апарату людини, ми маємо на увазі саме психічного хворих людей, розумовій діяльності яких притаманні такі аспекти, як марення, галюцинації, псевдогалюцинації, тощо. Проте, як би ми могли охарактеризувати психічну хворобу людини?

Чи не найкраще пояснення до даної теми, ми знайшли у підручнику «Психіатрія» від Кербиков О.В., Коркина М.В., Наджаров Р.А., Снежневский А.В. Психиатрия: «Кажучи про психічні хвороби, які також називаються психоз, мають на увазі таку зміну діяльності головного мозку, коли психічні реакції різко, грубо суперечать реальним ситуаціям (І.П. Павлов), що і характеризується, як розлад відображення реального світу та розлад поведінки. [...] Це і відображається, в тому, що настрої, відчуття, судження, рішення виникають при психозах із середини психічного апарату; у той час як при здоровому стані рішення, судження, почуття викликані внутрішніми збудниками та знаходяться в гармонійному зв'язку із реальним світом (В. Гринзер)». [11]

Варто зауважити, що при психічному захворюванні, значною мірою порушуються саме найбільш складні форми мислення (раціональне, абстрактне мислення, тощо), що і виходить із цитати наведеної вище. Проте, у залежності від захворювання порушуються (або ж не порушуються) безпосереднє відображення реальності (що стосується відчуттів). Даний аспект може варіюватися у різних пацієнтів та відповідно до їх захворювання, його симптоматики, періоду розвитку захворювання, тощо.

При появі психічного захворювання, внутрішні психічні процеси видозмінюються: порушується пам'ять, причинно-наслідковий зв'язок та сприйняття реального погіршується, людина здійснює невмотивовані та алогічні вчинки, простежується затьмарення свідомості. [11] Звісно, що лише спеціаліст здатен ідентифікувати певне захворювання, проте навіть вказані вище симптоми здатні продемонструвати нам ймовірні психічні відхилення. Зауважимо лише,

що є очевидним, що дані відхилення безпосередньо впливають на можливість людини формулювати когерентне повідомлення.

Характерні особливості невротичної психічної структури людини

Патологія станів головного мозку людини не завжди веде до тяжких та незворотніх процесів зміни психіки, трапляються також випадки психічних захворювань, при яких кореляція людини із зовнішнім світом не порушена (на відмінну від шизофренії, наприклад). (В. Гринзер)». [11] Наприклад, при неврозах людина все ще відчуватиме дискомфорт, проте це жодним чином не вплине на її фізичний стан та на її здатність функціонувати як соціальний суб'єкт. Натомість, у випадку ж психозу, пацієнти стають недієздатними (у період загострення, особливо), а інколи – небезпечними для оточуючих.

Характерні особливості пограничної

До цього моменту, ми визначили та класифікували існування **неврозів** та невротичної патології, які є відмінними від **психозів** (шизофренія, шизоподібний розлад, тощо). Проте, є також пограничний тип психічної структури людини. Як зазначається Отто Кернбергом, всесвітньовідомий психоаналітик, президент Міжнародної Психоаналітичної Асоціації (1997 – 2001), чи не найскладнішим в психіатрії є класифікація та диференціації пограничний розлад характеру. [8]

Класифікація та лікування даних розладів є надзвичайно складним процесом, що і відзначається (навіть) у цінних роботах, які покликані розглянути та дослідити специфіку існування та ідентифікації пограничного типу психічної структури людини. Хоча і зазначається, що генетична історія пацієнта не надто допомагає розрізнити невротичну, пограничну, або ж психотичну симптоматику. Проте, згідно із твердженням Отто Кернберга, є ймовірним, що саме структурний підхід у психіатрії допоможе нам краще ідентифікувати та простежити співвідношення між генетичною схильністю до захворювання та його проявами. [8]

Загалом для людей із пограничним розладом особистості є характерним імпульсивна поведінка, емоційна нестабільність, тривожність, схильність до самогубства та до завдання шкоди собі, тощо. Для діагностики пограничного розладу особистості, лікар-психіатр повинен засвідчити наявність як мінімум п'яти симптомів. [42] У МКБ-10 пограничний розлад особистості розглядається як підвид емоційно-нестійкого розладу особистості.

У межах даної роботи, ми не будемо вчитися працювати із проявами пограничної структури функціонування психічного апарату людини, адже дана тема не матиме жодного безпосереднього відношення до нашої роботи. Тим не менш, ми вважаємо важливим у загальних рисах описати межі та загальні особливості даного поняття.

2.2 Об'єкт дослідження та синдроми

Чому наше дослідження орієнтоване саме на роботу із текстами зібраними серед пацієнтів, які хворіють на шизофренію та шизоподібний розлад? Це зумовлено не лише інтересом (шизофренія є одним із найбільш відомих психічних захворювань), а радше можливістю простежити та відмітити певні структурні особливості розпаду мовлення, як такого. Певну сумбурність є можливим простежити саме на прикладах мовлення людей, хворих на шизофренію, адже дані аспекти яскраво увиразненні у даному випадку.

Тут також присутнє наше спостереження та думка, що у мовленні пацієнтів, хворих на шизофренію, все ж простежується певна логічність (попри свідчення про його розірваність та відсутність зв'язку). На нашу думку, враховуючи той факт, що психічна структура психотиків характеризується як та, при якій «настрій, відчуття, судження, рішення виникають із середини психічного апарату», є можливим простежити все ж внутрішню зумовленість та контекстуальну залежність навіть при маренні.

Відмітимо, що у Міжнародній класифікації захворювань 10-го перегляду (МКХ-10) шизофренія, шизотиповий розлад особистості, хронічні розлади

марення, гострі та перехідні психотичні розлади, шизоафективні розлади, маячний розлад та інші неорганічні психотичні розлади об'єднані в один блок класу «Психічні розлади та розлади поведінки» (F00-F99). [62] Із міркувань типологізації розглянемо ймовірність використання мовленнєвого матеріалу зібраного у кожному із згаданих вище захворювань. Тут ми також представимо фактори, які уможливають подальшу обробку матеріалу, або ж навпаки – унеможливають її.

Шизофренія

Шизофренія має різні форми та етапи розвитку захворювання, як такого. Наприклад, розрізняють наступні форми шизофренії – параноїдна, гебефренічна, кататонічна, недиференційна, проста. [60] Деякі із цих форм будуть зустрічатися і у даній роботі, зокрема, ми будемо опрацьовувати записи пацієнтів із параноїдною, гебефренічною шизофренією та не тільки.

Окремо варто наголосити на ключових аспектах перебігу такого захворювання як шизофренія. На перших етапах розвитку захворювання зустрічаються невротичні розлади пацієнтів, згодом з'являються афективні розлади, марення, та психо-органічні розлади психічної системи хворого (або ж хворої). Тобто, у пацієнта спочатку спостерігаються невротичні розлади психіки, астенічні та фобічні розлади, згодом з'являється марення та галюцинації (або ж псевдогалюцинації). Також, частим проявом шизофренії та її розвитку у пацієнта є розвиток синдрому Кандинського-Клерамбо, парафренія та апатичне слабоумство. [9] Можемо припустити, що для кожного із етапів характерні ті чи інші мовленнєві особливості.

Розрізняють шизофренічне та епілептичне слабоумство, яким характерні певні специфічні ознаки. В контексті нашого дослідження ми будемо розглядати саме шизофренічне слабоумство, тож згадаймо чим саме воно відрізняється від епілептичного та чому для нас є важливим розуміння даного аспекту. Саме при шизофренічному, або ж апатичному слабоумстві спостерігається інтелектуальна

індиферентність та бездіяльність. Варто зауважити – саме здатність до розумової діяльності можуть зберігатися ще тривалий час. [9] Звісно, що даний момент із шизофренічним слабоумством буде згадано у контексті даного дослідження.

Задля логічного завершення, зауважимо також і про відмінність апатичного (або ж шизофренічного) слабоумства та епілептичного. У випадку епілептичного – у хворого (або ж хворої) знижується пам'ять, а будь-які дрібні аспекти щоденного життя набувають неймовірної цінності, він або ж вона сприймають будь-які дрібні моменти, як надзвичайно важливі. З'являються гіперболізовані та виразні особливості мовлення. Процес мислення ускладнений, хворим надзвичайно складно закінчити думку. [9] Тож, в контексті розгляду мовленнєвого матеріалу увиразнити та розрізнити слабоумство даних двох типів не є складним.

Шизотиповий розлад особистості

Шизотиповий розлад особистості найчастіше протікає саме на субпсихотичному рівні та за своїм клінічним оформленням *нагадує початкові прояви психотичних форм хвороби* (власне, тому даний розлад і визначаю в одну групу). Хвороба розвивається повільно та поступово, із часом змінюючи звичайний стиль та спосіб життя людини. [60]

При шизоподібному розладі особистості поведінка пацієнтів може бути охарактеризована як чудернацька, хворим притаманне усамітнення та уникнення близьких контактів із оточуючими. Також, у пацієнтів простежуються аномалії мислення, неадекватний емоційний стан, мовлення можна охарактеризувати, як дивне та нелогічне; стосовно продуктивної симптоматики – зустрічаються ілюзії, з'являється дереалізація та деперсоналізація, істерія, тощо. [60] Тож, у даному випадку захворювання за симптоматикою цілком нагадує одну із форм шизофренії, проте все ж відрізняючись деякими аспектами (недарма шизотиповий розлад особистості також називають «проста шизофренія»).

Маячний розлад

Маячний розлад також входить в одну групу із шизофренією та шизоподібними розладами згідно із міжнародною класифікацією хвороб 10-го перегляду (МКХ-10). На відмінну від шизофренії, при маячному розладі простежується виключно систематизоване маячіння, проте будь-які інші сильні психотичні прояви відсутні. [5] Власне даний аспект і відрізняє шизофренію та маячний розлад, попри те, що маячіння притаманне для обох діагнозів.

У ході нашого дослідження, ми намагаємося дослідити саме внутрішні залежності та структурні закономірності у мовленні пацієнтів. Власне, тому ми не так враховуємо саме тип захворювання, адже більшою мірою нас цікавить саме структурний аспект функціонування мовлення, при якому притаманне марення.

2.3. Мовлення людей із психічними вадами: характерні особливості (у контексті психотерапевтичної та психіатричної практик)

Лікарі, які безпосередньо працюють із психічно хворими – психіатри та психотерапевти, неодноразово наголошують на тих чи інших особливостях мовлення пацієнтів хворих на шизофренію. Складно не погодитися, та не відзначити особливості, які здатні виокремити мовлення пацієнтів, яким притаманні мовленнєві відхилення від мовлення «звичайних», або ж «здорових» людей.

Ми можемо перерахувати (попередньо) характерні особливості саме мовлення при психічних відхиленнях (шизофренія, шизотиповий розлад марення, маячний розлад та інші). Варто зауважити, що поведінка нас не цікавить значною мірою, пригадаємо лише побіжно.

Оскільки більшість пацієнтів мовлення яких ми розглядаємо мають діагноз шизофренія, у даному розділі найбільшою мірою буде представлена саме специфіка мовлення пацієнтів із шизофренією. Проте, також розглянемо і специфіку функціонуванням та організацією марення у залежності від захворювання.

Проблеми із асоціативним процесом пацієнтів

Перш за все, варто наголосити на специфіці узагальнення та побудови асоціативного ряду, при шизофренії. У пацієнтів узагальнення та асоціативний ряд формується за несуттєвими, або ж неочевидними ознаками. [9]

Однією із яскраво виражених та відразу помітних особливостей мовлення є **порушення асоціативного процесу** у мовленні хворих шизофренією. Загалом проблеми із АП мовця спостерігаються у при будь-яких проявах психічних хвороб (марення, галюцинації, тощо). [11] Тож, варто зауважити та відзначити важливі для нашого дослідження порушення (+ в подальшому місці та розвиток даних порушень у контексті аналізу дискурсу).

Одним із найбільш помітних порушень АП – пришвидшення асоціацій. [11] Як приклад можемо навести наступний фрагмент мовлення (узято із напрацьованої бази даних): «такое ощущение что я потеряю весь мир и всё \\\ ну как бы из под ног уйдет весь \\\ просто наверно не то что там любовь не любовь а просто как ощущение вот этого человека придает мне в жизни моей сил хотя люблю я абсолютно другого человека уже двадцать-пять лет \\\ [ФІО] там он в интернете есть все хочу что бы он был моим семейным доктором но никак не получается \\\ но ничего \\\ главное знаете что не заикливаться наверно да».

У пацієнта спостерігається значною мірою неможливість концентруватися на одній конкретній ідеї. Разом із тим, однією із особливостей є – постійна поява все нових ідей, які все більше стають обумовлені випадковими зв'язками. [11] І хоча нам складно не погодитись із висновками видатних психіатрів (Кербиков О.В., Коркина М.В., Наджаров Р.А., Снежневский А.В.), проте варто висунути припущення – що навіть попри певну раптову зміну ідей, як стверджують автори [11], мовленню притаманна певна структурність. Як приклад приведемо наступну фразу (пацієнт, чоловік, орієнтовно 30 років, шизофренія): «да вот потом я **вызываю** \\\ ну какое-то состояние такое знаете целый день вот вчера да как-то веселое но при этом сам с собой разговариваю ну так как будто мне

класно но при этом энергия сегодня кушать не хочется вот \\ я ходил ходил потом походил ... **вызываю** скорую приезжаю по диабете и тут же псих бригада привозит тут меня доктор ...». У даному випадку ми можемо простежити і пришвидшення асоціативного процесу, і розірваність мовлення, проте – дана фраза показує, що мовленню психічно хворих притаманна певна концептуальна цілісність.

Попри те, що одним із головних критеріїв порушення АП – постійна зміна теми мовлення (дуже часто викликана різними асоціативними тригерами – співзвучність слів, випадковий візуальний образ, тощо), як зазначено у [5], у нас є певні зауваження. Перш за все, це тематична цілісність, висловлення певної ідеї, чи групи ідеї (зазвичай тригер один, наприклад, параноїдальний синдром, при якому пацієнти відчують, що за ними стежать) впродовж всього монологу (якщо такий мається).

Зауважимо, також, що будь-яке стороннє запитання або ж спроба змінити тему матиме незначний вплив на пацієнта (-ку). Вони неодмінно «повернуться до наболілого».

Інша проблема та ускладнення асоціативних процесів при шизофренії – уповільнення АП. [12] У даному контексті хворий, або ж хвора скаржаться на погіршення розумових процесів, складність формулювання та вираження думок як таких. На даному етапі дослідження, ми не можемо стверджувати, що дане ускладнення АП є релевантним до нашого дослідження.

Надмірна обґрунтованість мисленнєвих процесів та персеверація також є характерною для порушення АП. Для надмірної обґрунтованості характерна «в'язкість мислення, зниження його продуктивності», а також – втрата можливості пріоритетизації (у мовленні це виражено доданням великої кількості деталей при розповіді). [11] Персеверація також виражається у мовленні довготривалим представленням певної синтаксичної структури.

Із точки зору психіатричної практики порушення АП може бути зумовлене певними фізіологічними процесами, які утворюються в головному мозку. Зокрема – посилення АП у лікарській практиці пояснюється підвищенням збудження кори головного мозку із одночасним послабленням активного умовного гальмування та генералізацією збудження. І навпаки – гальмування АП (персеверація, загальне уповільнення) є обумовленим зниженням функції кори головного мозку із процесами гальмування, які превалюють над процесами збудження. [11] Тож, ми здатні зробити висновок про відношення та залежність процесів ментальних від фізичних (очевидно). Хоча тут є і цікаві дослідження, які здатні нашоствхнути нас на цікаві (як думається на даний момент) роздуми, а саме – синдром Кандинського-Клерамбо, або ж синдром психічних автоматизмів.

Маячіння

Як стверджує Коркіна [9], саме маячення є одним із найбільш яскраво виражених розладів мислення. За стандартним формулюванням, маячіння – це неправильно виведенні судження та заключення, які формують його «реальність». *Тобто, при маячінню хворі, немов потрапляють в певну фіктивну реальність, в межах якої і перебуває суб'єкт.*

Важливим є також і наголосити на відмінності маячіння від звичайних помилкових, або ж необґрунтованих суджень. Перш за все, маячіння може виникати лише у випадку прогресуючого психічного розладу (психотична структура, див вище). Також, маячіння – один із ключових симптомів шизофренії та шизоподібного розладів (які власне нас і цікавлять в ході даного дослідження).

При маячінні, пацієнт повною мірою впевнений у правоті своїх суджень. [9] Варто наголосити – власне марення формують «реальність» пацієнта. Тому будь-які аргументи та запевнення його, або ж її у неправильності сказаного, або

ж написаного – марні. Більше того, вони здатні викликати агресивну та невідконтрольну ситуацію.

Маячіння також неможливо спростувати для пацієнта (лише в окремих випадках). [9] Даний феномен може бути зумовленим тим, що маячіння і формують уявлення про реальність та свідомість хворого на шизофренію (у нашому випадку ми розглядаємо саме шизофренію).

Тим не менш деякі сучасні психіатри, стверджують, що маячіння – «цілковита втрата індивіда можливості міркувати або логічно мислити». [12]

Проте чи було б справедливим стверджувати, що маячіння є по своїй структурі нелогічним, які не несуть жодної фактичної інформації? Якщо із першим судженням (що стосується семантичної цінності та змістовності мовлення хворих на шизофренію), ми і могли б погодитись. Із другим судженням, нам і належить працювати в ході даної роботи, а певніш – його спростуванням або ж навпаки.

2.4 Лінгвістичні дослідження орієнтовані на мовлення людей із психічними розладами

Існує велика кількість досліджень, які стосуються специфіки функціонування та впорядкування мовлення людей із психічними вадами. Дані дослідження втілені як у межах психіатричної практики, так і в межах лінгвістичної, що лише підтверджує згадані аспекти актуальності нашого дослідження.

Одні із ключових та одних із найбільш корисних робіт (в тому числі із якими ми працювали) є все ж психіатричні роботи таких видатних психіатрів (представників «старої школи»), як Коркіна, Снежневський, Жмуров, Канабих та інші.

У даних роботах (попри те, що це суто психіатричні навчальні посібники) ми змогли знайти чималу кількість корисної інформації та пояснень, щодо функціональних особливостей мовлення людей із психічними вадами (у нашому

випадку: шизофренія, шизоподібний розлад, та інші). Чудовим прикладом розбору специфіки марення є робота Снежевського «Маячіння» [11], де детально та змістовно досліджено специфіка мовців (люди із психічними вадами) при маячінні. Головні концепти із даних робіт ми вже пригадували у підпункті 1.2 (див. вище).

В достатній мірі аргументовані та досліджені певні особливості функціонування саме мовлення людей із вище згаданими психічними вадами. Наприклад, змістовна та ґрунтовна робота психотерапевта Януша Вроблеля «Мова та шизофренія» [58], де автор розглядає процеси при шизофренії у їх залежності із *мовленнєвим актом*. Наприклад, зв'язок між конотацією та денотацією у мовленні та мовленнєвий акт при шизофренії, що значною мірою поглиблює наше уявлення про специфіку функціонування мовлення.

Проте, однаковою мірою корисними були і роботи в межах лінгвістики (комп'ютерної, зокрема), науки про дані та комп'ютерної науки, які покликані дослідити специфіку організації та функціонування мовленнєвого матеріалу. Наприклад, ми опрацювали такі сучасні роботи та дослідження як «Шизофренія та структура мови: лінгвістичний аспект» [29] та можемо стверджувати, що лінгвістичний аспект дослідження представлений якісно.

Згадані вище роботи є лише прикладами досліджень, які стосуються мовлення людей із психічними вадами, у різних сферах. Кожна із сфер може дійсно бути корисною для подальшої роботи та обробки мовленнєвого матеріалу мовців згаданих вище.

Висновки до другого розділу

Другий розділ має на меті представити та детально описати теоретичні ключові аспекти, які стосуються дослідження мовлення людей із психічними вадами. Ми представляємо поняття психотичної структури психічного апарату людини (яку нам і було б цікаво дослідити) та увиразнюємо особливості функціонування мовлення при ментальних хворобах.

У першому підпункті даного розділу ми також представили ключові аспекти, які є критично важливими для нашої роботи – поняття психотичної структури.

Ми увиразнили предмету дослідження, яким є мовлення людей із психічними вадами. Тут ми чітко сформулювали та увиразнили особливості тих категорій мовців із психічними вадами із мовленням яких ми зацікавлені працювати. Дані розлади описані у підпункті 2.2.

Саме в межах даної частини ми також сформулювали певні аспекти мовлення, які згадані у психіатричних роботах (сповільнення асоціативного процесу, специфіка марення та його функціонування).

Дані аспекти ми досліджували не лише в роботах написаних в науковому психіатричному дискурсі, а і у роботах психоаналітиків, лінгвістів, спеціалістів у комп'ютерних науках, які користуючись понятійним апаратом притаманному дискурсу кожній із сфер змогли лише увиразнити ті чи інші аспекти мовлення людей із психічними вадами (згаданими у підпункті 2.1). Ті та інші аспекти формулювання мовлення представлені у великій кількості робіт, лише деякі із яких згадані у підпункті 2.3.

РОЗДІЛ 3. ПОПЕРЕДНЯ ОБРОБКА ТА ПІДГОТОВКА МОВЛЕННЄВОГО МАТЕРІАЛУ

У третьому розділі нашої роботи ми представляємо власне практичний процес збору, аналізу та обробки даних. Нижче ми детально описали кожен із етапів та увиразнили його в межах нашого дослідження.

3.1. Створення корпусу для подальшої обробки даних

Сучасна лінгвістика маючи чималий ресурс для досліджень (квалітативний, квантитативний аналіз тексту, тощо) часто користується корпусом текстів, маючи на меті дослідити ті, чи інші закономірності у мові, мовлені. Зокрема, корпус текстів є важливим та корисним інструментом для дескриптивного дослідження використання мови як такої, а також – специфіку організації лексичних, граматичних, синтаксичних рівнів у мові та мовлені.

Нижче ми детально описуємо процес збору корпусу текстів, який надалі будемо опрацьовувати.

3.1.1. Збір та обробка мовленнєвого матеріалу

У ході нашого дослідження, ми маємо на меті дослідити структурні закономірності у мовленні із психічними відхиленнями (МХБ-10: шизофренія, шизоподібний розлад, тощо). Для цього створено 2 корпуси: корпус із записами людей із психічними розладами згаданими вище та референтний корпус із записами мовлення людей без психічних розладів (принаймні видимих, які б можна було класифікувати як марення).

До того, як починати обробку даних та роботу із даними як таку, нам потрібно перш за все вирішити, яким чином ми будемо структурувати та впорядковувати дату (текстову інформацію), яку і збираємось аналізувати. Для цього найзручнішим способом було б саме зібрати корпус текстів (який у нашому випадку має складатися із двох споріднених текстів – анотовані записи спонтанного мовлення двох категорій людей згаданих вище).

Корпус текстів – це будь-яка колекція зібраних текстів (написаних або ж промовлених), які є впорядковані (та складені) за допомогою програми та можуть бути проанотовані (тобто на яких є можливим простежити та відмітити певну лінгвістичну інформацію). [46] Саме це ми і потребує зробити перш за все – зібрати усі наші дані, із якими будемо надалі працювати, як певну доступну базу інформації.

Загалом обидва корпуси складаються із проанотованих аудіозаписів. Кількість записів для формування першого корпусу – 14 записів, кожен із яких може різнитися за якістю, довжиною, тощо. Основний корпус ми збирали із записів пацієнтів однієї із психіатричних лікарень Києва (назву лікарні та імена учасників воліємо не розголошувати із етичних міркувань). Референтний корпус складається переважно із проанотованих записів спонтанного (!) мовлення дикторів, журналістів, науковців, мовлення яких відзначається «чистотою» та структуризацією. Адже, нам здається, що таким чином було б найкраще простежити ту чи іншу (разючу) відмінність між мовленням даних різних мовленнєвих груп.

Перший корпус складається із 14 проанотованих записів та 13333 слів, або ж 59 137 символів. Можемо, уважати, що це орієнтовно 50 сторінок (якщо уважати, що 1 сторінка дорівнює 250 слів).

Другий корпус, на відмінну від першого, який є більш масивний, що стосується кількості записів, складається із всього лише 7 проанотованих записів. Проте, переважну більшість (6 записів) складають анотовані записи мовлення освічених людей, спеціалістів, мовлення яких (на нашу суб'єктивну думку) можна уважати як зв'язне та гарно структуроване. Тож, другий корпус складається із всього лише 10 тисяч слів (або ж 43 сторінки)), та 54 107 (!) символів.

Хочемо зауважити, що кількісний аспект у нашому випадку є надзвичайно цікавим. Адже попри невелику (відносну) кількість записів у другому корпусі,

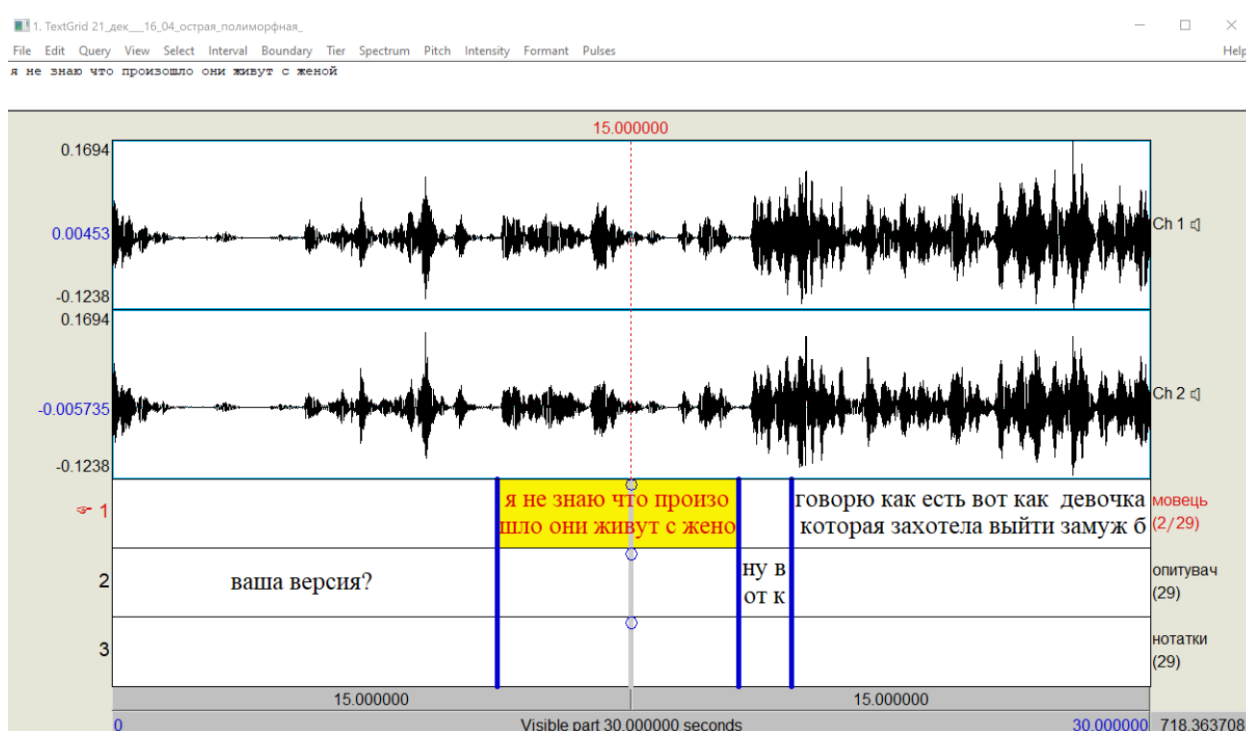
розміри другого і першого корпусів є майже однаковими. Даний аспект може свідчити не лише про наповнюваність повідомлення та його змістовність (у випадку мовців, мовлення яких складає другий корпус), а і про велику кількість слів, які мають мало значення (слова вставки, тощо) у випадку мовців у першому корпусі.

Попередньо, ми хотіли б зауважити та зробити висновок, що це може свідчити про структурну цілісність та «якість» повідомлення, яке притаманне для мовців із другого корпусу. І навпаки – про мовленнєву розгалуженість та «розбитість» мовлення у першому.

3.1.2 Формування корпусів

Для роботи із записами як такими нам потрібно проанотувати кожен із зібраних записів. Подальша обробка записів здійснюється у програмі Praat, за допомогою якої ми можемо розділити кожен із текстів на фрази, позначити наявність шуму, сторонніх звуків, тощо.

Рис. 2:



На рисунку 2 (див. вище) ми можемо простежити безпосередньо процес анотації запису. Для обробки кожного із файлів, ми відкриваємо кожен аудіо файл у програмі Praat. Потім, варто вказати рівні, згідно із якими будемо розподіляти текст (у програмі це називається «annotate layer») та прописуємо кількість потрібних нам рівнів анотації, а саме – опитувач, опитаний та нотатки.

Усі дані є збереженими у форматі .txt. Власне, кожен із наших корпусів (основний і референтний) складається із даних .txt файлів, які згодом будуть об'єднані в корпус завдяки бібліотеці NLTK у Python.

На сьогодні, велика кількість досліджень у галузі мовознавства здійснюється на базисі корпусу (зібрані попередньо дослідником (-цею) тексти та впорядковані за допомогою програми, або ж вже доступні корпуси мов). Загалом для роботи із великою корпусом та великою кількістю текстової інформації як такої можна використовувати такі ресурси як SketchEngine, або ж LangVox4.0 (від Lancaster University).

Враховуючи той факт, що обробку даних ми будемо здійснювати за допомогою мови програмування Python, найзручнішим способом зібрати корпус в межах нашого дослідження було б саме сформувати потрібну базу текстів безпосередньо у програмі. Таким чином ми зможемо працювати із текстом не відволікаючись та не витрачаючи зайвий час на те, щоб вирішувати будь-які ймовірні складнощі із файлами, доступом до них, тощо із іншої бази даних у нашу Python програму.

Процес створення корпусу текстів за допомогою Python NLP бібліотеки NLTK, виглядає наступним чином:

```
import os, os.path

import nltk.data
```

(1) Завантажуємо потрібні нам бібліотеки для роботи із текстом

```
path = os.path.expanduser('~/.nltk_data')
```

(2) Створюємо шлях до теки, де буде безпосередньо зберігатися корпус.

```
if not os.path.exists(path):  
    os.mkdir(path)  
    print("created")  
else:  
    print("already exists")
```

(3) Перевіряємо, чи файл було створено, або ж було використано попередньо створений файл.

```
newfile = nltk.data.load('corpora/patients/patient1.txt', format='text',  
encoding='utf-16')  
print(newfile)
```

(4) У теці, де буде знаходитися безпосередньо корпус текстів, створюємо додаткову теку “corpora”, де будуть знаходитися наші два корпуси текстів (один – корпус текстів мовців 1, другий – корпус текстів мовців 2). Це можна було б зробити, як за допомогою способу, який був використаний на початку створення корпусу для створення теки «nltk_data», або ж створити особисто потрібні теки. Якщо виникає будь-яка проблема із пошуком теки безпосередньо, радимо скористатися командою `os.path.abspath(path)`, для пошуку теки «nltk_data».

(5) За допомогою команди `nltk.data.load()` завантажуюмо потрібний текст. Також тут варто вказати, що інформація із файлів буде завантажуватися як байт-код (а не текст написаний кирилицею, латиницею). Це не має жодних проблем для нас, адже достатньо лише вказати `format='text'` і ми отримаємо знову придатний до читання текст.

Варто зауважити, що одним із аспектів даного етапу роботи, який значною мірою сповільнює процес роботи над обробкою записів (та має бути оптимізованим, наприклад, за допомогою системи розпізнавання голосу, тощо) є

саме анотація текстів власноруч. Проте, ми намагалися мінімізувати наявність будь-яких друкарських помилок, за допомогою уважного та ретельного вичитування тексту.

Код процесу формування корпусів можна знайти в Додаток 3.

3.2 Підготовка корпусів для подальшої обробки

У межах даного дослідження, ми збираємося опрацьовувати всю інформацію за допомогою мови програмування Python та в середовищі програмування, яке підтримує дану мову програмування. Перш за все відзначимо, що дана мова є ООП (об'єктно орієнтоване програмування).

ООП є одним із парадигм програмування, для якого програма складається із множини «об'єктів» (класи), які взаємодіючи між собою реалізують задані завдання. Даний вид програмування складається із чотирьох основних концептів, а саме – *інкапсуляція, успадкування, поліморфізм, абстракція*. [tag a resource] Більш детально про кожен із видів та його застосування (важливість) для нашого дослідження буде йтися нижче.

Також в мові програмування Python є наступні типи даних, які варто враховувати при обробці та збереженні даних та будь-якої інформації:

- str (імена, текст);
- int (кількісні показники, числа);
- float (показники, нумерація);
- bool (бінарна інформація, так \ ні питання);
- datetime (дати);
- category.

Наступним кроком із обробки даних текстів має бути чистка файлів від будь-якої зайвої інформації, яка може нам лише завадити у ході дослідження. Зайву інформацію ми знаходимо та видаляємо за допомогою кодування regex прописаного у одному із циклів програми.

Рис. 3:

```
class RetrievingData:

    def getting_word(self, text):
        cleaned = re.findall("\((.\n)*\)", text)
        cleaned = re.sub("\(d+[.]d+)\|([A-Za-z]+)\|(пацієнт)", "", text)
        cleaned1 = cleaned.translate({ord(i): None for i in "\"\" and \"\n\"}).split(" ")

        finalList = []
        for word in cleaned1:
            if word.isalpha():
                finalList.append(word)
        return finalList

    def stop_words(self, getting_word):
        pass
```

На рис. 3 зображено увесь процес чистки даних. Ми не бачимо потреби пояснювати кожен із кроків. Проте наголосимо радше безпосередньо на частині, яка відповідає прибиранню зайвої інформації із файлу. Перш за все, ми завантажуюємо усі дані в межах лапок, адже під час завантаження текстів із Praat проанотована інформація зосереджена саме в лапках:

```
cleaned = re.findall("\((.\n)*\)", text),
```

де `re.findall` – команда, яка знаходить та завантажує ту інформацію, яка збіглася із вказаним шаблоном; `\((.\n)*\)`, `text` – безпосередньо регулярний вираз, який знаходить потрібну текстову інформацію.

Проте, на жаль, під час використання даного виразу ми не змогли уникнути непотрібної інформації, а саме – показники часового проміжку відтворення фрази, ідентифікуючих слів таких як «опитувач», «пацієнт», «інше», тощо. За допомогою наступного рядка ми змогли прибрати будь-яку зайву інформацію:

```
cleaned = re.sub("\(d+[.]d+)\|([A-Za-z]+)\|(пацієнт)", "", text),
```

де `re.sub` – команда, яка вказує на те, що у `text`, потрібно знайти показники, які будуть співпадати із регулярним виразом вказаним на початку у дужках та замінити його на вираз, який вказано другим у дужках (у нашому випадку – порожнеча); `\(d+[.]d+)\|([A-Za-z]+)\|(пацієнт)` – безпосередньо регулярний

вираз, який має знайти потрібну інформацію (у нашому випадку зайву); "" – інформація знайдена за допомогою регулярного виразу має бути замінена на показник вказаний у лапках; text – вказані зміни та маніпуляцію мають бути здійснені на вказаному файлі.

На даному етапі, також варто прибрати стоп-слова (stop-words), які можуть вплинути на остаточні показники. Це буде здійснено для більш точного обрахунку показника TF-IDF, який буде використано в подальшому обрахунку ключових слів.

Нормалізація тексту

Під час роботи із великою кількістю текстового матеріалу, нам потрібно розуміти, та передбачувати певні мовленнєві аспекти, які можуть стати нам на заваді під час подальших обрахунків. Одним із цих таких аспектів є структурні особливості слов'янських мов, із якими ми і будемо працювати.

Слов'янські мови входять в групу синтетичних мов, які поділяються на флективні та аглютинативні мови. Враховуючи даний аспект, ми б хотіли наголосити на специфіці словотворення, а саме – у всіх індоєвропейських та семіто-хатських мовах дуже важливе значення при вираженні граматичних значень відіграє саме флексія, або ж словозміна [10].

Словозміна відображає системний аспект творення того, чи іншого слова відповідного до його синтаксичного зв'язку із іншими частинами речення, фрази, словосполучення, тощо. Це може ускладнити наше подальше дослідження, тож варто вирішити, яким чином ми могли б оптимізувати даний аспект роботи.

Вирішити даний аспект ми зможемо за допомогою опрацювання кожного токена та виокремлення його кореня. Таким чином ми зможемо надалі не лише обраховувати кількісний та якісний показник кожного токена. Ми вважаємо, що саме завдяки нормалізації тексту як такого, ми і зможемо обраховувати та вирізняти еквівалентний клас у мовленні як такий (а також взаємозалежності у двох окремих груп).

Вирізнити та вирахувати даний аспект, ми можемо за допомогою ідентифікації та обрахунку кореня кожного слова. Для обрахунку даного аспекту ми звернемося до аспекту ідентифікації кореня, або ж леми.

У NLP, для цього ми користуємося методами, які називаються стематизація (stemming) та лематизація (lemmatization). Обидва ці способи можуть допомогти отримати та вирахувати корінь слова для подальшої роботи із ним. Проте, не варто вважати, що стемінг та лематизація є тотожними та нам варто очікувати однакових результатів. Адже кожен із способів може дати нам різну інформацію.

Надалі ми більш детально розглянемо та аргументуємо доцільність використання кожного із них, оберемо, той метод, який буде використовуватися в межах нашого дослідження.

Стематизація

Один із ключових аспектів роботи із текстом, який написаний однією із мов згаданих вище – виокремлення кореня. У NLP, даний процес називають стематизація. За допомогою стематизації, ми можемо отримати виключно корінь кожного слова.

Хід стематизації, яку ми використовували під час даного дослідження, описано нижче:

```
ourStemmer = SnowballStemmer()
```

(1) Тут ми користуємось бібліотекою SnowballStemmer, яку завантажуюмо окремо за допомогою рядка `from nltk.stem.snowball import SnowballStemmer`. Вже надалі в дужках `SnowballStemmer()`, ми вписуємо ту мову, яка буде використовуватися під час стематизації.

```
list_with_stems = []
```

```
for i in classy.getting_word(listToString):
```

```
list_with_stems.append(ourStemmer.stem(i))

return list_with_stems
```

(2) Остання частина програми власне і стематизує потрібний нам список (тут: `listToString`). Тут кожен елемент (`i`) в списку попередньо обробленому та «прибраному», ми додаємо у бібліотеку (type: dictionary) `list_with_stems` за допомогою `.append` способу. Кожен із цих елементів ми попередньо стематизуємо, або ж прибираємо закінчення за допомогою способу `ourStemmer.stem(i)`. Вкінці ми повертаємо вже заповнений список `list_with_stems`.

Лематизація

Як було згадано вище, лематизація та стематизація значною мірою відрізняються саме підходом, який використовується для вирахування кореня кожної словоформи. [64]

Варто зауважити, що процес стематизації не задовольняє достатньо суттєвий аспект, який стосується «повернення» слова до його вихідної форми. Під час обробки таких слів як «хочу» та «хотіти», ми отримуємо різні результати під час стематизації, а саме – «хоч» та «хот» відповідно. Це зумовлено тим, що алгоритм `SnowballStemmer` не повертає слово до його перш початкової форми під час обробки, тож ми не отримуємо якісний підрахунок кількості використання того чи іншого слова (дієслова особливо) у тексті. Натомість, ми отримуємо два різні результати для кожного із слів.

Ми можемо урахувати даний аспект та використати лематизацію нашої текстової інформації, а не стематизацію. Під час лематизації наша програма буде повертати кожне із слів до основної форми, тож нам буде набагато простіше забезпечити подальшу обробку тексту.

Ми можемо спробувати задовільнити даний аспект обрахувавши леми за допомогою алгоритму прописаного нижче:

```
import pymorphy2
```

```
morph = pymorphy2.MorphAnalyzer()
```

(1) Спочатку ми завантажуємо бібліотеку `pymorphy2` за допомогою команди `import`. Завдяки даній бібліотеці ми зможемо визначити перш початкову форму кожного слова у списку. Наступний крок – завантажуємо аналізатор морфем (`pymorphy2.MorphAnalyzer()`) та асоціюємо його із змінною `morph` для подальшого використання даної функції.

```
list_with_lemmas = []
```

(2) Створюємо порожній список `list_with_lemmas`, в який ми і будемо додавати результати лематизації.

```
for p in classy.getting_word(listToString):
```

```
    list_with_lemmas.append(morph.parse(p)[0].normal_form)
```

```
return list_with_lemmas
```

(3) Процес лематизації відбуватиметься за допомогою виразу згаданого вище. Тут власне кожне слово (позначене, як «р») у списку `listToString`, буде аналізуватися відповідно до запису «`morph.parse(p)[0].normal_form`» та додаватися у список `list_with_lemmas`.

Враховуючи той, факт, що стематизація лише залишає корінь кожного слова, а не дає нам його перш початкову форму, ми схильні вважати, що саме завдяки лематизації буде можливим досягти кращих результатів обрахунку та аналізу нашого корпусу та записів, які його складають.

Висновки до третього розділу

У третьому розділі дослідження ми детально описали процес організації та формування корпусів (основний та референтний корпуси). Основний корпус складається із анотованих записів мовців із психічними вадами. Матеріал був

оброблений нами власноруч та сформований у корпус за допомогою мови програмування Python.

У підпункті 3.2 детально описаний процес безпосередньо підготовки даних для подальшої їх обробки: чистка даних та оформлення їх (лематизація) для подальшої обробки.

Варто наголосити, що для подальшої роботи ми обрали саме метод лематизації, а не стематизації.

На даний момент можемо із впевненістю зазначити, що мовленнєві дані ретельно оформлені та готові для подальшої їх обробки.

РОЗДІЛ 4. ОБРОБКА ЗІБРАНОГО МОВЛЕННЄВОГО МАТЕРІАЛУ

Завдяки роботі із великою кількістю зібраного матеріалу, ми здатні відстежити ті чи інші закономірності та внутрішні залежності у тексті (тут: анотовані корпуси текстів) та певні структурні відмінності\схожості між ними. Нижче описаний процес аналізу зібраного мовленнєвого матеріалу

4.1 Ключові поняття та їх місце у структуризації мовлення

Мета нашого дослідження простежити структурні особливості мовленнєвого матеріалу зібраного та обробленого у ході нашого дослідження за допомогою методів дистрибутивної семантики. Ми маємо на меті проаналізувати структуру висловлювання за допомогою увиразнення ключових слів у мовленні.

Ключові слова ми маємо на меті виявити за допомогою *латентного семантичного аналізу (ЛСА)*, як одного із методів, який може бути використаний для репрезентації та здобуття значення мовленнєвої одиниці у певному контексті (який представлений великим корпусом текстів). [39]

Варто також зауважити, що латентний семантичний аналіз є одним із методів обробки, який був сформульований саме на базисі дистрибутивної семантики.

Головний принцип роботи методу латентного семантичного аналізу є дослідити контекст вживання слова та встановити набір внутрішньо-зумовлених обмежень, за допомогою яких стане можливим встановити схожість у значеннях слів та словосполучень один із одним. За допомогою даного методу є можливим встановити такі залежності як когерентність тексту та\або дискурсу.

Латентний семантичний аналіз тексту може бути представленим двома способами [39]:

1. як спосіб визначити приблизну оцінку замінності слів у контексті використання у більших сегментах тексту;

2. як модель комп'ютерних процесів та репрезентацій, які лежать в основі значних частин набуття та використання знань.

Щоб зібрати дійсно якісні показники ключових слів для кожного із корпусів, ми обраховували їх без урахування стоп-слів. Для цього ми попередньо очистили кожен корпус за допомогою наступного коду програми:

```
stop_words = set(stopwords.words('russian'))
```

- (1) визначаємо бібліотеку стоп-слів.

```
filtered_sent = [w for w in word_tokens if not w in stop_words]
```

- (2) видаляємо усі стоп-слова для кожного із корпусів та зберігаємо результати у змінній `corpora_with_no_stopwords` (основний корпус без ключових слів) та змінній `corpora2_no_stopwords` (референтний корпус без ключових слів).

ЛСА модель представлена за допомогою токенів, які представлені у вигляді векторів та обрахунку їх TF-IDF значень, як указано на рис. 4. Саме за допомогою моделі описаної нижче ми і будемо виявляти ключові поняття для кожного із корпусів.

Одним із основних та початкових етапів латентного семантичного аналізу (для виявлення ключових понять) є обрахунок показника TF-IDF (eng. TF – term frequency, IDF – inverse document frequency).

За допомогою обрахунку TF-IDF ми здатні оцінити «важливість» слова (або ж, як у нашому випадку, токену) не за його квантитативними показниками (наприклад, завдяки Counter, обрахувати частоту використання кожного токену), а саме за його важливістю у межах усього документа як такого.

Таким чином ми будемо аналізувати кожен із наших текстів\документів (із яких і складається загальний корпус). Даний підхід є одним із перших етапів (та одним із найважливіших). [41]

Для обчислення TF-IDF нам потрібно помножити показник TF на показник IDF:

$$\text{tfidf}_{t,d} = \text{tf}_{t,d} * \text{idf}_t,$$

де TF (eng. term frequency) означає співвідношення входження певного слова (токена) та загальної кількості слів (не токенів) у конкретному документі, а IDF (en. inverse document frequency) використовується для обчислення якісного показника (конкретної одиниці).

Показник TF-IDF обчислюємо для кожного із корпусів окремо. Таким чином ми зможемо виразити якісний (а не кількісний) показник кожного із ключових слів.

На сьогодні наявні різні бібліотеки, за допомогою яких можливо розрахувати даний показник у Python (skit.learn, Tensor Flow). Ми будемо користуватися бібліотеками skit.learn.

Для обчислення показника TF-IDF за допомогою даної бібліотеки користуються двома способами TfidfTransformer та TfidfVectorizer. Загалом кожен із даних способів незначною мірою відрізняється, хоча результати обчислень однакові.

Для нашого дослідження, ми вирішили вирахувати TF-IDF за допомогою TfidfVectorizer.

Рис. 4:

```
vectorizer = TfidfVectorizer(use_idf=True, ngram_range=(1, 3))
X = vectorizer.fit_transform(text_no_stopwords)
```

Далі отримавши якісні показники значень, ми обчислюємо показник за допомогою зменшення лінійної розмірності через розрізання усіченої особливої величини (SVD). Процес обчислення зображено на рис. 5.

Рис. 5:

```
lsa = TruncatedSVD(n_components=21, n_iter=100)
```

За допомогою даного методу ми обрахували ключові поняття для текстів, які складають основний та референтний корпуси. Варто зауважити, що отримані показники вдало та якісно здатні передати ключові поняття кожного із текстів. Саме тому можемо уважати, що обрахунок було здійснено успішно.

У *четвертому* та *п'ятому* додатках ми представляємо список (скорочений) ключових слів для основного та референтного корпусів відповідно.

4.2 Візуалізація результатів

Базовим способом обробки є зібрати дистрибутивну інформацію у вигляді векторів та вирахувати дистрибутивну, або ж семантичну схожість показників векторних залежностей. [51] Ми здійснили даний обрахунок, який було описано вище.

Завдяки цьому етапу у нас є можливість визначити кількісний показник ключових слів у потрібному для нас тексті та спробувати простежити ймовірні закономірності (для текстів, які складають перший та другий корпуси).

Надалі для обрахунку та представлення даних ключових понять у наших текстах, ми замінимо кожне із ключових слів у тексті, щоб простежити ймовірні закономірності, на значення «KEYWORD».

Для візуалізації ключових слів та задля того, щоб простежити специфіку реалізації даних ключових слів, ми обрали метод хмари слів, де зібрані не лише ключові поняття, а і решта часто вживаних слів. Даний метод є чудовим способом для візуалізації вживаності токена у тексті.

Повною мірою даний метод описано на рис. 6. Лише прокоментуємо ключові аспекти обрахунку хмарини слів, а саме – для обрахунку використовуємо Python бібліотеки `wordcloud` та `matplotlib.pyplot`. Далі залишається лише увести потрібні показники та обрахувати показник для кожного із текстів.

Усі обрахунки та хмарки слів для кожного із текстів залишаємо в Додаток 1 (основний корпус, мовці із психічними вадами) та Додаток 2 (референтний корпус, мовці без психічних вад).

Рис. 6:

```
#wordcloud
from wordcloud import WordCloud
import matplotlib.pyplot as plt

cloud = WordCloud(width=1000, height=500).generate(' '.join(equiv))
plt.figure(figsize=(15, 8))
plt.imshow(cloud)
plt.axis('off')
plt.savefig('non-patient7' + '.png', bbox_inches='tight')
plt.show()
plt.close()
```

В ході даної візуалізації ми отримали достатньо цікаві результати, а саме – *мовленнєвий акт кожного із мовців представлений за допомогою зв'язних (більшою або ж меншою мірою) ключових слів та сполучників (до списку сполучників у нашому випадку можна додати слова-філери).*

Даний аспект є ключовим при формуванні мовленнєвого акту як для людей із психічними вадами (тут: шизофренія) та людей без психічних (виразних) вад. На прикладі результатів даного дослідження ми бачимо, що і повідомлення мовців із психічними вадами, і мовленнєвий акт мовців без вад, складається із набору ключових понять, які більшою або ж меншою мірою взаємопов'язані під час акту говоріння. Тобто, даний аспект є спільним для вказаних груп мовців.

Структурною закономірністю мовлення людей із психічними вадами вважаємо саме вказану понятійну зв'язність. Попри велику кількість вад мовлення, які зумовлені фізичним та психологічним станом здоров'я людини, аспект функціонування та взаємозалежність ключових слів та тригерів, які актуалізуються у мовленні, є незмінним.

Можемо лише припустити, що саме від того наскільки вдало ключові слова та поняття поєднані у дискурсі, буде залежати якість повідомлення як такого. Проте, це могло б стати темою окремого дослідження. Дану думку розвивати вважаємо недоречним.

Повністю код програми можна знайти за посиланням: <https://github.com/bonchardon/thesis>.

Висновки до четвертого розділу

Четвертий розділ складається із двох основних частин – частини, де описано пошук ключових слів та методи використані із цією метою та частини, де описано, яким чином ми збираємося візуалізувати ті закономірності, які можна було б виявити за допомогою даного методу.

Ми використали латентний семантичний аналіз для обрахунку ключових понять в кожному у кожному із корпусів. Латентний семантичний аналіз використовується для репрезентації та здобуття значення мовленнєвої одиниці у певному контексті (який представлений великим корпусом текстів). Даний метод існує на базисі саме дистрибутивної семантики та відповідного уявлення про мовленнєву одиницю.

Даний метод зміг успішно представити усі ключові поняття та головуючі ідеї для кожного із текстів (у кожному корпусі).

Варто зауважити, що попри складність сприйняття мовленнєвого матеріалу мовців із психічними вадами та неможливість (у більшості випадків) увиразнення головної теми, через незв'язність та неспроможність мовців вибудувати зв'язну лінію оповіді, за допомогою латентного семантичного аналізу ви все ж змогли дійсно зібрати ключові аспекти дискурсу притаманні мовленню у зв'язний список ключових слів.

Згаданий аспект наштовхує на роздуми, стосовно специфіки формування мовленнєвого акту у дискурсі окремої людини. Наше припущення – мовлення і

складається із ключових понять взаємно-пов'язаних під час акту говоріння. На прикладі результатів візуалізації кожного мовленнєвого матеріалу, які складають два корпуси (основний та референтний), ми бачимо, що незалежно від якості повідомлення та здатності мовця його вербалізувати, кожен із текстів складається із взаємозалежних тем та ключових понять. Адже кожна візуалізація (див. Додаток 1 та Додаток 2) (а від того кожен із прикладів мовленнєвого матеріалу, які складають основний та референтний корпуси) складається із ключових понять та сполучників, або ж не суттєвих лексичних одиниць (слова-філери, обмовки, тощо).

ВИСНОВКИ

Наша робота покликана дослідити структурні особливості мовлення людей із психічними вадами. Для цього ми зібрали корпус текстів, який складається із мовленнєвого матеріалу людей із психічними хворобами (шизофренія, шизоподібний розлад). Також, задля увиразнення результатів, нами був зібраний референтний корпус текстів із мовленнєвим матеріалом людей без психічних вад.

Для подальшої обробки корпусів ми використовували об'єктно орієнтовану мову програмування Python. Процес безпосередньо обробки поділяється на два етапи – попередня обробка та безпосередньо робота та аналіз текстів.

Варто зауважити важливість попередньої обробки та підготовки тексту, адже від якості мовленнєвого матеріалу і залежать подальші результати дослідження. Враховуючи той факт, що нам вдалося вирахувати достатньо цікаві закономірності, будемо уважати, що якість даного етапу можна уважати прийнятною.

Дане дослідження здійснюємо в межах парадигми дистрибутивної семантики та користуючися методикою сформульованою в межах дистрибутивного аналізу тексту.

Одним із способів для репрезентації та здобуття значення мовленнєвої одиниці у певному контексті (який функціонує згідно із засадами дистрибутивної семантики) є метод латентного семантичного аналізу. Саме завдяки даному методу ми сформулювали та увиразнили головуючі поняття та ключові слова притаманні для текстів.

Зауважимо на специфічних особливостях мовленнєвого матеріалу людей із психічними вадами, а саме неможливість (у більшості випадків) увиразнення головної теми через неспроможність мовців вибудувати когерентну лінію

оповіді. Мовленню також притаманне пришвидшення \ сповільнення асоціативних процесів, тощо.

За допомогою латентного семантичного аналізу ми все ж змогли дійсно зібрати ключові аспекти (слова) дискурсу у зв'язний список ключових слів. Скорочений список представлений у четвертому та п'ятому додатках.

Ми стверджуємо, що мовлення людей із психічними вадами складається із ключових понять взаємопов'язаних під час акту говоріння. Адже на прикладі результатів візуалізації кожного мовленнєвого матеріалу, які складають два корпуси (основний та референтний), можна стверджувати, що ключові слова найбільшою мірою реалізовані у мовленні.

Кожна візуалізація (а від того кожен із прикладів мовленнєвого матеріалу, які складають основний та референтний корпуси) складається із ключових понять та сполучників, або ж не суттєвих лексичних одиниць (слова-філери, обмовки, тощо).

Тут ми хотіли б наголосити саме на специфіці влаштування дискурсу і у випадку першого корпусу, і у випадку другого – специфіка функціонування є надзвичайно схожою.

Враховуючи даний аспект ми і хотіли б зробити висновок, що попри контекстуальну та структурну (у більшості випадків) відмінність мовлення при психічних розладах та мовлення, якому притаманна алогічність та проблеми із асоціативним зв'язком (більш детально даний аспект розглядаємо у другому розділі), мовлення даних двох груп мовців формується шляхом злиття ключових понять під час акту говоріння.

Варто також зауважити, що дане дослідження наштовхнуло нас на деякі роздуми, які ми не мали змоги реалізувати. Ми вважаємо, що є сенс розглянути нижче вказані аспекти у наступних дослідженнях на дану тему (або ж у ході продовження даної роботи, або ж в контексті робіт інших дослідників та студентів).

Для нас було б неймовірно цікавим задіяти способи обрахунку в межах теорії риторичної структури [44], яка передбачає розділення мовленнєвого повідомлення на нуклеус та сателіти (або ж важливе \ не важливе), для аналізу мовлення людей із психічними хворобами. Справедливо уважати, що дані методи могли б дати чималу кількість цікавих результатів та допомогти нам простежити зв'язність представлену контекстуально.

Враховуючи доцільність та якість результатів дослідження текстових даних за допомогою методу векторизації, ми б все ж хотіли спробувати обрахувати аспект аномалії за допомогою алгоритму IsolationForest, який використовують для визначення аномалій у наборі тексту представленого у векторному вигляді.

Також, ми б хотіли розглянути мовлення мовців із психічними вадами у контексті структуризації та відповідності до функціональних складових (за Романом Якобсоном). [60] На даний, момент у нас є лише орієнтовне уявлення стосовно даного аспекту, яке ми не наважились представити у даній роботі, оскільки бракувало аргументації.

Нам було б цікавим продовжити роботу в межах дослідження мовлення людей із психічними вадами у межах теоретичного аспекту (увиразнення функціональної та лінгвістичної складової мовлення людей із психічними вадами). Також, ми б хотіли продовжити аналіз та обробку за допомогою комп'ютерних методів та увиразнення ключових аспектів мовлення у межах інших лінгвістичних теорій (теорія риторичної структури, як зазначено вище).

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Боброва Екатерина Давидовна. "Роль теории речевых актов в становлении и развитии лингвистической прагматики" Russian Journal of Linguistics, no. 1, 2012, pp. 5-11.
2. Бурлачук Л. Ф., Кочарян А. С, Жидко М. Е. Психотерапия: Учебник для вузов. 2-е изд., стереотип. — СПб.: Питер, 2007. — 480 с: ил. — (Серия «Учебник для вузов»).
3. Ділі Джон. Основи семіотики/ Перекл. з англ. та наук. ред. Анатолій Карась - К.: Арсенал, 2000. - 232 с
4. История философии: Запад-Россия-Восток (книга четвертая. Философия XXв.).- М.: 'Греко-латинский кабинет' Ю.А. Шичалина, 1999.- 448с.
5. Самохвалов В. Психиатрия. – Издательский дом: Феникс, 2002
6. Синиця А. Логіко-філософський аспект теорії «мовленнєвих актів» Львівський національний університет імені Івана Франка, УДК 167.2, 2008
7. Каннабих, Ю. В. История психиатрии / Ю. В. Каннабих. — Москва : Издательство Юрайт, 2019. — 384 с. — (Антология мысли). — ISBN 978-5-534-10526-1. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://www.urait.ru/bcode/430718>
8. Кернберг О.Ф. Тяжелые личностные расстройства: Стратегии психотерапии. / Пер. с англ. М.И. Завалова. ² М.: Независимая фирма Класс , 2000. 464 с.
9. Коркина М.В., Лакосина Н.Д. Психиатрия. МЕДпресс-информ
10. Кочерган М. П. Вступ до мовознавства. — К.: ВЦ «Академія», 2006. — С.326
11. Кербиков О.В., Коркина М.В., Наджаров Р.А., Снежневский А.В. Учебник. – 2-е изд., перераб. – М.: Медицина, 1968. – 448 с.
12. Жмуров В.А. Психопатология. Часть I. Симптоматология Иркутск: ИГУ, 1986. - 165 с.
13. Шинкарук В. І. Семіотика // Філософський енциклопедичний словник

14. Alberto, P., Brasso, C., Morese, R. et al. Understanding communicative intentions in schizophrenia using an error analysis approach. *npj Schizophr* 7, 12 (2021). <https://doi.org/10.1038/s41537-021-00142-7>
15. An overview of the approaches and methods for analyzing a text from a discursive viewpoint || Concepcion Hernandex-Guerra
16. Austin, J. L. *How to do things with Words*, ed. J. O. Urmson and Marina Sbisa. Cambridge, MA: Harvard University Press, 1975.
17. Benjamin Bengfort, Rebecca Bilbro, and Tony Ojeda. 2018. *Applied Text Analysis with Python: Enabling Language-Aware Data Products with Machine Learning* (1st. ed.). O'Reilly Media, Inc.
18. Bird, Steven & Klein, Ewan & Loper, Edward. (2009). *Natural Language Processing with Python*.
19. Carrascal, Begoña and K. Korta. "Towards a logic of communicative intention" (2004).
20. Cohen, R. Analyzing the Structure of Argumentative Discourse. *Computational Linguistics*, 13(1-2): 11-24, 1987.
21. Covington, Michael & He, Congzhou & Brown-Johnson, Cati & Naci, Lorina & McClain, Jonathan & Sirmon-Taylor, Bess & Semple, James & Brown, John. (2005). Schizophrenia and the structure of language: The linguist's view *B. Schizophrenia research*. 77. 85-98. 10.1016/j.schres.2005.01.016.
22. Duch Charles, Frawley William. A system of semantic equivalence in specialized discourse. In: *Équivalences*, 9e année-n°2-3, 1978. pp. 29-35.
23. E. Grave*, P. Bojanowski*, P. Gupta, A. Joulin, T. Mikolov, *Learning Word Vectors for 157 Languages*
24. Frow, J. A. (1989). Formal method in discourse analysis. *Journal of Pragmatics*, 13, 333-341. ON EQUIVALENCE CLASSES
25. Gardner, Rod. (2008). *Conversation Analysis*. 10.1002/9780470757000.ch10.
26. Gee, James. (1999). *An Introduction to Discourse Analysis: Theory and Method*.

27. Glynos, Jason & Howarth, David & Norval, Aletta & Speed, Ewen. (2009). Discourse Analysis: varieties and methods.
28. Green, Mitchell, "Speech Acts", The Stanford Encyclopedia of Philosophy (Winter 2020 Edition), Edward N. Zalta
29. Grice, H.. "H. Paul Grice. 1957. 'Meaning.' *Philosophical Review* 66: 377–388, presently published by Duke University Press." (2014).
30. Guerra, Concepción. (2014). An overview of the approaches and methods for analysing a text from a discursive viewpoint. *Onomázein Revista de lingüística, filología y traducción*. 30. 237-246. 10.7764/onomazein.30.17.
31. Harris, Daniel W. ; Fogal, Daniel & Moss, Matt (2018). *Speech Acts: The Contemporary Theoretical Landscape*. In Daniel Fogal, Matt Moss & Daniel Harris (eds.), *New Work on Speech Acts*. Oxford, UK: Oxford University Press.
32. Harris, Z. (1952). Discourse Analysis. *Language*, 28(1), 1-30. doi:10.2307/409987
33. Hinzen, Wolfram & Rosselló, Joana. (2015). The linguistics of schizophrenia: thought disturbance as language pathology across positive symptoms. *Frontiers in psychology*. 6. 971. 10.3389/fpsyg.2015.00971.
34. J. Searle, John R. *Speech Acts: An Essay in the Philosophy of Language*. London: Cambridge University Press, 1969.
35. Jones, K.. "A statistical interpretation of term specificity and its application in retrieval." *J. Documentation* 60 (2004): 493-502.
36. Katz, Jerrold & Fodor, Jerry. (1963). The Structure of a Semantic Theory. *Language*. 39. 170-210. 10.2307/411200.
37. Kilgarriff, Adam. (1997). Putting Frequencies in the Dictionary. *Int. J. Lexicogr.* 10. 10.1093/ijl/10.2.135.
38. Kiparsky, P. *orphological Units: Stems*.
39. Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, 259-284.

- 40.Langford, C. (1938). Charles W. Morris. Foundations of the theory of signs. International encyclopedia of unified science, vol. 1, no. 2. The University of Chicago Press, Chicago 1938, vii 59 pp. Journal of Symbolic Logic, 3(4), 158-158. doi:10.2307/2267781
- 41.Lavin, Matthew. (2019). Analyzing Documents with TF-IDF. The Programming Historian. 10.46430/phen0082.
- 42.Leichsenring F, Leibing E, Kruse J, New AS, Leweke F. Borderline personality disorder. Lancet. 2011 Jan 1;377(9759):74-84. doi: 10.1016/S0140-6736(10)61422-5. PMID: 21195251
- 43.Lyons, J. (1995). Lexical meaning. In Linguistic Semantics: An Introduction. Cambridge: Cambridge University Press.
- 44.Mann W.C., Thompson S.A. (1987) Rhetorical Structure Theory: Description and Construction of Text Structures. In: Kempen G. (eds) Natural Language Generation. NATO ASI Series (Series E: Applied Sciences), vol 135. Springer, Dordrecht. https://doi.org/10.1007/978-94-009-3645-4_7
- 45.Manning, C., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511809071
- 46.McEnery, T., Xiao, Z., & Tono, Y. (2006). Corpus based language studies: An advanced resource book.
- 47.Mel'čuk, I., & Milićević, J. (2020). An Advanced Introduction to Semantics: A Meaning-Text Approach. Cambridge: Cambridge University Press. doi:10.1017/9781108674553
- 48.Murray, Robin (2008). Essential psychiatry (4th ed.). Cambridge University Press. ISBN 978-0-521-60408-6. OCLC 298067373.
- 49.Nematzadeh, A. et al. "Evaluating Vector-Space Models of Word Representation, or, The Unreasonable Effectiveness of Counting Words Near Other Words." Cognitive Science (2017): n. pag. Language Corpora MICHAEL STUBBS

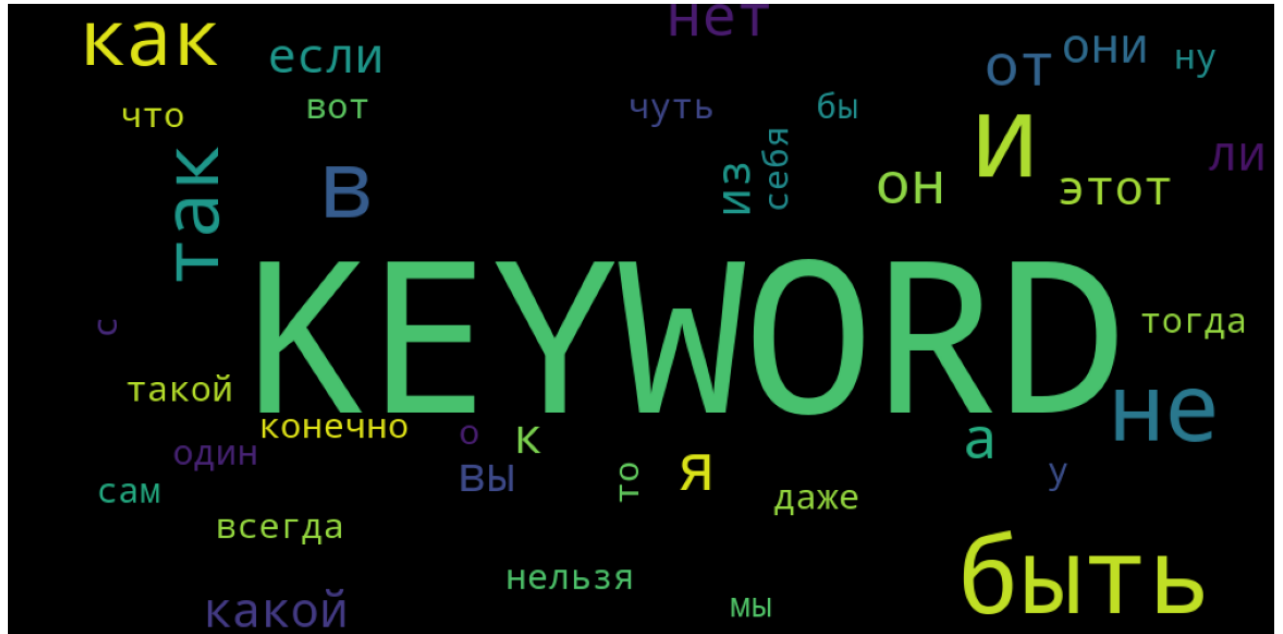
50. Récanati, F.. "On Defining Communicative Intentions." *Mind & Language* 1 (1986): 213-241.
51. Rieger, Burghard B. (1991). *On Distributed Representations in Word Semantics*
52. Robin Cohen. 1983. *A computational model for the analysis of arguments*. Ph.D. Dissertation.
53. Robin Cohen. 1984. *A computational theory of the function of clue words in argument understanding*. In *Proceedings of the 10th International Conference on Computational Linguistics and 22nd annual meeting on Association for Computational Linguistics (ACL '84/COLING '84)*. Association for Computational Linguistics, USA, 251–258.
54. *Steps for Creating a Specialized Corpus and Developing an Annotated Frequency-Based Vocabulary List* by Marie-Claude Toriida
55. Trappes-Lomax, Hugh. (2008). *Discourse Analysis*. 10.1002/ 9780470757000.
56. Wetherell, Margaret; Taylor, Stephanie and Yates, Simeon J. eds. (2001). *Discourse Theory and Practice: A Reader*. London: Sage Publications Ltd.
57. Wildfeuer, Janina. (2013). *Formal Approaches to Discourse Analysis*. *Zeitschrift für Semiotik*. 35. 393-417.
58. Wrobel Janusz (1989). *Language and schizophrenia*
59. Zellig S. Harris (1954) *Distributional Structure*, *WORD*, 10:2-3, 146-162, DOI: 10.1080/00437956.1954.11659520
60. Jakobson, Roman. "Linguistics and Poetics". *Volume III Poetry of Grammar and Grammar of Poetry*, Berlin, Boston: De Gruyter Mouton, 2010, pp. 18-51. <https://doi.org/10.1515/9783110802122.18>

СПИСОК ЕЛЕКТРОННИХ РЕСУРСІВ

61. Шизофрения, шизоаффективное и шизотипическое расстройства, хронические бредовые психозы (В.В. Дунаевский) <http://www.s-psy.ru/obucenie/kurs-psihiatrii/5-kurs-lecebnyj-fakultet/elektronnyj-ucebnik-po-psihiatrii/tema-no9-sizofrenia/sizofrenia-v-v-dunaevskij->

62. Britannica, The Editors of Encyclopaedia. "Ordinary language analysis".
Encyclopedia Britannica, 22 Mar. 2009,
<https://www.britannica.com/topic/ordinary-language-analysis>.
63. ICD-10 Version: 2015 [https://icd.who.int/browse10/2015/en#/F20 -- MKB-10](https://icd.who.int/browse10/2015/en#/F20--MKB-10)
64. How to process textual data using TF-IDF in Python by Mayank Tripathi:
<https://www.freecodecamp.org/news/how-to-process-textual-data-using-tf-idf-in-python-cd2bbc0a94a3/>
65. Stemming and Lemmatization in Python
datacamp.com/community/tutorials/stemming-lemmatization-python
66. The definition of psychosis
cla.purdue.edu/academic/english/theory/psychoanalysis/definitions/psychosis.html

Мовець 7:



Формування корпусу текстів (приклад)

```
"""
Making the corpus #1
"""

path = os.path.expanduser('~/.nltk_data')
if not os.path.exists(path):
    os.mkdir(path)
    #print("created")
#else:
    #print("already exists")

#print(os.path.exists)

getting_data = path in nltk.data.path

corpus = nltk.data.load('corpora/patients/patient16.txt', format='text', encoding='utf-8')
try_text = word_tokenize(corpus)
listToString = ' '.join(map(str, try_text))
#print(listToString)

all_texts = nltk.data.load('corpora/patients/all_texts.txt', format='text', encoding='utf-8')
try_text_all = word_tokenize(all_texts)
all_words = ' '.join(map(str, try_text_all))
#print(all_words)
#print(os.path.abspath(path))
```

Ключові поняття із другого корпусу (мовлення людей із психічними вадами)

'это', 'мочь', 'просто', 'говорить', 'хотеть', 'который', 'сказать', 'жить', 'очень', 'хороший', 'понимать', 'день', 'наркотик', 'ключ', 'дать', 'жизнь', 'смотреть', 'идти', 'первый', 'ходить', 'написать', 'время', 'приехать', 'просить', 'пойти', 'понять', 'каждый', 'любить', 'рука', 'деньга', 'папа',	'доктор', 'бояться', 'искать', 'уходить', 'чувствовать', 'нормальный', 'остальной', 'реально', 'паспорт', 'гривна', 'некоторый', 'интересный', 'равно', 'давно', 'прийти', 'дурка', 'маленький', 'вести', 'помнить', 'кухня', 'молиться', 'ломбард', 'мыться', 'разный', 'наверное', 'мозг', 'попадать', 'полностью', 'убирать', 'улица', 'несколько', 'тусовка', 'нравиться',	'радость', 'кролик', 'советовать', 'вернуть', 'немного', 'стих', 'домой', 'класный', 'паразит', 'иерусалим', 'церковный', 'стараться', 'беспокоить', 'хватать', 'побои', 'газ', 'ставить', 'лечиться', 'дубликат', 'конкретный', 'здоровый', 'шизофрения', 'чёрный', 'слишком', 'готовить', 'этаж', 'спросить', 'адвокат', 'записывать', 'давление', 'холодно', 'конфликт',	'смочь', 'послать', 'нести', 'одержимый', 'норбеков', 'биби', 'семьсот', 'выход', 'психологически', 'секунда', 'устроить', 'заснуть', 'место', 'дорогой', 'помощник', 'закрывать', 'вставить', 'одиннадцать', 'ценить', 'близкий', 'ответственность', 'бросать', 'общение', 'находиться', 'экспериментировать', 'стимулятор', 'програмка', 'санитар', 'ремонт', 'приговор', 'довольно', 'старый',	'пузырёк', 'годиться', 'ванная', 'агрессия', 'чумачество', 'путешеств', 'раскинуть', 'доказывать', 'звено', 'работящий', 'наркота', 'продумать', 'шизофреник', 'правило', 'напиток', 'оплатить', 'покурить', 'скрытный', 'разъехаться', 'владимирович', 'раскрыть', 'хлам', 'зламаль', 'дюма', 'личность', 'пожалуй', 'здавать', 'проституция', 'собираться', 'средство', 'окно', 'вчерашний',
--	--	--	--	---

Ключові поняття із першого корпусу (мовлення людей без психічних вад)

'это',	'дело',	'тронутый',	'стандарт',	'заговорить',
'всё',	'считать',	'выборы',	'нынешний',	'эксперимент',
'говорить',	'написать',	'реально',	'различать',	'волошин',
'ещё',	'собственно',	'чёткий',	'вспоминать',	'новогодний',
'знать',	'шестой',	'включая',	'подпадать',	'коммерческий',
'авария',	'слово',	'повод',	'болгарский',	'якутский',
'очень',	'образ',	'абсолютный',	'забрать',	'наблюдение',
'разный',	'делать',	'площадка',	'номер',	'минимум',
'друг',	'наш',	'заниматься',	'ремонт',	'основание',
'мощность',	'например',	'принять',	'людиковский',	'мельтешить',
'далее',	'произойти',	'текст',	'пожить',	'приплыть',
'персонал',	'сторона',	'напомнить',	'некуда',	'хронометраж',
'работа',	'провести',	'начинать',	'бывало',	'тривиальный',
'доклад',	'страна',	'обсуждение',	'народный',	'пройтись',
'время',	'беслан',	'ошибка',	'лингвистика',	'собираться',
'вообще',	'власть',	'круг',	'предыдущий',	'плавать',
'понимать',	'изменение',	'оставаться',	'фарм',	'низка',
'идти',	'общаться',	'сожаление',	'кусок',	'параллельно',
'просто',	'специалист',	'карельский',	'предпринять',	'тонкий',
'иметь',	'путин',	'ребята',	'склониться',	'спица',
'ситуация',	'практически',	'оказаться',	'готовность',	'практика',
'жить',	'часть',	'точно',	'длина',	'смысл',
'работать',	'анализ',	'квартира',	'коронация',	'отбалансировать',
'должный',	'остановка',	'коллективный',	'восьмой',	'образование',
'языковой',	'история',	'возможно',	'сессия',	'болгарин',
'поэтому',	'тысяча',	'ругать',	'расследовать',	'дзасохв',
'думать',	'акимов',	'изучать',	'образоваться',	'актовый',
'пойти',	'называться',	'горный',	'формализм',	'изъясняться',
	'уволить',	'французский',	'происходить',	'системник',
	'пытаться',	'невероятный',	'будить',	'автоматика',
	'примерно',	'направление',	'урезать',	'собеседник',
	'исландия',			'нежелание',