

# ТЕОРЕТИЧНІ ЗАСАДИ ТЕРМІНОЗНАВСТВА ТА ЛЕКСИКОГРАФІЇ

## НОРМУВАННЯ ТА СТАНДАРТИЗАЦІЯ ТЕРМІНОЛОГІЇ

УДК 800

Тарас Кияк

Київський національний університет імені Тараса Шевченка

### ВУЗЬКОГАЛУЗЕВІ ТЕРМІНИ ЯК ОСНОВА ФОРМУВАННЯ ТА КВАЗІРЕФЕРУВАННЯ ФАХОВИХ ТЕКСТІВ

© Кияк Т., 2008

**Фахова мова – сукупність усіх мовних засобів, уживаних в обмеженій професійною ділянкою сфері комунікації для забезпечення взаєморозуміння між людьми, які працюють у цій галузі. Функціонування фахової мови забезпечує докладно визначена термінологія. Словник професійних текстів містить вузькогалузеві лексеми, міжгалузеві наукові терміни, професіоналізми (виробничу лексику), номенклатурні назви, професійний жаргон і загальноновживані лексеми. Рівень абстракції фахової мови залежить від рівня насиченості тексту термінологією. Терміни – найчастіше вживані в тексті одиниці, які можна використати, конструюючи квазівитаг.**

**Professional language is the complex of all language means being used in professionally restricted area of communication in order to ensure the mutual understanding among people working in the given sphere. The functioning of professional language is provided with accurately determined terminology. Nevertheless the vocabulary of professional text includes, apart from narrow-branch lexical units, inter-branch general scientific terms, professionalisms (industry words), nomenclatures, professional jargon and the lexis of general use. The level of abstract professional language depends on the level of the text's terminological saturation. Terms are the most frequently used lexical units in a text, which can be used while constructing a quasiabstract.**

Дослідження фахових мов (інакше: «спеціальні мови», «технолекти») має об'єктом особливості мови певної галузі науки. При цьому фахова мова до певної міри протиставляється загальноновживаній мові, хоча словниковий запас першої бере початок саме із другої. Але якраз проблеми, пов'язані зі специфічною лексикою фахових мов, породили необхідність виділення окремої науки – термінознавства. У той час, як наука про фахові мови має на меті дослідження власне мови певної сфери науки і техніки, предметом дослідження термінознавства є не мова як така, а саме термін як єдність поняття і назви. Таким чином, термінознавство охоплює і понятійну сферу певної фахової галузі.

На перших порах дослідження фахових мов стосувалося тільки професійної лексики. Лише недавно воно виокремилася у власну гілку лінгвістики, хоча залишається дуже тісно пов'язаним із

термінознавством. Фаховою мовою можна називати сукупність усіх мовних засобів, які застосовуються у професійно замкнутій сфері комунікації з метою забезпечення порозуміння між людьми, які працюють у цій сфері [3, с. 139]. До цієї дефініції варто додати ще й той факт, що функціонування фахової мови забезпечує винятково чітко встановлена термінологія.

Усю лексику фахового тексту можна поділити на 4 види:

- 1) терміни даної галузі, які мають власну дефініцію;
- 2) міжгалузеві загальнонаукові термінологічні одиниці (у т.ч. терміни суміжних наук) – наприклад, поширені терміни філософії, політології, математики, філології і т.п.;
- 3) напівтерміни або професіоналізми, до яких можна віднести і номенклатури, щоправда, професіоналізми, як правило, можуть тлумачитися, на протипагу номенклатурним одиницям (пор. професіоналізм «Beißzange» та номен «Opel-Vectra»);
- 4) професійні жаргонізми, які не претендують на точність та однозначність, мають великий рівень образності та емоційно забарвлене значення (напр. «Нехе» – «підйомник матеріалів на будові», «Schießbude» – «ударний інструмент джазової капели»).

Якщо погодитися з тим, що вузькогалузєва лексика найважливіша і найбільш уживана в даних текстах, то цілком логічно напрошується той факт, що вона містить у собі найвагомішу інформацію. Якщо так, то вона має стати базою для компресії тексту, найперше для побудови об'єктивного і навіть автоматичного реферату.

У зв'язку з інформаційним бумом, коли загальна кількість інформації подвоюється через кожні 8–10 років, виникає проблема її опрацювання. Тут маємо чергове протиріччя: з одного боку, вчений зобов'язаний стежити за новими досягненнями у своїй спеціальності, з іншого, – він не в змозі віднайти й засвоїти інформацію необхідного об'єму. Та йому не обов'язково знайомитися з усією літературою, де можлива присутність для вченого інформаційних «шумів», надлишкових чи відомих знань.

У науково-технічній літературі зазвичай розрізняють три типи інформації: 1) фактографічну – інформація про факти, явища, процеси, події; 2) логіко-теоретичну – повідомлення про методи отримання фактографічної інформації, висновків із фактів, посилання на джерело інформації; 3) оцінкову – вираження авторського ставлення до повідомлення. Зрозуміло, що найважливішою інформацією з точки зору оброблення тексту є фактографічна інформація, яка виражається переважно спеціальними одиницями підмови науки.

Розрізняють два види компресії тексту – анотування та реферування. Анотування – це максимально коротка характеристика матеріалу. Провести необхідний відбір літературних джерел допоможуть користувачеві їхні відповідні реферати, тобто короткі викладення наукової праці, книги, статті, які передають головні ідеї, напрям наукового пошуку. Такий реферат може бути складений перекладачем, спеціалістом, інформаційним працівником.

На практиці розрізняють кілька видів анотування та реферування, причому реферат може включати в себе як попередній коментар усі етапи продукування анотації. Поетапний алгоритм має такий вигляд: 1) предметна рубрика (в т.ч. визначення відповідної галузі); 2) назва теми; 3) вихідні дані джерела; 4) стисла характеристика матеріалу; 5) критична оцінка першоджерела; 6) стислий виклад змісту; 7) висновки автора; 8) коментар референта. Зауважимо, що анотування включає в себе перші п'ять етапів, у той час як етапи 6–8 є прерогативою процесу реферування, яке, повторимося, може включати й перші п'ять етапів.

З ростом кількості джерел росте й число створювачів рефератів. Тому виникає необхідність розробки методики автоматичного реферування. При цьому всі методи опираються на віднайдення ключових слів, у функції яких у науково-технічних текстах виступають термінологічні одиниці. При цьому терміни слугують основою формального аналізу семантики тексту.

Прикладом методики спрощення аналізу слугує алгоритм компресії інформації тексту. Під компресією тексту мається на увазі один із можливих варіантів його квазіреферування, тобто складання реферату на основі відбору й комбінування готових фрагментів тексту. Автоматизація процесу реферування – надзвичайно велика наукова й практична потреба не тільки задля досягнення мети (отримання реферату), скільки завдяки розробці методології автоматичного визначення найсуттєвішого, цінного в науково-технічній інформації.

Виділення таких факторів є не що інше, як оцінка фактів, певною мірою оцінка ефективності того чи іншого наукового дослідження.

Визначальною гіпотезою підходу до проблеми квазіреферування виступає припущення, що найважливіша фраза тексту має найбільше число лексичних зв'язків з іншими фразами даного тексту. Під лексичним зв'язком мається на увазі присутність однакових компонентів (основ) тексту, який розглядається.

Далі формується матриця лексичної суміжності, де фіксуються всі загальні елементи масиву в лінійній послідовності. Окрім того, можна побудувати таку матрицю виходів, де фіксуються циклічні шляхи з метою виявлення структури абзаців. З допомогою такої методики можна перевірити стилістичну правильність побудови абзаців. У більшості випадків у кінці абзацу є речення, що містить найменшу кількість лексичних одиниць, і навпаки, абзац вводиться переважно лінгвістично найбільш вагомими фразами. Таким чином, у квазіреферат потраплять найважливіші фрази абзацу.

Далі встановлюється функціональна вага цілих речень  $F$  у межах всього досліджуваного масиву:

$$F = t_i / l(n-1),$$

де  $t_i$  – кількість лексичних зв'язків речення;  $l$  – кількість вершин (основ) у реченні;  $n$  – кількість речень у тексті.

До даної формули слід внести деякі зміни, якщо враховувати не лише кількість лексичних зв'язків між реченнями, але і їх функціональну вагу в межах пропонованого бінарного семантичного дерева, яке репрезентує повну теморематичну структуру кожного речення, і де враховується ступінь «віддаленості» слова від вершини дерева всього речення.

Наступний етап полягає у «відкиданні» речень з найменшою функціональною вагою. Так робиться до тих пір, поки не залишаться речення, кількість яких відповідає замовленому обсягові реферату.

Аналогічним чином було проаналізовано понад тридцять науково-технічних текстів різних фахових мов. Вислідом роботи стали реферати розміром 8-10 речень. За твердженням більшості фахівців, квазіреферати вийшли досить задовільними. Більш того, свідченням надійності пропонованого алгоритму можна вважати той факт, що в реферат, як правило, увійшли найчастотніші слова тексту. Таким чином, статистика підтверджує семантичну вагу не лише речень, але й слів у тексті.

Пропонована модель репрезентує дуалізм формування людської думки в інформаційному режимі «нове-відоме», що має вагоме значення для вирішення низки питань штучного інтелекту [2]. Водночас це важливо з огляду моделювання семантичних властивостей людини, що так важливо для побудови автоматизованих систем обробітку тексту [1, с. 9].

У сфері автоматичних сфер управління такі роботи спрощують процеси фіксації даних, покращують збереження інформації, прискорюють пошук і групування необхідних відомостей, упорядковують документообіг завдяки вилученню з обігу проміжних та нерелевантних даних.

Автоматичне квазіреферування можна вважати також одним із перших кроків до автоматичної оцінки науково-технічної інформації, до автоматичного створення високоефективних фактографічних систем.

Всезростаючий інформаційний обмін передбачає не лише збільшення науково-технічних публікацій, у тому числі й перекладів, а й підвищення їх якості. І хоча в даний час видається багато словників, довідникової літератури й посібників з перекладу, вони не завжди надають достатню допомогу перекладачеві в подоланні всіх труднощів, оскільки торкаються лише окремих випадків, конкретних питань, не вникаючи в проблему оформлення всього тексту перекладу. Таким чином, бачимо цілу палітру можливостей запропонованого алгоритму виявлення лексико-семантичних зв'язків між реченнями тексту, ступеня його термінологічності як одного із методів квазіреферування фахових текстів.

1. Городецький Б. Ю. *Актуальные проблем прикладной лингвистики* / Б. Ю. Городецький // *Новое в зарубежной лингвистике. – Вып. XII прикладная лингвистика. – М.: «Радуга», 1983. – С. 5–22.*
2. Кияк Т. Р. *Мотивированность лексических единиц (количественные и качественные характеристики)* / Т. Р. Кияк. – Львов: *Высшая школа*, 1988.
3. Hoffman Lothar. *Fachsprachen als Subsprachen/ Fachsprachen. Walter de Gruyter. Berlin-New York, 1998.*