

ПОЗИЦІЙНА ІНТЕРФЕРЕНЦІЯ ТЕРМІНІВ ЯК СТРУКТУРНА ХАРАКТЕРИСТИКА НАУКОВОГО ТЕКСТУ

Позиційна інтерференція термінів, тобто присутність різних термінів в одному відрізку наукового тексту, розглядається як вияв зв'язку між відповідними спеціальними поняттями. Статистичне обстеження текстів свідчить, що існує кореляція між інтенсивністю позиційної інтерференції термінів та комунікативною складністю тексту.

Ключові слова: термін, науковий текст, комунікативна структура тексту.

У науковому тексті приблизно 30–40% слововживань припадає на термінологічну лексику. Текст навіть невеликого розміру — стаття — містить іноді до трьохсот – п'ятисот різних термінів. Це не випадково. Адже поміж усіх мовних засобів саме термін є головним носієм фахових знань. Саме він передає інформацію, заради якої створюється текст. У зв'язку з цим виникає питання: чи існують закономірності, які регулюють уживання термінів у тексті, а якщо існують, то чи не допоможе їхнє пізнання глибшому розумінню механізмів текстотворення?

Це питання стає дедалі актуальнішим, оскільки в останні роки увага мовознавців усе більше спрямовується на функціонально-комунікативні аспекти застосування мови, зокрема на дискурс як найважливішу категорію комунікації. Досліджуються загальні закони спілкування за допомогою мови, мовні засоби комунікації, комунікативні ситуації, комунікативна стратегія й тактика, структура дискурсу тощо¹.

Здебільшого методика таких досліджень полягає в застосуванні загально-теоретичних положень і міркувань при розв'язанні конкретного завдання й обґрунтуванні одержаних результатів більшою чи меншою кількістю прикладів. До вивчення комунікативних стратегій або тактик підходять, аналізуючи ментальні процеси, пов'язані зі створенням тексту.

В основі нашої статті лежить дещо інша ідеологія. Ми виходимо з припущення, що головним чинником, який формує науковий текст, отже й усі його складники, і визначає когнітивні, комунікативні, тематичні аспекти тексту, є термін. Тому, на нашу думку, саме йому належить чільне місце серед компонентів тексту, аналіз яких може дати відповідь на низку питань, пов'язаних зі структурою та функціонуванням наукового дискурсу.

Дослідження наукового дискурсу неодмінно має включати аналіз термінологічного складника тексту: структурних, семантичних характеристик термінів, їхньої поведінки в тексті. Останній є результатом (матеріальним продуктом) певних когнітивних та комунікативних процесів. Аналізуючи текст, зокрема його термінологічний складник, ми отримуємо можливість певною мірою рекон-

¹ *Полюжин М. М.* Сучасні напрямки й підходи до вивчення комунікативних актів дискурсу // *Іноземна філологія.* — 2004. — Вип. 113. — С. 165–171.

струювати ці процеси й визначити деякі властивості, характеристики, ознаки тексту, які не мають прямого матеріального вираження, такі як тематична чи когнітивна структура, комунікативна стратегія й тактика тощо. Звичайно, не можна ставити знака рівності між текстом і дискурсом. Проте, незважаючи на різні трактування самого поняття дискурсу², ніхто не може заперечувати, що чільним компонентом дискурсу є текст. До того ж текст є єдиним складником дискурсу, який має матеріальне втілення. Тому поки що адекватні об'єктивні наукові методи дослідження опрацьовані лише для тексту.

Особливо перспективним для розв'язання конкретних завдань теорії дискурсу є аналіз взаємодії термінів у тексті з урахуванням усіх терміновживань.

У статті термін може повторюватися до двохсот–трьохсот разів (тому він є головним маркером зв'язності в науковому тексті). У різних уживаннях один і той самий термін виконує різні комунікативні функції. Усі вживання термінів у науковому тексті належать до двох категорій: інформативних чи неінформативних.

Інформативне вживання розуміємо як використання терміна, що доносить до читача більш-менш суттєву інформацію про референт: про його ознаки, властивості, поведінку, застосування тощо. Розрізняємо два типи інформативного вживання терміна: домінантне та підпорядковане.

Під домінантним уживанням розуміємо випадки, коли мета речення — передати інформацію про відповідний референт (тут і в наступних прикладах виділено термін, уживання якого розглядається): *An active control system designed to augment system damping was also tested; The active control system was successful in augmenting the damping of the fundamental system modes, and was found to be robust with respect to changes in rotor-speed and airspeed; Soft-inplane rotor systems generally have reduced damping margins and lower stability boundaries than stiff-inplane rotor systems; Soft-inplane rotor system had two sets of elastomeric dampers [...].*

У разі підпорядкованого вживання головна комунікативна функція речення спрямована на опис іншого референта, а термін, який нас цікавить, є, наприклад, назвою ознаки цього референта: *Conventional helicopters with three or more blades have soft-inplane rotor systems; The lag-mode frequency and damping of the new soft-inplane rotor system were measured [...].*

Розмежування між домінантним і підпорядкованим використанням терміна має певну аналогію з тема-рема-тичною класифікацією речень, але тема-рема-тичний статус терміна не завжди відповідає розглянутим типам.

Під неінформативним уживанням розуміємо таке використання терміна, яке передає інформацію хіба що про сам факт існування референта. Типовим прикладом є метатематичне вживання. Тут функція терміна — інформувати читача не про властивості референта, а про те, як він представлений у тексті, у яких працях розглядається тощо: *The high heating rates caused by the turbulence are illustrated in Fig. 2; Fig. 33 shows the measured heat-transfer rate [...]; The larger bluntnesses were studied by Stetson [...]; Zanchetta et al. measured transition on a blunt cone at zero AOA (angle of attack).*

До варіантів метатематичного вживання належить згадування мікро-референта у зв'язку з описом мети, структури тексту: *This article is to provide a survey on the data mining techniques [...], with an emphasis on database-oriented techniques.*

² Nunan D. *Introducing Discourse Analysis*.— London, 1993.— 134 p.; McCarthy M. *Discourse analysis for language teachers*.— Cambridge University Press, 1991.— 214 p.; Brown G. and Yule G. *Discourse analysis*.— Cambridge University Press, 1983.— 288 p.

Інший приклад неінформативного вживання — анонсування референта, тобто його перше згадування без висвітлення суттєвих ознак. Типові контексти — констатація актуальності певного поняття, перелік досліджуваних об'єктів тощо: Researchers in many different fields have shown great interest in *data mining*; Due to the rapid evolution of multimedia, more and more *digital images* are being captured and stored; Most bicycles fall into one of four main categories: *utility*, *touring*, *racing*, and *mountain bicycles*.

За попередніми підрахунками, кількість неінформативних уживань не перевищує 3%. Серед інформативних уживань приблизно 60% припадає на домінуючі, 40% — на підпорядковані.

Найчастіше в кожному відрізку тексту вживаються паралельно два чи кілька різних термінів (пересічному мовознавцю зрозуміти зміст як окремих технічних термінів, так і статті в цілому досить важко, але завдання полягає лише в тому, щоб простежити поведінку терміна в тексті — як він переходить з одного речення в інше, з якими своїми «родичами» зустрічається на шляху цієї мандрівки):

(1) (Since a rule can be defined explicitly or be extracted by a knowledge discovery process, it is important to differentiate a deduction rule from a generalized one.

(2) (A rule which is explicitly defined by a user or an expert is a deduction rule; whereas a rule which is generalized from a database is a generalized rule.

(3) (Both deduction rules and generalized rules can be primitive-level or high-level rules.

(4) (However, since a generalized rule summarizes data from a database, it reflects a general fact in the current database but does not enforce a constraint on the possible databases.

(5) (This contrasts with a deduction rule or an integrity constraint which states a rule (or a constraint) that potential database must follow.

(6) For example, a generalized rule, «all of the teaching assistants are graduate students» states a fact but does not claim that all the teaching assistants must be graduate students nor reject the possibilities of hiring an undergraduate student as a teaching assistant.

(7) However, the same rule, stated as a deduction rule or an integrity constraint, will reject the possibility of allowing an undergraduate student to serve as a teaching assistant in a consistent database.

Тут у семи реченнях ужито 10 спеціальних термінів: *constraint*, *data*, *database*, *deduction rule*, *generalized rule*, *high-level rule*, *integrity constraint*, *knowledge discovery process*, *primitive-level rule*, *rule*. Деякі речення містять по кілька термінів.

Про що свідчить наявність двох чи кількох різних термінів в одному реченні або в суміжних реченнях?

Паралельне вживання двох чи кількох термінів в одному відрізку тексту — це іманентна властивість наукового тексту. Вона зумовлена переважно когнітивними чинниками. Між окремими референтами існує зв'язок. Якщо цей зв'язок є релевантним для теми, він знаходить своє відображення в позиційній близькості відповідних термінів. Наприклад, у наведеному вище фрагменті тексту в першому реченні стверджується, що при одержанні інформації з бази даних важливо розрізняти два типи правил — дедуктивні та узагальнені. Таким чином, фіксується зв'язок між мікрореферентами «*knowledge discovery process*», «*rule*», «*deduction rule*», «*generalized rule*» (мікрореферентом тут названо референт, який відповідає окремому терміну³). Друге речення дає визначення дедуктивних та узагальнених правил. При цьому експлікується їхній зв'язок з мікрореферентом «*database*». У наступному реченні пропонується подальша класифікація правил.

Рідше паралельне вживання термінів пояснюється суто комунікативними чинниками. Наприклад, порівнюються чи протиставляються два референти:

³ Скороходько Е. Ф. Неоніми в когнітивній та комунікативній структурі наукового тексту (на матеріалі англійської мови) // Мовознавство.— 2004.— № 1.— С. 45–53.

However, unlike *Fourier descriptor*, only the few lower order moment descriptors have physical interpretation. Іноді таке порівняння чи протиставлення охоплює весь текст. Наприклад, автор описує конструкцію та функціонування дизельного двигуна, послідовно порівнюючи його з більш відомим карбюраторним (бензиновим) двигуном: [...] This is a fundamental difference between a *diesel engine* and a *gasoline engine*. In a *gasoline engine*, a spark is used to ignite the combustible mixture, while in a *diesel engine* the fuel ignites by itself just from contact with air — air that is very hot because it has been highly compressed [...].

Терміни вживаються паралельно (незалежно від того, чи є між ними зв'язок, чи ні) і там, де йдеться про структуру статті: In Section 2, a *data model* is constructed for *knowledge-rich databases*, which consists of both *deductive database* and *knowledge discovery components*. In Section 3, four basic categories of *query answering mechanisms* in a *knowledge-rich database* are introduced, according to the combinations of *data vs. knowledge queries* and *direct vs. intelligent query answering mechanisms*.

Уживання двох чи кількох термінів в одному відрізку тексту як вияв зв'язку між (мікро)референтами назвемо позиційною інтерференцією термінів⁴. Розглянемо на дещо більшому прикладі інтерференцію двох термінів у науковій статті⁵ (див. табл. 1). Тут наведено речення, в яких ужито терміни *discretisation* «дискретизація» та *truncation* «округлювання». У першому стовпчику вказано номер речення, у третьому й четвертому — який з двох термінів наявний у відповідному реченні (d — *discretisation*, t — *truncation*).

Шосте речення, у якому вперше з'являються терміни *discretisation* та *truncation*, міститься у «Вступі» аналізованої статті. Воно лише анонсує відповідні мікрореференти. Звернімо увагу на те, що це речення відділяють від наступного входження терміна *discretisation* 89 інших речень, а від входження *truncation* — 101 речення. У той же час між іншими входженнями відстані значно менші. Початкові або кінцеві входження, відірвані від основної частини ланцюжка входжень, назвемо нерегулярними⁶.

Відрізок тексту від першого до останнього входження певного терміна (без урахування нерегулярних входжень) стосується відповідного мікрореферента. Назвемо цей відрізок T-протяжністю. У нашому прикладі референта «*discretisation*» стосується відрізок [96–305], а «*truncation*» — [108–249].

Стаття в цілому містить 310 речень (назва статті, її розділів та підрозділів тут не враховується). Отже, T-протяжність кожного з цих термінів охоплює більшу частину тексту. Разом з тим термін *discretisation* експліцитно присутній лише в одинадцяти, а *truncation* — у десяти реченнях (послідовність речень, кожне з яких містить певний термін, назвемо T-ланцюжком⁷).

Решта речень найчастіше стосується підпорядкованих мікрореферентів або, навпаки, мікрореферентів, щодо яких «*discretisation*» та «*truncation*» є підпорядкованими.

Наприклад, речення 97–98 стосуються мікрореферентів «*thresholding*», «*weighted aggregation*» та «*transformation*», які уточнюють і розвивають інформацію, наведену в попередньому реченні:

(97) The latter two operations are replaced by *thresholding* in the case of Hopfield networks.

⁴ Скороходько Е. Ф. Позиційна інтерференція термінів в англomовному науковому тексті: гіпероніми та гіпоніми // Вісн. Харків. нац. ун-ту : Романо-германська філологія. Методика викладання іноземних мов. — 2004. — № 635. — С. 164–169.

⁵ A Discrete Fully Recurrent Network of Sigmoid Units for Associative Memory and Pattern Classification // Intern. Journ. of Pattern Recognition and Artificial Intelligence. — 2002. — Vol. 16. N 5. — P. 527–550.

⁶ Скороходько Е. Ф. Зазнач. праця. — С. 45–53.

⁷ Там же.

(98) The *weighted aggregation* determines similarity and the *transformation* introduces a nonlinearity and also bounds the output.

Низка речень (9, 17, 62, 63 та ін.) стосується мікрореферента «learning algorithm», якому підпорядковані як «discretisation» так і «truncation»:

(9) Two *learning algorithms* for training such a network to store fixed points or attractors are proposed.

(17) [...] the general use of Recurrent Neural Networks is not as extended as that of Feedforward Neural Networks due to the complexity of the development of the *learning algorithms*.

(62) Following this is a detailed description of two *learning algorithms*.

(63) Simulations described in the section that follows demonstrate the effectiveness of the *learning algorithms* and the properties of the network in terms of basins of attraction.

Таблиця 1

| № | Інтерференція термінів <i>discretisation</i> та <i>truncation</i> | Термін | |
|-----|---|--------|---|
| | | d | t |
| 6 | This paper describes two learning algorithms for a network whose units take the dot product of the input with a weight vector, followed by a <i>tanh</i> transformation and a <i>discretisation</i> transformation in the form of rounding or <i>truncation</i> . | + | + |
| 96 | The complete transformation for one unit will consist of a weighted aggregation [...], a transformation [...] and <i>discretisation</i> . | + | |
| 99 | Both are necessary since [...] <i>discretisation</i> is required to produce a finite number of possible states [...]. | + | |
| 101 | <i>D</i> is the <i>discretisation</i> operator. | + | |
| 108 | An obvious way of carrying out <i>discretisation</i> is to use <i>truncation</i> which produces equally spaced discrete values. | + | + |
| 109 | <i>Truncation</i> makes good sense because it determines the number of digits that need to be retained. | | + |
| 110 | Then <i>T</i> , the <i>truncation</i> operator, is defined as [...]. | | + |
| 112 | For <i>truncation</i> to one place after the decimal we get [...]. | | + |
| 123 | If we assume <i>truncation</i> to one place after the decimal the state vectors will lie in (0,0). | | + |
| 124 | Rather than <i>truncation</i> we may also use rounding. | | + |
| 127 | Rounding can be expressed in terms of <i>truncation</i> as follows [...]. | | + |
| 128 | The formulations of the training algorithms are expected to be very similar whether <i>truncation</i> or rounding is used. | | + |
| 144 | For a vector <i>x</i> to be a fixed point we require that [...] where <i>D</i> is the <i>discretisation</i> operator which so far has been <i>truncation</i> or rounding. | + | + |
| 154 | For <i>truncation</i> the equation is [...]. | | + |
| 164 | The curve is very close to the $y = x$ line where it crosses leading to multiply fixed points if <i>discretisation</i> is carried out. | + | |
| 165 | There is a discrete fixed value wherever the curve is no more than 0.1 above the $y = x$ line in the case of <i>truncation</i> . | | + |
| 182 | Let us assume that the <i>discretisation</i> function is rounding. | + | |
| 195 | The <i>truncation</i> function may be approximated in several ways. | | + |
| 213 | If we assume <i>truncation</i> to one place behind the decimal, <i>d</i> is the output for one unit if and only if [...]. | | + |
| 226 | The actual training algorithm for <i>truncation</i> as <i>discretisation</i> is [...]. | + | + |
| 231 | Rather than <i>truncation</i> we may also use rounding. | | + |
| 249 | <i>Truncation</i> is used for <i>discretisation</i> . | + | + |
| 305 | Additional work may consist of implementing a more general <i>discretisation</i> that is not linear. | + | |

Зв'язок між поняттями «discretisation» та «truncation» виявляється, зокрема, в одночасному входженні відповідних термінів до тих самих речень (6, 108, 144, 226, 249). Підпорядкування референта «truncation» референту «discretisation», так само як їх обох референтів «learning algorithm», експліковано в реченні 6.

Таким чином, науковий текст має складну комунікативну структуру, компонентами якої є фрагменти тексту (Т-протяжності), що стосуються окремих мікрореферентів. Між цими фрагментами існують два типи відношень: включення (одна Т-протяжність є частиною іншої) та перетину (дві Т-протяжності мають спільні речення, хоч не збігаються повністю).

Формальним виявом складності комунікативної структури є позиційна інтерференція термінів. Інтенсивність інтерференції виражається кількістю Т-протяжностей, між якими існує те чи інше відношення, та обсягом відрізків тексту, спільних для кількох Т-протяжностей.

Розглянемо інтерференцію термінів у складі двох невеликих текстів приблизно одного обсягу, відображену на рис. 1. На горизонтальній осі вказані номери речень тексту, на вертикальній представлені терміни в тій послідовності (знизу вверху), у якій автори вводять відповідні поняття. З рисунка видно, на які ділянки тексту припадають входження термінів, проти кожного з них схематично представлена його Т-протяжність. Дві вертикальні риски вказують на її початок та кінець, тобто перше й останнє входження терміна (для *fuselage* це речення 1 та 8). Бічні грані прямокутника позначають місця тексту, яким передують, відповідно, 25% та 75% входжень терміна (для *fuselage* це проміжок між першим і другим реченнями й шосте речення). Таким чином, прямокутник охоплює 50% центральних входжень терміна. Маленький квадрат у межах прямокутника позначає центральну точку ланцюжка, медіану — місце, перед яким налічується 50% уживань терміна (для *fuselage* — третє речення). Оскільки для ілюстрації нашого підходу обрано невеликі тексти, кількість уживань окремих термінів виявилася малою. Тому наведені характерні точки Т-протяжностей іноді зливаються.

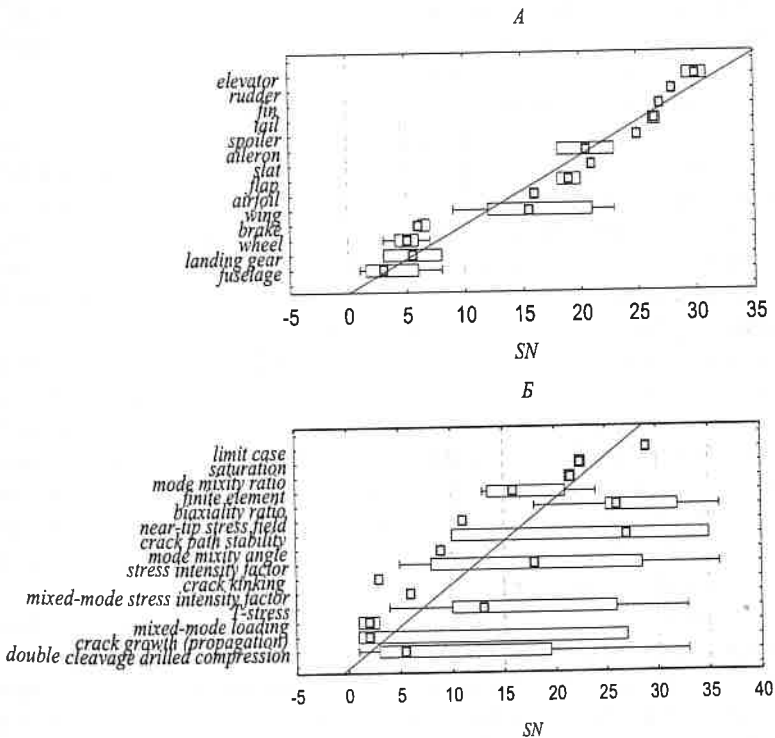


Рис. 1. Співвідношення позицій термінів у тексті

За показник локалізації фрагмента тексту, який стосується певного мікрореферента, прийнято відповідну медіану. Навкісні лінії на обох частинах рисунка ілюструють загальну тенденцію переходу авторської розповіді від одного мікрореферента до іншого. В обох випадках чітко простежується певна послідовність опису: фіюзеляжу присвячений відрізок тексту з першого до восьмого речення, шасі — з третього до восьмого, крило розглядається в реченнях з дев'ятого до двадцять третього і под. Окремі відступи від обраного порядку не порушують загальної картини.

Проте між текстами А та Б існує істотна відмінність: Т-протяжності другого тексту мають у середньому більшу довжину порівняно з першим, вони частіше перетинаються, включаються одна в одну. Це є ознакою інтенсивнішої інтерференції термінів. Інтенсивність інтерференції виражається кількістю Т-протяжностей, між якими існує те чи інше відношення, та обсягом відрізків тексту, спільних для кількох Т-протяжностей.

Якщо врахувати, що текст А — науково-популярний, а Б — суто науковий, теоретичний, то причину, можливо, слід шукати у відмінностях їхньої когнітивної та / або комунікативної структури. Комунікативно найпростіший текст — такий, у якому кожній підтемі відповідає окремий відрізок. Проте глобальний референт наукового тексту⁸ має складну структуру. Між мікрореферентами існують різноманітні зв'язки, що не дозволяє описувати їх ізольовано. Ведучи мову про один із них, автор змушений залучати до викладу інші. Виявом зв'язку між кількома референтами є вживання відповідних термінів в одному реченні чи в одному ланцюжку суміжних речень. Перетин Т-ланцюжків або входження одного ланцюжка до іншого — це індикатор зв'язку між мікрореферентами.

Можна припустити, що інтенсивність позиційної інтерференції термінів у науковому тексті віддзеркалює складність його комунікативної структури, яка, у свою чергу, залежить від двох чинників — складності глобального референта (отже, і когнітивної структури тексту) та обраної автором стратегії побудови тексту. Більша складність когнітивної структури приводить до більшої кількості зв'язків між мікрореферентами. Тому в наукових текстах важливу роль відіграють рекурсивні тактики, коли після опису одного об'єкта автор переходить до опису другого і, використовуючи інформацію про цей другий об'єкт, знову повертається до першого. Таким чином, інтенсивність позиційної інтерференції термінів, виражена кількістю та протяжністю спільних фрагментів Т-ланцюжків, може слугувати за міру комунікативної складності тексту.

Для перевірки цієї гіпотези проаналізовано 25 англомовних статей трьох категорій: науково-популярних навчального характеру, теоретичних оглядових та теоретичних монографічних (таких, що містять результати оригінальних досліджень). Ми виходили з того, що найпростішими в комунікативному аспекті є навчальні тексти, автори яких викладають матеріал максимально послідовно, залишаючи осторонь окремі деталі, не експлікуючи менш важливі зв'язки між поняттями тощо. Оглядові статті складніші за навчальні, але їм також властива чітка послідовність та відносна простота викладу.

З кожного тексту відібрано по 20–40 термінів, які автор вважав найсуттєвішими для характеристики теми та її підтем, а саме терміни, присутні у назвах розділів, підрозділів, параграфів. Було збережено порядок їхньої появи у цих назвах. Таким чином зафіксовано задекларований автором порядок висвітлення окремих підтем. Кожній підтемі відповідає один або кілька мікрореферен-

⁸ Глобальний референт розуміємо як організовану сукупність спеціальних понять — мікрореферентів.

тів, перелік яких назвемо авторським профілем тексту⁹. Авторський профіль виконує дві функції: він певною мірою характеризує як тему тексту, так і його тематичну структуру, принаймні послідовність розгортання теми.

Для кожного з текстів було підраховано кілька показників. Перший з них — середній локальний зв'язок, його визначаємо як середню довжину відрізка тексту, спільного для двох суміжних Т-протяжностей.

Він безпосередньо характеризує інтенсивність позиційної інтерференції термінів у тексті. Для спрощення підрахунків беремо до уваги інтерференцію лише суміжних Т-протяжностей, наприклад, *fuselage* та *landing gear*, *landing gear* та *wheel*, *wheel* та *brake* тощо.

Термін *fuselage* вперше з'являється у першому реченні, востаннє — у восьмому, *landing gear* уживається від третього до восьмого речення. Отже, спільний відрізок тексту налічує шість речень. Для Т-протяжностей *landing gear* та *wheel* спільними є п'ять речень, для *wheel* та *brake* — два (останній термін має усього два входження). У цій частині тексту кожна наступна Т-протяжність є частиною попередньої, тобто тут спостерігаємо відношення включення. Це відбиває специфіку зв'язків між відповідними референтами: кожний наступний об'єкт є частиною попереднього. Звідси випливає, що глобальна тема тексту має ієрархічну структуру.

Т-протяжності термінів *biaxiality ratio* та *finite element* ілюструють інший тип відношення — перетин. Перша протяжність охоплює речення 18–36, друга — 13–24. Спільним є відрізок тексту з вісімнадцятого до двадцять четвертого речення.

У подальшому для узагальнення та порівняння результатів обстеження текстів різної довжини замість номера речення, в якому вжито термін, або кількості речень у відрізку використовуватимуться так звані унормовані величини. Довжина будь-якого тексту приймається за одиницю. Розташування терміна в тексті визначається як відстань відповідного речення від початку тексту за формулою SN_i/SN_{max} , де SN_i — номер i -го речення, що містить термін, SN_{max} — номер останнього речення, тобто обсяг тексту. У цьому випадку відстань десятого речення в тексті загальним обсягом п'ятдесят речень, так само як тридцятого речення в тексті обсягом сто п'ятдесят речень, дорівнює 0,2. Відповідно визначається й довжина фрагмента тексту.

Результати спостереження представлені у табл. 2. В останньому стовпчику наведено інтервал, в якому з 95% ймовірністю наявний середній локальний зв'язок інших текстів відповідної категорії.

Таблиця 2

| Локальний зв'язок у текстах різних категорій | | | |
|--|--------------------|-------------------|-----------------------|
| Категорія тексту | Кількість термінів | Локальний зв'язок | |
| | | середній | 95% довірчий інтервал |
| навчальний | 146 | 0,073 | 0,048 – 0,098 |
| оглядовий | 220 | 0,190 | 0,155 – 0,225 |
| монографічний | 270 | 0,244 | 0,207 – 0,281 |

Інтенсивність позиційної інтерференції термінів, мірою якої є локальний зв'язок, найбільша у монографічних текстах і найменша в навчальних. За даними статистичного аналізу (t -тесту), ймовірність того, що відмінність між ве-

⁹ Скороходько Э. Ф. Роль системо- і текстообумовлених характеристик терміна в частотном индексировании научных текстов // Научно-техническая информация. Сер. 2. Информационные процессы и системы. — 2002. — № 8. — С. 1–6.

личинами локального зв'язку навчальних та оглядових текстів є випадковою, не перевищує 0,000001, а оглядових та монографічних — 0,039808. Звичайно, існує ймовірність того, що в інших навчальних текстах інтенсивність локальної позиційної інтерференції термінів виявиться більшою, ніж у теоретичних оглядових, а в оглядових — більша, ніж у монографічних, проте така ймовірність мала.

Другий показник — коефіцієнт рангової кореляції Спірмена між задекларованим та фактичним порядком опису мікрореферентів, тобто між порядком появи термінів у назвах розділів та підрозділів і порядком їхнього вживання в тексті. Рангову кореляцію використовують, якщо аналізуються не кількісні, а порядкові характеристики. У нашому випадку послідовність появи термінів у назвах структурних компонентів тексту порівнюється з послідовністю появи відповідних термінів у тексті. Коефіцієнт рангової кореляції визначає, якою мірою автору вдається додержуватися бажаної послідовності опису. Складність когнітивної структури наукового тексту змушує постійно робити відступи: ведучи мову про певний референт, доводиться залучати необхідні дані, що стосуються інших референтів. Тому коефіцієнт кореляції, характеризуючи ступінь послідовності опису, одночасно вказує на інтенсивність позиційної інтерференції термінів.

Було використано два варіанти кореляційного аналізу. У першому послідовність термінів у авторському профілі порівнювалась із позиціями в тексті відповідних медіан. Наприклад, перші три розділи статті *Airplanes* (рис. 1А) мають назви *fuselage*, *landing gear*, *wheel*. Їм присвоюються ранги, відповідно 1, 2, 3. Ці терміни зустрічаються в таких реченнях: *fuselage* — 1, 2, 4, 8; *landing gear* — 3, 8; *wheel* — 3, 4, 5, 6, 7. Медіани (центральні точки Т-ланцюжків) дорівнюють: *fuselage* — 3, *landing gear* — 5,5, *wheel* — 5; вони ж отримують ранги згідно з їхньою позицією в тексті, тобто відстанню від його початку: *fuselage* — 1, *landing gear* — 3, *wheel* — 2 (якщо назва розділу містить кілька спеціальних термінів, вони отримують один ранг). Таким чином, у супереч намаганню автора спочатку вести мову про шасі, а потім — про колеса, він практично був позбавлений такої можливості, оскільки відповідні референти тісно пов'язані. Тому входження терміна *wheel* вклинюється в Т-ланцюжок терміна *landing gear*. Коефіцієнт кореляції коливається від нуля (повна відсутність зв'язку) до одиниці (максимальний зв'язок). У цьому прикладі він дорівнює 0,5, що свідчить про певний відступ від задекларованої послідовності опису.

В іншому варіанті послідовність термінів у авторському профілі порівнювалась з позиціями усіх входжень відповідних термінів (тобто тут враховувався весь Т-ланцюжок). За ранг входження приймається номер речення, яке містить термін. У наведеному прикладі коефіцієнт кореляції дорівнює 0,25. Менше значення коефіцієнта свідчить, що другий варіант аналізу, враховуючи всі входження, чутливіший до інтерференції термінів.

У текстах усіх категорій зафіксовано позитивний, але менший за одиницю коефіцієнт кореляції (табл. 3, *p*-значення — це ймовірність помилки, тобто того, що в інших текстах відповідної категорії кореляція буде відсутня).

Таблиця 3

| Категорія тексту | Кореляція між задекларованим і фактичним порядком опису референтів | | | | | |
|------------------|--|----------------------|--------------------|--------------------------|----------------------|--------------------|
| | Перший варіант | | | Другий варіант | | |
| | кількість термінів | коефіцієнт кореляції | <i>p</i> -значення | кількість терміновживань | коефіцієнт кореляції | <i>p</i> -значення |
| навчальний | 164 | 0,886 | 0,000000 | 2528 | 0,866 | 0,000000 |
| оглядовий | 228 | 0,670 | 0,000000 | 3111 | 0,402 | 0,000000 |
| монографічний | 265 | 0,540 | 0,000000 | 3899 | 0,038 | 0,018545 |

Отже, в усіх трьох категоріях задекларований автором порядок представлення мікрореферентів (підтем) лише частково збігається з тим, який реально існує в тексті. У навчальних текстах розбіжності виявилися меншими, у монографічних — більшими. Аналіз за другим варіантом показав, що в монографічних текстах кореляція незначна. Це свідчить, зокрема, про інтенсивну позиційну інтерференцію термінів.

Рис. 2 ілюструє інтерференцію в типовій статті монографічного типу¹⁰ (зірочками позначені нерегулярні входження). Стаття налічує 588 речень, авторський профіль — 40 термінів. Середній локальний зв'язок між суміжними термінами авторського профілю в цій статті становить 0,266, коефіцієнти кореляції: перший варіант — 0,348, другий — 0,045.

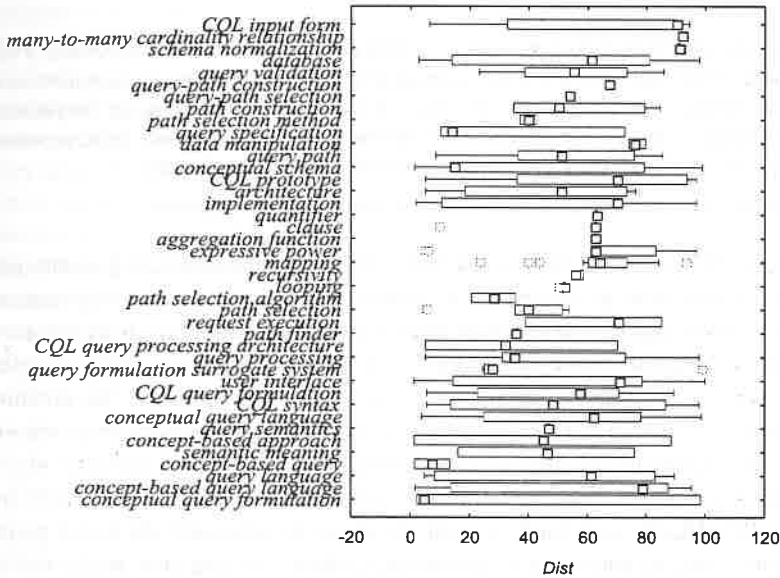


Рис. 2. Позиційна інтерференція термінів у науковій статті

Таким чином, результати експерименту свідчать на користь обох висунутих гіпотез: інтенсивність позиційної інтерференції термінів, тобто кількість та протяжність спільних фрагментів Т-ланцюжків, віддзеркалює комунікативну (а опосередковано — й когнітивну) складність наукового тексту; складність комунікативної структури тексту, яка виявляється як у інтенсивності позиційної інтерференції термінів, так і в порушенні послідовності викладу, у навчальних текстах найменша, у монографічних — найбільша. З цього випливає, зокрема, що інтерференція термінів може виконувати функцію стилеметричного показника.

Е. Ф. SKOROKHOD'KO

POSITIONAL TERM INTERFERENCE AS A STRUCTURAL CHARACTERISTICS OF A SCIENTIFIC TEXT

The positional interference of terms in a scientific text, i.e. the presence of different terms in the same stretch of sentences, is regarded as a demonstration of the relations between corresponding concepts. It is shown that the strength of this interference correlates with the communicative complexity of a text.

Keywords: term, scientific text, communicative structure of a text.

¹⁰ Enriching the conceptual basis for query formulation through relationship semantics in databases // Information Systems.— 2001.— Vol. 26. N. 6.— P. 445–475.