

## КОМП'ЮТЕРНА ЛІНГВІСТИКА

О. Багумян\*

(Київ)

УДК 161.2.81'322.5

## ТИПИ ТЕКСТОВОЇ СЕГМЕНТАЦІЇ: ЛІНГВОКОГНІТИВНИЙ АСПЕКТ

*In the article we have focused primarily on the ways of modern English texts segmentation in applied aspects. Having described briefly a variety of textual segmentation, we can differentiate the basic textual units and possible reference relations among them. Each coherence relation described below is associated with one or more constraints that must be met for it to hold.*

Слово текст (лат. *textus*) означає тканину, сплетення, об'єднання. Текст – це об'єднана за змістом послідовність знакових одиниць, основними властивостями якої є зв'язність і цілісність. Така послідовність знаків визнається одиницею вищого рівня, що є інформаційним і структурним цілим [1: 7—9]. Основними завданнями при вивченні структури тексту є визначення критеріїв текстових одиниць, їхнє вицленування; визначення їхнього статусу, суттєвих характеристик; визначення мовного оформлення цих одиниць, міжфразових зв'язків і відносин між ними [3: 7—19].

Сьогодні існує кілька типів сегментації тексту: *формальне, об'ємно-прагматичне (композиційне, композиційно-стилістичне), структурно-смісловне (семантико-структурне, об'єктивно-суб'єктивне, лінійне-нелінійне, тематичне) членування тексту.*

Таке різноманіття типів текстової сегментації пояснюється тим, що текст – це певна глобальна структура, що містить глибинну і поверхневу структури. Під *глибинною структурою* ми розуміємо певне семантичне чи понятійне утворення, що є відправною точкою при породженні поверхневих структур. *Поверхнева структура* – лінгвістична форма, у яку втілена глибинна структура. Глибинну структуру можна порівняти з програмою, що диктує вибір мовних засобів. Однак відносини між глибинною і поверхневою структурами не вичерпуються відносинами підпорядкування. Поверхнева структура може впливати на глибинну структуру, актуалізувати її [9: 56—58].

При *формальному членуванні* тексту досліджується лише поверхнева структура. У наявності виявляється мінімум вихідної інформації – границі речень, установлені розділовими знаками та деякими сполучниками, якщо перед ними немає коми. Сегментування за пунктуаційними знаками є найпростішою і разом з тим лінгвістично обгрунтованою дією. Саме слово «препинание» (рос.) означає «зупинка», «перерва», а розділові знаки усередині текстової тканини розділяють її частини й вказують на співвідношення між ними. Далі йде розчленування речення на гіпотетичні частини – сегменти, границями яких є дієслова, розділові знаки й сполучники (крім розділових знаків і сполучників між прикметниками, що визначають певний іменник). На цьому ж етапі встановлюються зв'язки між словами всередині фрагментів речення [6: 43—46].

Процедури подальшого сегментування неоднакові для різних типів складних речень: складні сполучникові речення сегментуються за допомогою сполучників і сполучникових слів; у безсполучникових реченнях сегментування речення відбувається за наявності в ньому предикативної пари з урахуванням знаків пунктуації і збігається з членуванням його на предикативні одиниці. Однак основна структурна одиниця при формальному (автоматичному) аналізі тексту – це речення (*ibid.*).

При *семантико-структурній сегментації* тексту виділяються дві основні семантико-структурні одиниці – висловлювання (реалізоване речення) і надфразова

---

\* © О. Багумян, 2006

єдність (ряд висловлень, об'єднаних семантично і синтаксично в єдиний фрагмент). Надфразові єдності в свою чергу об'єднуються в більш великі фрагменти-блоки [3: 21—24], що забезпечує цілісність тексту завдяки реалізації дистантних і контактних смислових і граматичних зв'язків.

При *композиційній сегментації* розглядаються одиниці якісно іншого плану – абзаци, параграфи, глави, розділи, підзаголовки й ін. Ці одиниці вичленюються візуально на формальному рівні (шрифти, символи, пропуски і т. п.), а на семантичному це означає введення нової теми чи ідеї. І.Р.Гальперін визначав такий тип членування тексту як об'ємно-прагматичний, і пояснював, що цей тип сегментації дозволяє збільшити текстову значущість певної фрази (фраз) за рахунок її графічного виділення, висування в самостійний абзац [5: 51].

Одиниці семантико-структурного й композиційного рівня перебувають у взаємозв'язку й взаємозумовленості; в окремому випадку вони навіть у «просторовому» відношенні можуть збігатися, накладаючись один на одного, наприклад, міжфразова єдність і абзац, хоча при цьому вони зберігають свої власні відмінні ознаки. Синонімом поняття „надфразова” єдність є структурно-смислова одиниця тексту – складне синтаксичне ціле. Якщо складне синтаксичне ціле дорівнює повному тексту твору, то воно може мати (чи не мати) *заголовок*, що визначає основний суб'єкт чи об'єкт тексту. Складне синтаксичне ціле звичайно складається з трьох частин: *зачину, розгортання* – термін Т.І.Сільман, *кінцівки*. Однак складне синтаксичне ціле може бути двоскладовим, тому що можлива відсутність зачину чи кінцівки.

Надфразова єдність (чи складне синтаксичне ціле) організується через темарематичну послідовність. У темарематичній послідовності здійснюється покорова тематизація реми. Єдність теми можна розглядати в межах мікротемати й теми всього твору. Перехід від однієї теми до іншої (мікротемати) є границями надфразових єдностей [7: 8—22].

Деякі дослідники намагаються визначити більші одиниці тексту, ніж НФС, у єдиний предикативно-релятивний комплекс. У складі такої логіко-семантичної цілісної системи надфразові єдності, відрізняючись лише відносною самостійністю, функціонують як засоби передавання різноманітної інформації й по-різному взаємодіють один з одним у розгортанні загального комунікативного смислу. Принцип побудови цієї комплексної одиниці в її мовному оформленні визначається не кількістю надфразових єдностей, що до неї входять, а тими відносинами, що виникають між ними в тексті, характером організації інформації [2: 255—258].

У цілому предикативно-релятивний комплекс можна охарактеризувати як складну багатокomпонентну ієрархічну систему протипоставлених і взаємозалежних відносин для виявлення в тексті додаткового смислу [5: 61].

Семантико-синтаксичне членування збігається з членуванням іншого рівня – композиційно-стилістичним. Перше здебільшого визначається прагматичною настановою самого тексту, друге – цілком обумовлено прагматичною настановою автора. Що стандартніший текст за своєю мовною організацією, то більше зближаються ці рівні членування (синтаксичне й абзацне), що менш стандартний текст, то більше між ними розбіжностей, а іноді й протиріч.

Абзац є одиницею *композиційно-стилістичного членування* тексту й розглядається як частина тексту, що розташована між двома відступами. Абзац і міжфразова єдність (складне синтаксичне ціле) у чомусь схожі за зовнішніми ознаками, але не тотожні по суті. Складне синтаксичне ціле – це темарематична послідовність, що відкривається фразо-зачином (чи стрижневою фразою, що містить смисл усього цілого). Саме фрази-зачини складних цілих, будучи стягнутими разом, утворюють смислову канву тексту. В абзаці може й не бути зачину як такого. Стрижнева фраза абзацу (головна в тематичному, логічному, змістовному плані) може стояти на початку абзацу, наприкінці абзацу чи сама виступати в ролі окремого абзацу. Більш того, в абзаці

може бути кілька стрижневих фраз, якщо він великий за обсягом і вміщає в собі ряд тема-рематичних послідовностей. Абзац може розірвати одне тема-рематичне об'єднання (одне складне синтаксичне ціле). Обсяг і структура абзацу цілком пов'язані з волею автора, його настановою, його особистими пристрастями, особливою манерою написання [3: 26—42].

В основі *тематичної сегментації* тексту лежить ідея про те, що семантичний простір тексту – це ієрархія тем, семантичне представлення яких здійснюється за допомогою лексичного матеріалу аналізованого тексту. При тематичній сегментації тексту здійснюються операції з виділення лексичних ядер висловлень, установлення зв'язків між ними, об'єднання їх у теми й виявлення зв'язків уже між лексичними ядрами окремих тем [2: 225—265].

Якщо в тексті переважає принцип лінійної зв'язності (тобто представлення явищ предметного світу відбувається у вигляді послідовного ланцюга дій і подій), членування тексту визнається об'єктивним, чи іконічним. Якщо автор, відповідно до власного бачення світу, виділяє якісь факти, що видаються йому особливо важливими, сегментацію тексту варто вважати суб'єктивною [9: 56—58].

До певної міри з об'єктивним і суб'єктивним членуванням тексту стикається запропоноване О.Л.Каменською лінійне / нелінійне членування. Якщо перше ґрунтується на поступових переходах від однієї теми до іншої (у межах єдиної макротемі), як це має місце в першому з наведених вище прикладів, можна говорити про лінійний розвиток тексту. Руйнування хронологічної послідовності подій, їхньої логіки руху дослідник розглядає як нелінійне членування тексту (*ibid.*).

Текстова сегментація має теоретичне, практичне і прикладне значення в сучасній лінгвістиці. Членування тексту є одним з початкових етапів його аналізу як «зверху», так і «знизу», дослідження загальних властивостей тексту, таких як цілісність, зв'язність і відокремленість тексту та їхня взаємодія в плані забезпечення характеристик тексту як цілого. Можливий також і зворотний шлях – дослідження формування тексту, процесу його «породження», коли дослідження починається з аналізу мовних факторів, даних у тексті, а закінчується – текстовими одиницями.

Під час вивчення текстових одиниць установлюються й класифікуються зв'язки між ними (нанизування, зчеплення, розщеплення, склеювання, обрамлення, розгортання та інші), що розширює теоретичні основи для наступного етапу вивчення тексту (В.Г.Борботько), наприклад, для градації зв'язків на семантично сильні й слабкі (Н.І.Теплицька), виділення явищ парцеляції й інтеграції в межах тематично зв'язаних відрізків тексту [4: 36] тощо.

На основі сегментації тексту здійснюється класифікація текстів, адже поняття «текст» може бути застосовано не тільки щодо цілісного літературно оформленого твору, але й до його частини, достатньо самостійної з погляду мікротеми та мовного оформлення. Так, можна говорити про текст глави, розділу, параграфа, текст вступу, висновку та інші [1: 7—9].

Якщо текстова композиційна сегментація проводиться паралельно з тематичним членуванням тексту, то такий аналіз дає змогу встановити тематичні блоки в тексті з виходом у поверхневі структури тексту [2: 225]. Результати дослідження доцільно враховувати при сприйнятті тексту адресатом: цей процес буде набагато ефективнішим, тому що «розмір частини звичайно розрахований на можливості читача сприймати обсяг інформації «без втрат» [5: 51].

Застосування позиційних методів для виконання операцій з автоматичної обробки текстів (зокрема, для індексування й реферування) також прив'язано до композиційної структури тексту. Воно спирається на задане автором зовнішнє членування тексту, де виділяються заголовки, окремі заголовки, параграфи, абзаци, виноска та інше. Вихідною передумовою використання таких методів слугує положення про те, що інформаційна значущість лінгвістичної одиниці визначається її позицією в певному тексті. До

інформаційно важливих позицій у документі належать, насамперед, заголовок, заголовки окремих частин, висновки, резюме, вступ. Відповідно до цього здійснюється вибір ключових речень, що відносяться до основних елементів змісту документа [8: 317].

Таким чином, членування тексту може бути як суб'єктивним, тому що воно завжди інтенціональне й екстенціональне, так і об'єктивним, оскільки обумовлене необхідністю відображення світу в його упорядкованості й побудові. Сегментація тексту пов'язана з характером людського мислення, що включає одночасно взаємодоповнювальні операції аналізу й синтезу інформації, що надходить. Єдність формальної, змістовної та комунікативної сторін тексту відбивається в поєднанні типів текстового членування.

#### Література

1. Аспекты общей и частной лингвистической теории текста. – М.: Наука, 1982.
2. Бабенко Л.Г. Филологический анализ текста. Основы теории, принципы и аспекты анализа. – М.: Академический Проект; Екатеринбург: Деловая книга, 2004.
3. Валгина Н.С. Теория текста. – М.: Логос, 2004.
4. Ванников Ю.В. Синтаксические особенности русской речи (явление парцелляции). – М.: Наука, 1969.
5. Гальперин И.Р. Текст как объект лингвистического исследования. – М.: Наука, 1981.
6. Грязнухина Т.А., Дарчук Н.П. и др. Синтаксический анализ научного текста на ЭВМ: Монография. – К.: Наукова думка, 1999.
7. Папина А.Ф. Текст: его единицы и глобальные категории. – М.: Едиториал УРСС, 2002.
8. Рожественский Ю.В. Лекции по общему языкознанию. – М.: Добросвет, 2000.
9. Тураева З.Я. Лингвистика текста. – М.: Просвещение, 1986. – 127 с.

*Лариса Бондарчук*, к.філол.н.,\*  
Київський національний економічний  
університет ім. В. Гетьмана (Київ)  
УДК 161.2.81'322.373

### ОСОБЛИВОСТІ ЛОКАЛІЗАЦІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ УКРАЇНСЬКОЮ МОВОЮ

*У статті актуалізовано проблему українізації програмного забезпечення в Україні, проаналізовано її передумови, розглянуто основні термінопоняття цього напрямку лінгвістичного дослідження. Відзначено, що основними системами і способами адаптації української мови в міжнародне комп'ютерне середовище є транслітерація, транскрипція й спеціальні програми трансліту.*

На сучасному етапі Україна робить великі кроки на шляху входження до інформаційного простору, її починають пізнавати як державу з новими технологіями національного та всесвітнього масштабу. Це означає, що вона має гідно представляти себе на європейському та світовому рівні. Країна повинна мати програмне забезпечення високого рівня, а також свою мову в інформаційному просторі. Оскільки програмне забезпечення в Україні переважно є іншомовним, виникає потреба локалізації програмного продукту, перекладу його українською. Мова повинна бути бездоганною, оскільки на технологічно-прикладному рівні з нею знайомляться представники багатьох країн, особливо у всесвітній мережі Інтернет.

---

\* © Л. Бондарчук, 2006