

КОРПУС ТЕКСТОВ УКРАИНСКОЙ ПЕРИОДИКИ

1. Корпус текстов украинской периодики (КТУП) реализуется как часть проекта по созданию синхронно-диахронного корпуса текстов украинского языка XX в.

Целью КТУП является получение объективной картины современного украинского языка периодических изданий и, учитывая представленность в газетных источниках всех стилей и жанров языка, получение в какой-то степени общей картины украинского языка.

КТУП как подкорпус является автономным машиночитаемым массивом текстов периодических изданий, написанных на украинском языке за период с 1985 до 2000 г. С учетом характера истории Украины в конце XX в. в этом временном отрезке далее выделяются следующие хронологические периоды:

- а) 1985 - 1991 гг.;
- б) 1991 -1994 гг.;
- с) 1994 - 2000 гг..

В процентном отношении КТУП должен составлять приблизительно 20-25% от всего синхронно-диахронного корпуса, структура которого имела бы следующий вид (первоначальный общий объем - 10 млн. слов):

• научные тексты	30//25%	3000//2500 тыс.сл.
• тексты художественной литературы	30//35%	3000//3500 тыс.сл.
• тексты периодических изданий	25//20%	2500//2000 тыс.сл.
• диалектные тексты	10//15%	1000//1500 тыс.сл.
• тексты устной речи	5%	500 тыс.сл.
всего:	100%	10 млн. сл.

2. Одной из важнейших характеристик текстового материала, входящего в КТУП, является стилистико-тематическая характеристика. Стилистически тексты должны репрезентировать все стили украинского языка с учетом так называемого высокого, среднего и низкого уровня фактического материала. К высокому уровню относятся профессиональные периодические издания, научно и научно-популярная периодика (*Мовознавство, Наука і життя*), а также культурологические издания, рассчитанные на подготовленного читателя (*І, Сучасність*). Средний уровень составляют периодические издания общественно-политического и культурного характера (*Урядовий кур'єр, Жінка, Сільські вісті*). Низкий уровень - популярные журналы и газеты сенсационного, бульварного характера, частично «желтая пресса» (*Товарний вісник,*

Експрес). Тематически предусмотрено следующее общее разграничение текстов периодических изданий:

- а) общетематические;
- б) специально-тематические.

К общетематическим относятся ежедневные издания, главным образом периодика низкого уровня (~20% всех текстов). Специально-тематические охватывают разные области человеческого опыта и знаний; соответственно речь идет о текстах среднего и высокого уровня, которые извлекаются из следующих изданий:

- общественно-политических;
- политико-экономических;
- экологических и географических;
- художественных;
- исторических;
- философских и религиозных;
- литературо- и языковедческих;
- естественнонаучных.

3. При создании КТУП необходимо учесть территориальный аспект. Этот аспект влияет на формирование репрезентативного корпуса украинского языка, поскольку в зависимости от территориальной характеристики конкретного текста одного и того же стиля и жанра обнаруживаются территориальные специфические признаки украинского общенародного языка. А наличие текстов украинской диаспоры в КТУП даст возможность сопоставить и проанализировать языковой материал с точки зрения заимствований или калек из языка-окружения и эксплицировать «слабые / сильные места» языка, где происходят или нивелиационные процессы, или же процессы обогащения системы. В территориальном отношении в КТУП предусмотрена следующая структура:

I. Государственные периодические издания.

1. Печатный орган Верховной Рады Украины (*Голос України*).
2. Печатный орган Правительства (*Урядовий кур'єр*).

II. Общеукраинские периодические издания.

III. Региональная периодика (*Центр, Северо-запад, Юг, Восток, Диаспора*).

4. Можно предположить, что представленность оригинального и переводного текстов в границах одного корпуса лучше отобразит языковую картину, однако в КТУП, с учетом его функционального назначения, целесообразно остановиться на оригинальных текстах, и основополагающими (но не абсолютными) критериями отбора текстов в языковом аспекте считать:

- текст написан на современном украинском литературном языке;
- текст оригинальный (не переводной);

- текст написан автором в возрасте свыше 18 лет (для взрослых);
- текст написан в Украине или в диаспоре.

5. Традиционно корпусный текст состоит из двух частей: поля библиографического описания и поля собственно текста. Система библиографического описания предполагает экспликацию следующей информации:

1. Название: *название издания - газеты, журнала, бюллетеня, вестника.*
 2. Автор/авторы: *имя, фамилия.*
 3. Заглавие: *название конкретного текста.*
 4. Том: *№ тома, из которого взят текст.*
 5. Выпуск: *№ выпуска.*
 6. Издательство: *название издательства или издатель.*
 7. Место издания: *информация о местонахождении редакции периодического издания.*
 8. Дата: *полная дата периодического издания — месяц /день /год.*
 9. Страницы: *№ страницы / строки цитаты (если исходный документ не является электронным).*
 10. Тип текста: *прозаический/стихотворный.*
 11. Предметная характеристика текста: *указание на предметную принадлежность текста.*
 12. Стил ь текста: *научный, научно-популярный, научно-дидактический; публицистический; художественный; разговорный.*
 13. Назначение текста: *указание на аудиторию, для которой предназначен текст.*
 14. Количественные параметры: *количество словоединиц.*
 15. Файловая информация: *указание, где находится текст на магнитных носителях.*
 16. Иная информация.
- Собственно текст предполагает лингвистическую аннотацию.

А.Н.Дмитриева(Санкт-Петербург)

НЕКОТОРЫЕ СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ МАКЕДОНСКОГО ЯЗЫКА (АНАЛИЗ ТЕКСТА НА ФОНЕТИЧЕСКОМ УРОВНЕ)

Фонетическая система. Македонский язык наряду с болгарским, сербохорватским и словенским принадлежит группе южных славянских языков. Его фонетическая система включает 31 фонему: 5 гласных (/a/, /e/, /i/, /o/, /u/), 7 сонантов (/j/, /l/, /lʲ/, /m/, /n/, /nʲ/, /r/), 19 шумных согласных, из которых 18 противопоставлены по глухости-звонкости: взрывные /p/-/b/, /t/-/d/, /k/-/g/, /kʲ/-/gʲ/ шелевые /f/-/v/, /s/-/z/, /sʲ/-/zʲ/, аффрикаты /c/-/dz/, /cʲ/-/dzʲ/ и непарный глухой шелевой /x/. Существует также слогообразующий вариант /r/ (далее /эг/).