

ЧАСТОТИ ПОВТОРЮВАНOSTІ БУКВ І БІГРАМ У ВІДКРИТИХ ТЕКСТАХ УКРАЇНСЬКОЮ МОВОЮ

Упродовж тривалого часу дешифруванню криптограм допомагає частотний аналіз, суттю якого є вивчення статистичних закономірностей появи символів та їх сполук у вихідних та шифрованих повідомленнях. Саме щоб ускладнити частотний аналіз, в криптографії з'явилися шифри, які приводять до рівномірного розподілу символів у криптограмі. Сьогодні принципи частотного аналізу широко застосовуються в програмах для підбору паролів і дають змогу на декілька порядків скоротити час пошуку.

Очевидно, частотний аналіз вимагає насамперед еталонних частот повторюваності букв абетки, на якій написані відкриті тексти, і частот повторюваності n -грам ($n \geq 2$). Для російської, англійської та майже всіх європейських мов середньостатистичні частоти повторюваності букв, біграм, триграм можна знайти в літературних джерелах [1-7]. На жаль, для української мови в літературі наведені лише частоти повторюваності букв [8-10]. Тому мета цієї роботи – дослідити частоти повторюваності букв і біграм української мови на основі вибраних випадково текстів української мовою.

Тексти вибирались зі сторінок україномовних сайтів і належать до п'яти стилів сучасної української мови: розмовно-побутового, художнього, наукового, публіцистичного та ділового. Загальний об'єм текстів склав 900123 слів (приблизно 6 Мб). Перед розрахунками проводилась спеціальна обробка текстів за сформованими правилами. Так, вилучались: імена, прізвища, власні назви, слова написані на латиниці, дати, наявні в тексті фрагменти окремих слів (неповні слова), цифрові дані, скорочення одиниць вимірювання будь-яких величин, аббревіатура; перенесення слів, будь-які розділові знаки; відступи на початках абзаців.

Оскільки в текстах на будь-якій мові найбільш вживаним є пропуск між словами, то на першому етапі частоти букв і біграм підраховувались з врахуванням пропуску, а на другому етапі – без його врахування. Відзначимо також, що на першому етапі при оформленні текстів між кожними двома словами було залишено тільки один пропуск, аби запобігти у тексті виникненню біграми, складеної з двох пропусків.

Для фільтрації цифрових даних та розділових знаків, а також для підрахунку частот літер, біграм, триграм та n -грам була розроблена програма «Grammz». Обчислення проводились на 4-х ПЕОМ, оснащених процесорами Intel Core 2 Duo з частотою 3 ГГц та ОЗУ 4096 МБ. Обробка текстів та проведення усіх підрахунків зайняла приблизно одну годину.

Результати обчислених середньостатистичних частот букв та пропуску між словами в українській мові наведені у таблицях 1 і 2.

Таблиця 1

Середньостатистичні частоти букв та пропуску між словами в українській мові

—	0,138	і	0,044	д	0,027	г	0,013	ж	0,007
о	0,086	р	0,043	л	0,027	ч	0,011	ю	0,008
н	0,068	е	0,042	п	0,025	х	0,011	є	0,005
а	0,064	с	0,037	з	0,020	ї	0,010	щ	0,004
и	0,055	к	0,033	я	0,019	ц	0,010	ф	0,003
в	0,046	м	0,029	ь	0,016	ш	0,005	г	0,000
т	0,045	у	0,027	б	0,013	й	0,009		

Таблиця 2

Середньостатистичні частоти букв без врахування пропуску між словами в українській мові

о	0,100	р	0,050	д	0,031	ч	0,012	ш	0,006
н	0,079	е	0,048	п	0,029	х	0,013	є	0,005
а	0,075	с	0,043	з	0,023	ц	0,012	ф	0,004

и	0,064	к	0,039	я	0,021	ї	0,011	щ	0,004
в	0,053	м	0,034	ь	0,018	й	0,010	г	0,000
і	0,052	л	0,032	б	0,015	ю	0,009		
т	0,052	у	0,032	г	0,015	ж	0,008		

Наочне уявлення про частоту букв дає діаграма повторюваності букв української абетки. Важливою характеристикою цього розподілу є наявність максимумів та мінімумів, завдяки чому можна умовно розділити букви на класи високої, середньої, низької і дуже низької частот їх використання:

- пропуск, О,А,Н – висока частота;
- И,Т,В,Е,Р,І,С,К,М – середня частота;
- Д,Л,У,П,Я,З,Ь,Г,Ч,Б,Х – низька частота;
- Ю,Ж,Й,Ї,Є,Ф,Ш,Щ – дуже низька частота.

Частота повторюваності букви Г в текстах українською мовою дорівнює $6 \cdot 10^{-5}$.

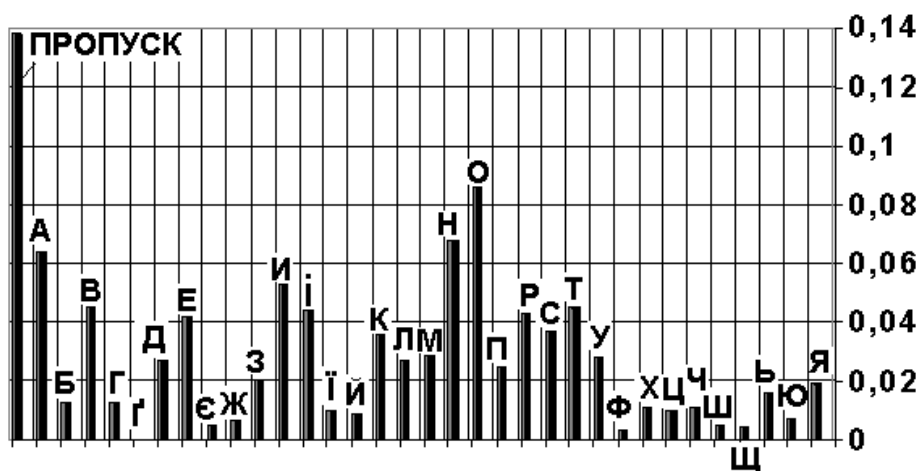


Рис.1 Діаграма частот повторюваності букв української абетки

Повний розподіл біграм за частотами наведений у таблиці 3. Символ __ відповідає пропуску між словами, буква у першому стовпці таблиці вказує першу букву біграми, буква у першому рядку – її другу букву (найбільші частоти виділені жирним шрифтом). Нерівномірність появи *n*-грам (і навіть слів) обумовлена характерними особливостями відкритих текстів українською мовою, а саме наявністю великої кількості повторень окремих фрагментів тексту: коренів слів, їх закінчень, суфіксів, загалом слів, окремих фраз. Так, для української мови такими звичними фрагментами є наступні біграми:

- з урахуванням пропуску між словами
И__ ; __П; __В; О__ ; А__ ; НА; І__ ; НО; СТ; __З; __Н;
- без урахування пропуску між словами
НА; НО; СТ; ОВ; КО; РО; НИ; ЕР; АН; ОМ; ПР.

Усі можливі біграми, які утворюються з букв української абетки, поділяються на два типи: припустимі *n*-грами, що зустрічаються хоч в одному якомусь тексті, та заборонені *n*-грами, що не зустрічаються в жодному тексті. Подібна класифікація *n*-грам використовується у детермінованих моделях відкритих текстів. Відзначимо, що деякі біграми, хоч і мають нульову частоту, але дуже рідко все ж таки зустрічаються в текстах українською. Наприклад, такими є біграми МГ ($1,68 \cdot 10^{-6}$), ЯБ ($6,72 \cdot 10^{-6}$), УЕ ($3,36 \cdot 10^{-6}$) та ін. Аби відрізнити ці рідкі біграми від дійсно заборонених, останні у таблиці позначені зірочкою.

Таблиця 3. Середньостатистичні частоти повторюваності біграм в українській мові із урахуванням пропуску між словами

	А	Б	В	Г	Ґ	Д	Е	Є	Ж
А	0,0000	0,0010	0,0030	0,0013	0,0000	0,0019	0,0000	0,0015	0,0008
Б	0,0014	0,0000	0,0001	0,0001	0,0000	0,0000	0,0013	0,0002	0,0000
В	0,0068	0,0000	0,0002	0,0000	*	0,0008	0,0017	0,0000	0,0004
Г	0,0020	0,0000	0,0000	0,0000	*	0,0000	0,0002	0,0000	0,0000
Ґ	0,0000	*	0,0000	*	*	0,0000	0,0000	0,0000	*
Д	0,0021	0,0003	0,0005	0,0001	0,0000	0,0001	0,0019	0,0000	0,0008
Е	0,0003	0,0004	0,0009	0,0006	0,0000	0,0016	0,0000	0,0000	0,0010
Є	0,0000	0,0000	0,0001	0,0000	0,0000	0,0002	0,0000	0,0000	0,0000
Ж	0,0008	0,0000	0,0000	0,0000	0,0000	0,0001	0,0021	*	0,0000
З	0,0051	0,0009	0,0013	0,0002	0,0000	0,0004	0,0004	0,0000	0,0000
И	0,0000	0,0003	0,0030	0,0003	0,0000	0,0007	0,0000	0,0002	0,0001
І	0,0003	0,0005	0,0046	0,0004	0,0000	0,0047	0,0000	0,0004	0,0004
Ї	*	0,0000	0,0002	0,0000	*	0,0000	*	*	0,0000
Й	0,0000	0,0002	0,0001	0,0000	0,0000	0,0002	0,0000	0,0000	0,0000
К	0,0035	0,0000	0,0011	0,0000	*	0,0000	0,0003	*	0,0000
Л	0,0037	0,0000	0,0000	0,0001	0,0000	0,0002	0,0030	0,0000	0,0000
М	0,0040	0,0000	0,0001	0,0000	0,0000	0,0000	0,0024	0,0000	0,0001
Н	0,0118	0,0000	0,0001	0,0000	0,0000	0,0004	0,0044	0,0001	0,0000
О	0,0000	0,0044	0,0090	0,0060	0,0000	0,0051	0,0002	0,0003	0,0018
П	0,0012	0,0000	0,0000	0,0000	*	0,0001	0,0033	0,0000	0,0000
Р	0,0065	0,0001	0,0004	0,0005	0,0000	0,0001	0,0056	0,0000	0,0004
С	0,0008	0,0000	0,0013	0,0000	*	0,0000	0,0010	0,0000	0,0000
Т	0,0065	0,0000	0,0023	0,0000	*	0,0000	0,0057	0,0001	0,0001
У	0,0002	0,0002	0,0021	0,0006	0,0000	0,0016	0,0000	0,0006	0,0004
Ф	0,0004	0,0000	0,0000	0,0000	*	0,0000	0,0003	0,0000	*
Х	0,0007	0,0000	0,0001	0,0000	*	0,0001	0,0000	*	*
Ц	0,0002	0,0000	0,0000	0,0000	*	0,0000	0,0027	*	*
Ч	0,0024	0,0000	0,0000	0,0000	*	0,0000	0,0015	*	*
Ш	0,0003	0,0000	0,0001	0,0000	*	0,0000	0,0009	*	*
Щ	0,0001	*	0,0000	*	*	0,0000	0,0006	*	*
Ь	0,0000	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Ю	0,0000	0,0000	0,0005	0,0000	0,0000	0,0003	0,0000	0,0003	0,0000
Я	0,0000	0,0000	0,0006	0,0005	*	0,0005	0,0000	0,0002	0,0002
—	0,0031	0,0044	0,0140	0,0021	0,0000	0,0077	0,0021	0,0007	0,0006

	З	И	І	Ї	Й	К	Л	М	Н
А	0,0013	*	0,0000	0,0011	0,0012	0,0028	0,0058	0,0035	0,0073
Б	0,0000	0,0007	0,0011	0,0000	0,0000	0,0001	0,0010	0,0002	0,0009
В	0,0002	0,0075	0,0052	0,0000	0,0000	0,0004	0,0013	0,0001	0,0030
Г	0,0000	0,0003	0,0007	*	*	0,0001	0,0009	0,0000	0,0003
Ґ	0,0000	0,0000	0,0000	0,0000	*	0,0000	0,0000	0,0000	0,0000
Д	0,0001	0,0022	0,0019	0,0000	0,0000	0,0006	0,0009	0,0002	0,0029
Е	0,0013	0,0000	0,0000	0,0001	0,0005	0,0027	0,0014	0,0024	0,0070
Є	0,0000	*	*	0,0002	*	0,0001	0,0000	0,0005	0,0001
Ж	0,0000	0,0007	0,0002	*	*	0,0001	0,0004	0,0000	0,0009
З	0,0000	0,0005	0,0005	0,0000	0,0000	0,0004	0,0004	0,0007	0,0022
И	0,0008	*	0,0000	0,0002	0,0025	0,0028	0,0016	0,0034	0,0033
І	0,0017	*	0,0000	0,0023	0,0022	0,0010	0,0022	0,0005	0,0033
Ї	0,0000	*	*	0,0007	0,0000	0,0002	0,0000	0,0002	0,0009
Й	0,0000	*	0,0000	0,0000	*	0,0001	0,0002	0,0001	0,0009
К	0,0000	0,0046	0,0024	*	*	0,0000	0,0011	0,0000	0,0006
Л	0,0001	0,0048	0,0030	*	*	0,0002	0,0000	0,0000	0,0001
М	0,0000	0,0047	0,0025	0,0000	*	0,0004	0,0002	0,0000	0,0005
Н	0,0001	0,0089	0,0061	*	*	0,0010	0,0000	0,0000	0,0061
О	0,0024	*	0,0000	0,0036	0,0002	0,0022	0,0031	0,0070	0,0048
П	0,0001	0,0011	0,0022	*	*	0,0000	0,0015	0,0000	0,0001
Р	0,0002	0,0057	0,0033	0,0001	0,0000	0,0013	0,0000	0,0017	0,0014
С	0,0000	0,0021	0,0011	*	0,0000	0,0014	0,0019	0,0001	0,0019
Т	0,0000	0,0070	0,0032	0,0000	*	0,0008	0,0001	0,0000	0,0013
У	0,0003	*	0,0000	0,0000	0,0001	0,0025	0,0015	0,0009	0,0007
Ф	0,0000	0,0000	0,0005	*	*	0,0000	0,0000	0,0000	0,0000
Х	0,0000	0,0002	0,0007	*	*	0,0000	0,0000	0,0000	0,0005
Ц	0,0000	0,0003	0,0042	*	*	0,0000	0,0001	0,0000	0,0001
Ч	0,0000	0,0022	0,0003	*	*	0,0001	0,0001	0,0000	0,0031
Ш	0,0000	0,0010	0,0004	*	*	0,0002	0,0002	0,0000	0,0004
Щ	0,0000	0,0001	0,0001	*	*	*	*	*	0,0000
Ь	0,0000	*	0,0000	0,0000	0,0000	0,0042	0,0000	0,0002	0,0021
Ю	0,0000	*	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Я	0,0006	*	0,0000	0,0000	0,0000	0,0024	0,0002	0,0010	0,0008
—	0,0106	0,0000	0,0048	0,0012	0,0020	0,0055	0,0011	0,0065	0,0104

	О	П	Р	С	Т	У	Ф	Х	Ц
А	0,0000	0,0010	0,0030	0,0034	0,0041	0,0004	0,0001	0,0017	0,0021
Б	0,0014	0,0000	0,0006	0,0004	0,0001	0,0029	*	0,0002	0,0000
В	0,0052	0,0004	0,0002	0,0016	0,0003	0,0010	0,0000	0,0001	0,0001
Г	0,0063	0,0000	0,0012	0,0000	0,0000	0,0005	0,0000	0,0001	0,0000
Г	0,0000	*	0,0001	*	*	0,0000	0,0000	*	*
Д	0,0053	0,0005	0,0008	0,0003	0,0001	0,0021	0,0000	0,0001	0,0000
Е	0,0008	0,0007	0,0087	0,0014	0,0015	0,0000	0,0002	0,0006	0,0006
Є	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Ж	0,0002	0,0000	0,0000	*	0,0000	0,0005	0,0000	0,0000	0,0001
З	0,0009	0,0005	0,0006	0,0001	0,0000	0,0008	0,0000	0,0000	0,0000
И	0,0000	0,0007	0,0017	0,0043	0,0031	0,0000	0,0001	0,0052	0,0012
І	0,0003	0,0001	0,0007	0,0022	0,0015	0,0000	0,0001	0,0003	0,0001
Ї	0,0000	*	0,0000	0,0000	0,0000	*	0,0000	0,0005	0,0000
Й	0,0005	0,0000	0,0000	0,0008	0,0001	0,0000	0,0000	0,0000	0,0000
К	0,0096	0,0000	0,0025	0,0004	0,0017	0,0023	*	0,0000	0,0008
Л	0,0033	0,0000	0,0000	0,0000	0,0001	0,0008	0,0000	0,0000	0,0000
М	0,0046	0,0010	0,0001	0,0003	0,0000	0,0026	0,0000	0,0000	0,0000
Н	0,0115	0,0000	0,0000	0,0018	0,0021	0,0020	0,0008	0,0000	0,0005
О	0,0001	0,0017	0,0064	0,0057	0,0021	0,0001	0,0002	0,0007	0,0008
П	0,0070	0,0000	0,0071	0,0002	0,0001	0,0004	0,0000	0,0000	0,0001
Р	0,0095	0,0001	0,0000	0,0007	0,0007	0,0021	0,0000	0,0003	0,0001
С	0,0015	0,0020	0,0000	0,0000	0,0115	0,0018	0,0002	0,0002	0,0004
Т	0,0051	0,0000	0,0028	0,0002	0,0005	0,0022	0,0000	0,0000	0,0001
У	0,0000	0,0008	0,0011	0,0012	0,0017	0,0000	0,0000	0,0002	0,0000
Ф	0,0014	0,0000	0,0003	0,0000	0,0000	0,0003	0,0000	*	0,0000
Х	0,0014	0,0000	0,0003	0,0000	0,0001	0,0002	0,0000	0,0000	0,0000
Ц	0,0001	0,0000	0,0000	0,0000	0,0007	0,0000	0,0000	0,0000	0,0001
Ч	0,0004	*	0,0000	0,0000	0,0000	0,0003	0,0000	0,0000	0,0000
Ш	0,0005	0,0000	0,0000	0,0000	0,0003	0,0003	0,0000	0,0000	0,0000
Щ	0,0028	*	*	*	*	0,0001	*	*	*
Ь	0,0015	0,0000	0,0000	0,0019	0,0004	*	0,0000	0,0000	0,0000
Ю	0,0000	0,0000	0,0000	0,0000	0,0022	0,0000	0,0000	0,0000	0,0000
Я	0,0000	0,0000	0,0002	0,0002	0,0009	0,0000	0,0000	0,0004	0,0000
—	0,0045	0,0159	0,0051	0,0095	0,0080	0,0038	0,0016	0,0009	0,0030

	Ч	Ш	Щ	Ь	Ю	Я	—
А	0,0011	0,0004	0,0001	*	0,0009	0,0001	0,0133
Б	0,0000	0,0001	0,0000	*	0,0000	0,0000	0,0003
В	0,0003	0,0001	0,0000	*	0,0000	0,0008	0,0078
Г	0,0000	0,0000	*	*	0,0000	0,0000	0,0002
Г	0,0000	0,0000	*	*	*	0,0000	0,0000
Д	0,0002	0,0000	0,0000	0,0001	0,0000	0,0003	0,0020
Е	0,0004	0,0001	0,0000	*	0,0001	0,0001	0,0060
Є	0,0000	0,0000	0,0000	*	0,0003	0,0000	0,0022
Ж	0,0000	0,0000	0,0000	*	0,0000	0,0000	0,0008
З	0,0000	0,0000	0,0000	0,0004	0,0000	0,0002	0,0032
И	0,0018	0,0004	0,0006	*	0,0000	0,0004	0,0161
І	0,0010	0,0008	0,0001	*	0,0004	0,0009	0,0116
Ї	0,0000	0,0000	0,0000	*	*	*	0,0070
Й	0,0000	0,0001	0,0000	*	0,0000	0,0000	0,0057
К	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0024
Л	0,0000	*	*	0,0044	0,0010	0,0023	0,0003
М	0,0008	0,0000	0,0000	0,0000	0,0000	0,0002	0,0044
Н	0,0001	0,0005	0,0000	0,0013	0,0005	0,0056	0,0018
О	0,0013	0,0005	0,0001	0,0000	0,0022	0,0002	0,0139
П	0,0000	0,0000	0,0000	*	0,0006	0,0001	0,0001
Р	0,0000	0,0005	0,0000	0,0002	0,0002	0,0007	0,0008
С	0,0000	0,0000	*	0,0038	0,0000	0,0033	0,0007
Т	0,0000	0,0000	0,0000	0,0050	0,0002	0,0005	0,0012
У	0,0006	0,0001	0,0001	*	0,0006	0,0000	0,0095
Ф	0,0000	0,0000	*	0,0000	0,0000	0,0000	0,0000
Х	0,0000	0,0000	*	*	*	*	0,0071
Ц	*	0,0000	*	0,0007	0,0002	0,0005	0,0001
Ч	0,0001	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002
Ш	0,0000	0,0001	*	0,0000	0,0000	0,0000	0,0002
Щ	*	*	0,0000	*	*	*	0,0001
Ь	0,0000	0,0005	0,0000	*	0,0000	0,0000	0,0045
Ю	0,0006	0,0000	0,0000	*	0,0002	*	0,0036
Я	0,0001	0,0000	0,0000	*	0,0001	0,0000	0,0098
—	0,0021	0,0006	0,0028	*	0,0000	0,0024	*

Цікавою для криптоаналізу є й таблиця 4 сполученості букв, що вказує найуживаніших її лівих та правих «букв-сусідів» за порядком спадання частот. Там же наведена доля голосних та приголосних у відсотках, розмічених попереду та позаду кожної букви

Таблиця 4

Сполученість букв української мови

Голосні	Приго- лосні	Ліворуч	Буква	Праворуч	Голосні	Приго- лосні
0,012	0,988	З,Т,Р,В,Н	А	__,Н,Л,Т,М	0,064	0,936
0,807	0,193	І,З,А,__,О	Б	У,О,А,Е,І	0,750	0,250
0,801	0,199	А,И,І,О,__	В	__,И,А,І,О	0,772	0,228
0,906	0,094	У,Е,А,__,О	Г	О,А,Р,Л,І	0,823	0,177
0,500	0,500	Б,__,	Ґ	Р	0,214	0,786
0,894	0,106	У,А,І,О,__,	Д	О,Н,И,А,У	0,678	0,322
0,004	0,996	Л,П,Н,Р,Т	Е	Р,Н,__,К,М	0,040	0,960
0,911	0,089	О,І,У,__,А	Є	__,Т,М,Ю,Д	0,197	0,803
0,737	0,263	__,Д,А,Е,О	Ж	Е,Н,А,__,Й	0,716	0,284
0,957	0,043	Е,А,І,О,__,	З	А,__,Н,В,О	0,558	0,442
0,000	1,000	Л,Р,Т,В,Н	И	__,Х,С,М,Н	0,017	0,983
0,001	0,999	Р,Ц,__,В,Н	І	__,Д,В,Н,І	0,109	0,891
0,994	0,006	І,А,__,І,О	Ї	__,Н,І,Х,В	0,205	0,795
0,995	0,005	Е,А,__,І,И	Й	__,Н,С,О,Л	0,204	0,796
0,581	0,419	Е,И,А,Ь	К	О,И,А,Р,__,	0,792	0,208
0,660	0,340	И,С,І,О,А	Л	И,Ь,А,О,І	0,859	0,141
0,853	0,147	Е,И,А,__,О	М	И,О,__,А,У	0,913	0,087
0,517	0,483	О,Н,Е,А,__,	Н	А,О,И,Н,І	0,800	0,200
0,011	0,989	Г,П,Р,К,Н	О	__,В,М,Р,Г	0,086	0,914
0,596	0,404	А,М,О,С,__,	П	Р,О,Е,І,Л	0,683	0,317
0,580	0,420	А,__,О,П,Е	Р	О,А,И,Е,І	0,813	0,187
0,700	0,300	І,А,И,О,__,	С	Т,Ь,Я,И,П	0,355	0,645
0,514	0,486	Ю,И,А,__,С	Т	И,А,Е,О,Ь	0,709	0,291
0,015	0,985	Т,К,М,Б,__,	У	__,К,В,Т,Д	0,063	0,937
0,416	0,584	Е,О,С,Н,__,	Ф	О,І,А,Е,Р	0,876	0,124
0,922	0,078	Е,О,__,А,И	Х	__,О,І,А,Н	0,740	0,260
0,625	0,375	О,К,И,А,__,	Ц	І,Е,Ь,Т,Я	0,832	0,168
0,886	0,114	І,А,О,И,__,	Ч	Н,А,И,Е,О	0,758	0,242
0,588	0,412	О,Р,Н,__,І	Ш	И,Е,О,Н,І	0,736	0,264
0,826	0,174	І,О,К,И,__,	Щ	О,Е,А,И,__,	1,000	0,000
0,000	1,000	Ц,Н,С,Л,Т	Ь	__,К,Н,С,О	0,142	0,858
0,640	0,360	У,П,А,Л,О	Ю	__,Т,Ч,В,Д	0,094	0,906
0,106	0,894	І,Л,__,С,Н	Я	__,К,М,Т,Н	0,030	0,970
0,676	0,324	Я,І,А,О,И	__	П,В,З,Н,С	0,160	0,840

Аналіз останньої таблиці підтверджує, що для української мови, як і для решти європейських мов, притаманне чергування голосних та приголосних. Якщо дослідити інші тексти, може бути присутня певна різниця в цифрах наведених частот букв, що пояснюється, по-перше, довжиною досліджуваного тексту, а по друге його тематикою. Наприклад, загалом мало жвивана буква Ф може стати досить частою в технічних текстах, бо використовується в таких словах, як функція, диференціал, дифузія, коефіцієнт і т. п. Ще більші відхилення від традиційного вживання окремих букв спостерігаються в деяких художніх творах, особливо у віршах.

Висновки. Проведені дослідження заповнюють пробіл в області частотного аналізу стосовно наявності ймовірностей біграм та сполучень найуживаніших букв української мови. Отримані експериментальні дані цілком можуть бути використані, наприклад, для формування послідовності перевірки комбінацій літер в процесі підбору пароля для кодованого тексту, щоб скоротити час пошуку.

Список літератури

1. Бабаш А.В., Шанкин Г.П. Криптография. Москва, СОЛОН-Р, 2002, 511 с.
2. Алферов А.П., Зубов А.Ю., Кузьмин А.С., Черемушки А.В. Основы криптографии. Москва, Гелиос, 2002, 480 с.
3. Яглом А.М., Яглом И.М. Вероятность и информация. Наука, гл.ред. физ.-мат. лит., Москва 1973, -511 с.
4. Пиотровский Р.Г. Информационные измерения языка, Л., «Наука», 1968, 237 с.
5. Яглом И.М., Добрушин Р.Л., Яглом А.М. Теория информации и лингвистика. Вопросы языкознания, 1960, №1, с. 100-110.
6. Лебедев Д.С., Гармаш В.А. О возможности увеличения скорости передачи телеграфных сообщений. Электросвязь, 1958, №1, с. 68-69.
7. Шеннон К. Предсказание и энтропия английского печатного текста. В кн. «Работы по теории информации и кибернетике» М., ИЛ, 1963, с. 669-686.
8. Вербіцький О.В. Вступ до криптології. Видавництво науково-технічної літератури, Львів 1998, 247 с.
9. Перебийніс В.І., Муравицька М.П., Дарчук Н.П. Частотні словники та їх використання. К.: Наукова думка, 1983.

*Рецензент: д.т.н., проф. Давлет'янс О.І.
Надійшла 02.06.2010 р.*