ISSN ***2413-5593 (Print); ISSN 2663-6530 (On-line)***
2025. № 48 : 142-150
Social Sciences. *Arts and Humanities*
PROBLEMS OF SEMANTICS, PRAGMATICS AND COGNITIVE LINGUISTICS
Taras Shevchenko National University of Kyiv, Ukraine
http://semantics.knu.ua/index.php/prblmsemantics

УДК 811.134.2’271’276:[004.934+003.035]

Стаття надійшла до редакції [Article received] – 01.08.2025 р.
Фінансування [Financing] – самофінансування [self-financing]
Перевірено на плагіат [Checked for plagiarism] – 03.08.2025 р.
Оригінальність тексту [The originality of the text] – 97 %
http://doi.org//10.17721/2663-6530.2025.48.11

АВТОМАТИЧНЕ РОЗПІЗНАВАННЯ ІСПАНСЬКОЇ МОВИ: СОЦІОЛІНГВІСТИЧНІ ЧИННИКИ ТОЧНОСТІ ТА ТИПОЛОГІЯ ПОМИЛОК

Юлія Анатоліївна Тарасенко *(м. Київ, Україна)*
y.trsnk@gmail.com

магістрантка кафедри романської філології
Київський національний університет імені Тараса Шевченка
(Міністерство освіти і науки України)
01601, м. Київ, бульвар Тараса Шевченка, 14

У статті досліджено ефективність автоматичного розпізнавання іспанської мови (ASR) на матеріалі корпусу із 304 аудіозаписів мовців різного віку, статі та з різними акцентами. Метою дослідження є оцінка точності системи Google Speech-to-Text, виявлення типових помилок та визначення впливу соціолінгвістичних чинників на якість транскрипції. Для аналізу використано метрики WER та CER, а також кількість замін, видалень і вставок. Результати показали середню точність 94,7 %, при цьому основним типом помилок стали заміни лексем. Найточніше система розпізнавала мовлення носіїв північнопіренейського акценту, а найнижчий рівень коректності спостерігався у підлітків та носіїв аргентинського варіанту іспанської. Практичне значення дослідження полягає у можливості вдосконалення ASR-моделей з урахуванням діалектних та соціальних характеристик мовців.

Ключові слова: *автоматичне розпізнавання мовлення, іспанська мова, ASR, WER, CER, акцент, соціолінгвістика.*

(Current issues in cognitive linguistics [Aktual'ni pytannja kognityvnoi' lingvistyky])
Automatic Speech Recognition of Spanish: Sociolinguistic Factors of Accuracy and Error Typology (in Ukrainian)
[Avtomatychne rozpiznavannia ispanskoi movy: sotsiolingvistychni chynnyky tochnosti ta typolohiia pomylok]

ISSN *2413-5593 (Print); ISSN 2663-6530 (On-line)*
2025. № 48 (УДК 81) : 142-150
ПРОБЛЕМИ СЕМАНТИКИ, ПРАГМАТИКИ ТА КОГНІТИВНОЇ ЛІНГВІСТИКИ
Київський національний університет імені Тараса Шевченка, Україна
http://semantics.knu.ua/index.php/prblmsemantics

AUTOMATIC SPEECH RECOGNITION OF SPANISH: SOCIOLINGUISTIC FACTORS OF ACCURACY AND ERROR TYPOLOGY

Yuliia Anatoliivna Tarasenko *(Kyiv, Ukraine)*
y.trsnk@gmail.com

Master Student at Department of Romanic Philology
Taras Shevchenko National University of Kyiv
(Ministry of Education and Science of Ukraine)
14 Taras Shevchenko Blvd., Kyiv, Ukraine, 01601

The article investigates the effectiveness of Automatic Speech Recognition (ASR) for Spanish based on a corpus of 304 audio recordings of speakers of different ages, genders, and accents. The aim of the study is to evaluate the accuracy of Google Speech-to-Text, identify common errors, and determine the impact of sociolinguistic factors on transcription quality. The analysis employs WER and CER metrics, as well as the number of substitutions, deletions, and insertions. The results revealed an average accuracy of 94.7 %, with substitutions being the predominant error type. The highest accuracy was achieved for speakers with a northern peninsular accent, while the lowest was observed in teenagers and speakers of the Argentinian variety of Spanish. The practical value of this study lies in the possibility of improving ASR models by taking into account dialectal and social characteristics of speakers.

Keywords: *automatic speech recognition, Spanish, ASR, WER, CER, accent, sociolinguistics.*

Вступ. Автоматичне розпізнавання мовлення (Automatic Speech Recognition, ASR) упродовж останніх десятиліть стало важливим інструментом для розвитку технологій обробки природної мови, освітніх застосунків, медичних сервісів і цифрових комунікацій [7, с. 11]. Хоча сучасні системи досягають високих показників точності для англійської мови, дослідження щодо функціонування ASR для іспанської мови залишаються обмеженими, особливо в аспекті впливу соціолінгвістичних чинників на якість розпізнавання [3, с. 12]. Це визначає наукову й практичну **актуальність дослідження**.

(Актуальні питання когнітивної лінгвістики [Aktual'ni pytannja kognityvnoi' lingvistyky])
Автоматичне розпізнавання іспанської мови: соціолінгвістичні чинники точності та типологія помилок (Українською) **[Avtomatychne rozpiznavannia ispanskoi movy: sotsiolingvistychni chynnyky tochnosti ta typolohiia pomylok]**

Розвиток ASR має тривалу історію: від експериментів А. Бела (1950-ті) до впровадження марковських моделей (1970–1980-ті) та сучасних нейромережевих підходів [9, с. 154]. Значний внесок у становлення технології зробили Б. Вінцюк, Д. Редді, Л. Рабінер. У сучасний період активно розробляються відкриті корпуси, серед яких важливе місце посідає проєкт *Common Voice* [4, с. 11]. Для англійської мови проведено численні порівняльні дослідження точності ASR, однак для іспанської така робота є менш систематичною, що створює дослідницький інтерес[4, с. 11].

Метою є оцінка точності системи Google Speech-to-Text під час автоматичного розпізнавання іспанської мови та визначення впливу соціолінгвістичних чинників (таких як вік, гендер, акцент) на результати транскрипції.

Завданнями є:

1. сформувати корпус іспанських аудіозаписів з різними соціолінгвістичними характеристиками;
2. здійснити автоматичну транскрипцію за допомогою Google Speech-to-Text;
3. проаналізувати показники точності (WER, CER) та типи помилок;
4. виявити кореляцію між якістю розпізнавання і соціолінгвістичними параметрами мовців.

Об’єктом дослідження є процес автоматичного розпізнавання іспанського мовлення.

Предметом – точність розпізнавання та типи помилок у транскрипції залежно від соціолінгвістичних характеристик мовців.

Методи дослідження. Під час дослідження були залучені аналітичний метод шляхом попереднього опрацювання уже навявних джерел, описовий метод, а також метод автоматизованої обробки мовлення, кількісний аналіз (WER, CER), якісний аналіз помилок (класифікація на заміни, пропуски, вставки) та соціолінгвістичний аналіз (оцінка впливу віку, гендеру та акценту на точність розпізнавання).

Наукова новизна дослідження полягає у помилок ASR для іспанської мови з урахуванням віку, гендеру та акценту мовців на основі відкритого корпусу *Common Voice*. Встановлено закономірності, що пояснюють відмінності у точності розпізнавання між різними соціолінгвістичними групами.

© ***Tarasenko Yu. A. [Tarasenko Yu. A.],*** *y.trsnk@gmail.com*

Основний зміст. Емпіричною базою дослідження став корпус *Common Voice*, розроблений Mozilla, що є відкритим ресурсом для збирання багатомовних зразків мовлення[5, с. 11]. Із корпусу було відібрано 304 аудіозаписи іспанською мовою загальним обсягом понад 3 години звучання [2, с. 219]. Записи репрезентують мовців різного віку (підлітки, молодь, середній та старший вік), статі (чоловіки й жінки) та з різними акцентами (північнопіренейський, андалузький, канарський, аргентинський). Такий відбір забезпечив різноманітність соціолінгвістичних характеристик та дозволив виявити їхній вплив на якість автоматичного розпізнавання.

Для забезпечення порівнюваності результатів усі аудіозаписи було приведено до однакових технічних параметрів (моно, 16 кГц, формат WAV). Тексти-оригінали (референс-транскрипції) було очищено від пунктуаційних та графічних варіантів, які не впливають на зміст. Особливу увагу приділено нормалізації: усі числівники подано літерами, скорочення розгорнуто, абревіатури уніфіковано. Це дозволило уникнути викривлень при розрахунку метрик.

Для транскрипції використано сервіс *Google Speech-to-Text*, який базується на глибинних нейронних мережах і є одним із найточніших комерційних рішень [1, с. 13]. Розпізнавання здійснювалося у режимі *default*, без додаткових мовних моделей. Отримані транскрипти автоматично зіставлялися з референсними текстами для подальшого аналізу [8, с. 4].

Якість автоматичного розпізнавання визначалася за допомогою двох основних метрик:

- $Word\ Error\ Rate = (Insertions + Substitutions + Deletions) \div Total\ Words\ in\ Correct\ Transcript \times 100\%$

де:

$Insertions$ – кількість доданих слів,

$Substitutions$ – кількість замінених слів,

$Deletions$ – кількість видалених слів,

$Total\ Words\ in\ Correct\ Transcript$ – кількість слів у цільовій транскрипції [6, с. 169].

- Формула для Character Error Rate (CER) дуже схожа на WER, але обчислюється на рівні символів, а не слів.

ISSN ***2413-5593 (Print); ISSN 2663-6530 (On-line)***
2025. № 48 : 142-150
Social Sciences. *Arts and Humanities*
PROBLEMS OF SEMANTICS, PRAGMATICS AND COGNITIVE LINGUISTICS
Taras Shevchenko National University of Kyiv, Ukraine
http://semantics.knu.ua/index.php/prblmsemantics

CER = $(Insertions + Substitutions + Deletions) \div$ Total Characters in Correct Transcript ×100 %

де:

Insertions– кількість доданих символів,

Substitutions – кількість замінених символів,

Deletions – кількість видалених символів,

Total Characters in Correct Transcript – загальна кількість символів у еталонній транскрипції [3, с. 24].

У середньому система продемонструвала досить високий рівень коректності. Середній показник Word Error Rate (WER) становив 5.28 %, що відповідає приблизно 94.7 % правильно розпізнаних слів. Це означає, що з кожних ста слів лише близько п’яти були відтворені неправильно.

Середній показник Character Error Rate (CER) склав 2.56 %, що демонструє низький рівень відхилень на рівні символів. Таким чином, навіть у випадках, коли окремі слова були розпізнані з помилкою, написання залишалося близьким до правильного.

Аналіз показав, що найбільшу частку становлять заміщення (Substitutions). Наприклад, для мовців 30-х років заміщення становили в середньому 12.4 %, тоді як видалення – лише 0.9 %, а вставки взагалі не фіксувалися. Подібна тенденція простежувалася і в інших вікових групах: у 20-річних заміщення становили 25.8 %, а видалення та вставки не перевищували 0 %. У підлітків (teens) спостерігався найбільш критичний показник: заміщення сягали 60 %, тоді як видалень і вставок не було. У мовців 50-х років заміщення становили 50 %, без жодного видалення чи вставки.

Подібна закономірність проявлялася і при аналізі за іншими змінними. За гендером у чоловічих голосах заміщення склали в середньому 19 %, тоді як видалення – лише 0.5 %, а вставок не було зовсім. У категорії “other” спостерігалася аналогічна картина: усі помилки були пов’язані з заміщенням (60 %).

За акцентом заміщення також переважали: у носіїв північнопіренейського акценту вони становили 12.4 %, видалення – лише 0.9 %; у носіїв центросурпіренейського – 27 % заміщень при відсутності інших типів помилок;

© ***Tarasenko Yu. A. [Tarasenko Yu. A.],*** *y.trsnk@gmail.com*

у ріоплатенському варіанті іспанської – 30 % заміщень, також без видалень і вставок.

Таким чином, можна зробити висновок, що Google Speech-to-Text прагне «заповнити» мовний потік будь-якими словами, навіть якщо вони не відповідають еталону. Це призводить до того, що система рідко пропускає чи додає слова, натомість основним джерелом помилок стають неправильні заміни лексем.

Приклади типових розбіжностей

1. Еталон: *“por su puesto”*
   Розпізнано: *“por supuesto”*
   Спостерігаємо заміщення, адже система обрала більш поширений варіант написання, що фонетично збігається. WER для цього прикладу становив 66.6%, тоді як CER був лише 7.7%, оскільки на рівні символів різниця мінімальна.
2. Еталон: *“una figura emergente nos eclipsará enseguida”*
   Розпізнано: *“una figura emergente no se eclipsa enseguida”*
   Тут маємо одразу кілька розходжень: заміщення (*“nos” → “no”*), а також граматичне спрощення (*“eclipsará” → “se eclipsa”*). Попри це, CER залишився невисоким (9%), адже система правильно зберегла більшість символів, але інтерпретувала дієслівну форму по-іншому.
3. Еталон: *“después suspiró”*
   Розпізнано: *“después suspiro”*
   Система замінила дієслівну форму на іншу, знявши наголос на останньому складі. Це приклад морфологічного заміщення, що не змінює кардинально сенсу, але є суттєвим для лінгвістичного аналізу.
4. Еталон: *“a caballo corredor y hombre reñidor poco les dura el honor”*
   Розпізнано: *“a caballo corredor y hombre reñido el honor”*
   Тут спостерігаємо скорочення фрази: система “втратила” частину (*“poco les dura”*), але зберегла основну структуру. Це приклад рідкісного випадку видалення. CER тут становив 25.8 %, що вказує на більшу кількість відхилень, але водночас розпізнаний текст залишається граматично зв’язним.

   Результати за соціальними групами

© Тарасенко Ю. А. [Tarasenko Yu. A.], *y.trsnk@gmail.com*

Поділ результатів за віком, статтю та акцентом показав цікаві закономірності.

- За віком: найнижчий показник WER спостерігався у мовців 30-х років (13.2 %), тоді як найвищий – у підлітків (60 %). Це пояснюється більшою чіткістю дикції у дорослих мовців та можливими особливостями молодіжної вимови.
- За статтю. Чоловічі голоси розпізнавалися значно точніше (WER = 19.5 %), ніж голоси з категорії *other* (WER = 60%). Така різниця може бути зумовлена як якістю записів, так і меншою кількістю представників у другій групі, що вплинуло на стабільність результатів.
- За акцентом: найточніше система працювала з носіями північнопіренейського акценту (WER = 13.3 %), а найгірше – з носіями агрентинського варіанту іспанської мови (WER = 30 %). Це узгоджується з відомими труднощами обробки агрентинського варіанту іспанської через особливості фонетики (наприклад, *yeísmo*).

Висновки. Послідження показало, що сучасні системи автоматичного розпізнавання мовлення (ASR), зокрема Google Speech-to-Text, демонструють доволі високий рівень точності для іспанської мови. Середній показник WER становив 5.28 % (приблизно 94.7 % правильно розпізнаних слів), а CER – 2.56 %, що свідчить про незначні відхилення на рівні символів. Основним джерелом помилок були заміщення (Substitutions), тоді як видалення та вставки зустрічалися рідко. Соціолінгвістичний аналіз показав, що точність розпізнавання залежить від віку, гендеру та акценту мовця: найнижчий WER спостерігався у дорослих носіїв північнопіренейського акценту, а найвищий – у підлітків та носіїв аргентинського варіанту іспанської.

Практичне значення дослідження полягає у тому, що результати можна застосовувати для підвищення ефективності мовних технологій у сфері освіти, автоматичного створення субтитрів, перекладу та аналізу мовних даних. Отримані дані дозволяють розробникам ASR адаптувати моделі під різні вікові групи та регіональні варіанти мови, а також враховувати специфіку вимови для покращення точності транскрипцій.

© ***Tarasenko Yu. A. [Tarasenko Yu. A.],*** *y.trsnk@gmail.com*

Разом з тим, дослідження має певні обмеження: обмежений розмір корпусу (304 аудіозаписи), використання лише одного сервісу ASR та можливий вплив якості запису і шумового середовища на точність.

Перспективи подальших досліджень включають порівняння декількох систем ASR для іспанської мови, розширення корпусу з урахуванням більшої кількості акцентів та вікових груп, а також дослідження методів автоматичного навчання моделей із врахуванням типових помилок, що дозволить підвищити точність розпізнавання для різних категорій мовців.

Література:

1. Дудченко, І. В. (2020). *Голосове управління комп'ютером на основі глосарію за допомогою алгоритмів розпізнавання мови* [Дипломний проєкт, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»]. https://ela.kpi.ua/server/api/core/bitstreams/781f9949-a7c4-4033-bcd6-403b6449e866/content
2. Наход, О. (2025). *Автоматичне розпізнавання українського мовлення на основі глибокого навчання.* https://doi.org/10.36074/logos-24.01.2025.043
3. Самвеляян, А. Р. (2021). *Розробка системи автоматичного розпізнавання українського мовлення* [Дипломна робота, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»]. https://ela.kpi.ua/server/api/core/bitstreams/af954b61-6f2e-47b7-963b-aac8648b500f/content
4. Вінцюк, Т. К., Сажок, М. М., Селюх, Р. А., Федорин, Д. Я., Юхименко, О. А., & Робейко, В. В. (2018). *Автоматичне розпізнавання, розуміння та синтез мовленнєвих сигналів в Україні,* (6), 7–24. https://nasplib.isofts.kiev.ua/handle/123456789/161562
5. Ardila, R., Branson, M., Davis, K., Kohler, M., Meyer, J., Henretty, M., Morais, R., Saunders, L., Tyers, F., & Weber, G. (2019). *Common Voice: A massively-multilingual speech corpus.* https://arxiv.org/abs/1912.06670
6. Gómez Seibane, S., San Martín, M., Herras, J., & Mata, G. (2024). *Is ASR a suitable tool for creating spoken linguistic corpora in European Spanish? Procesamiento del Lenguaje Natural, 73,* 165–176. https://corpusrural.es/publicaciones/2024/GomezSeibane-et-AL-SEPLN-2024.pdf
7. Jurafsky, D., & Martin, J. H. (2018). *Speech and language processing.* Stanford University. https://web.stanford.edu/~jurafsky/slp3/
8. Maison, L., & Estève, Y. (2023, August). *Some voices are too common: Building fair speech recognition systems using the Common Voice dataset.* In *Interspeech 2023* (ISCA). Dublin, Ireland. https://hal.archives-ouvertes.fr/hal-04163615

© Тарасенко Ю. А. [Tarasenko Yu. A.], *y.trsnk@gmail.com*

9. Rufiner, H. L., & Milone, D. H. (2004). *Sistema de reconocimiento automático del habla. Ciencia, Docencia y Tecnología, XV*(28), 151–177. https://www.redalyc.org/articulo.oa?id=14502806

References:

1. Dudchenko, I. V. (2020). *Holosove upravlinnia komputerom na osnovi hlosariiu za dopomohoiu alhorytmiv rozpiznavannia movy* [Diploma project, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"]. https://ela.kpi.ua/server/api/core/bitstreams/781f9949-a7c4-4033-bcd6-403b6449e866/content
2. Nakhood, O. (2025). *Avtomatychne rozpiznavannia ukrains'koho movlennia na osnovi hlybokoho navchannia.* https://doi.org/10.36074/logos-24.01.2025.043
3. Samvelian, A. R. (2021). *Rozrobka systemy avtomatychnoho rozpiznavannia ukrains'koho movlennia* [Diploma thesis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"]. https://ela.kpi.ua/server/api/core/bitstreams/af954b61-6f2e-47b7-963b-aac8648b500f/content
4. Vintsiuk, T. K., Sazhok, M. M., Seliukh, R. A., Fedorin, D. Ya., Iukhymenko, O. A., & Robeiko, V. V. (2018). *Avtomatychne rozpiznavannia, rozuminnia ta syntez movlennievykh syhnaliv v Ukraini. Upravliuiuchi systemy i mashyny, (6),* 7–24. https://nasplib.isofts.kiev.ua/handle/123456789/161562
5. Ardila, R., Branson, M., Davis, K., Kohler, M., Meyer, J., Henretty, M., Morais, R., Saunders, L., Tyers, F., & Weber, G. (2019). *Common Voice: A massively-multilingual speech corpus.* https://arxiv.org/abs/1912.06670
6. Gómez Seibane, S., San Martín, M., Herras, J., & Mata, G. (2024). *Is ASR a suitable tool for creating spoken linguistic corpora in European Spanish? Procesamiento del Lenguaje Natural, 73,* 165–176. https://corpusrural.es/publicaciones/2024/GomezSeibane-et-AL-SEPLN-2024.pdf
7. Jurafsky, D., & Martin, J. H. (2018). *Speech and language processing.* Stanford University. https://web.stanford.edu/~jurafsky/slp3/
8. Maison, L., & Estève, Y. (2023, August). *Some voices are too common: Building fair speech recognition systems using the Common Voice dataset.* In *Interspeech 2023* (ISCA). Dublin, Ireland. https://hal.archives-ouvertes.fr/hal-04163615
9. Rufiner, H. L., & Milone, D. H. (2004). *Sistema de reconocimiento automático del habla. Ciencia, Docencia y Tecnología, XV*(28), 151–177. https://www.redalyc.org/articulo.oa?id=14502806

© ***Tarasenko Yu. A. [Tarasenko Yu. A.],*** *y.trsnk@gmail.com*