

# ГЕНЕРУВАННЯ ПРАВИЛ РОЗСТАВЛЯННЯ НАГОЛОСІВ У БАГАТОМОВНОМУ АСПЕКТІ

Микола Сажок

Міжнародний науково-навчальний центр інформаційних технологій та систем

40 просп. Академіка Глушкова, Київ 252022

Електронна пошта: mykola@uasoiro.freenet.kiev.ua

*Mykola Sazhok. Generating Stress Rules in Multilingual Aspect.* The risen problem of stresses, particularly in Ukrainian, often confuses text-to-speech and recognition engines. Given analysis of stress in Ukrainian generalizes knowledge about this important aspect. Cases when stress moving in a word changes its meaning and/or its role in sentence are considered. Revealed properties allow for building heuristics and statistics based rules predicting a stress position in most cases and resolving several types of ambiguous stresses in words. Within entire paper cross-lingual stress aspects are discussed on example of Ukrainian and English. It is shown how parts of word influent on stress position and how narrow context helps to resolve stress ambiguities.

## 1 Вступ

Роль наголосів в усній мові важко переоцінити. Властивість наголосу коригувати та й зовсім змінювати зміст промовленого широко використовується у спілкуванні людей.

Хибний наголос різке слух і часто призводить до непорозуміння при сприйнятті усної мови людиною. Тому в усномовних системах для практичного використання при автоматичному синтезі правильний наголос значно поліпшує розбірливість синтезованого усномовного сигналу, а у випадку розпізнавання та смислової інтерпретації усномовного сигналу нехтування наголосами взагалі неприпустиме.

Розмаїтість типів наголосів значно ускладнює аналіз наголосу в цілому, тому вирішено зосередитися на наголосі саме у словах. Попри особливості наголосу в різних мовах, можна виокремити чимало спільних рис, притаманних розставленню наголосів як у споріднених, так і у віддалених мовах.

Використання знань про наголос в системах автоматичного синтезу, розпізнавання та інтерпретації усномовного сигналу завжди було на часі. Проте, для мов, де наголос є надто нерегулярним і не позначається на письмі, зокрема для української та англійської, дослідження з цього питання потребують більше передумов та ретельності. Особливо це стосується для мов, що допускають велику кількість словоформ, до яких відноситься українська.

## 2 Загальний аналіз природи наголосу

Що ж таке по суті є наголос? Чіткого визначення цієї просодичної характеристики напевно не існує. Очевидно, що це явище є чимось принципово

важливим в усній мові. Також помічено, що під час наголошення вимовляння найбільш відповідає фонетичним нормам та є найближчим до написання. Тобто в наголошеній позиції фонема-алофон звучить (реалізується) найбільш чітко, найменшою мірою піддаючись спотворенням та редукції.

В загальному випадку, наголос може бути сильнішим та слабшим, виокремлюючись на тлі як цілої фрази, речення, так і лише окремого слова або сусідніх складів. Отже, наголоси формуються на трьох рівнях: фрази, слова або синтагми та на субсловарному мікрорівні або на рівні складів.

Зосереджуючись на рівні не вище слова, зазначимо, що тут наголоси бувають основні та другорядні, які баланують на межі мікрорівня.

Позиція основного наголосу у слові рахується мовознавцями в залежності від мови частіше з кінця слова, але буває і з початку. Тут у деяких мов спостерігається регулярність. Хоча при вимовлянні фрази в цілому слово може втратити наголос, набуваючи другорядного наголосу зовсім не там, де розташовується основний. Типовою ілюстрацією такої властивості є приклад з французької "moulen rouge", де жирні літери позначають основний наголос, а похилі літери – другорядний.

Простежується також залежність наголосу від типу мовлення. В першу чергу ритм, потім темп, розмовність впливають на формування наголосів, їх плавання, перехід у слабкіший стан. Вплив ритму на наголоси чудово простежується на зразках поезії, пісенної творчості, коли диктат ритму інколи зовсім не зважає на граматично правильний наголос.

Прикладом цього є уривок з української народної пісні: "Ой у лісі калина, калина, калина", у той час, як словник подає "калина"

В українській мові основний наголос може перебувати на останньому або другому з кінця складі. Досить поширений також дактилічний наголос, тобто третій з кінця. Але цим не обмежується і в деяких словоформах позиція наголосу може перебувати аж на сьомому з кінця складі як-от у слові "спізнюватимемося".

Англійська мова також має нерегулярний наголос. Відлік позиції наголосу ведеться з початку слів, як для більшості германських мов, і припадає основний наголос переважно на перший склад. Помічено також тенденцію англійською мови до перероблення наголосів запозичених слів, наприклад французьких, на свій лад.

Наголос тісно пов'язаний з іншими просодичними характеристиками, особливо інтона-

ційними. Простежується чітка залежність різких змін інтонаційного контуру саме одночасно з початком наголошеного складу. Безумовно, кожна усна мова своєрідно реалізує наголос, проте крос-лінгвістичні порівняння в багатьох випадках дозволяють адаптувати знання про наголос з однієї мови на іншу.

### 3 Автоматичне розставлення наголосів

Побудувати алгоритм розставлення наголосів для мов з нерегулярним наголосом у словах і просто і складно. Здавалося б, що може бути простішого, ніж знайти слово з тексту в словнику, де завчасно проставлені наголоси та поставити наголос у шуканому слові. Проте такий шлях вирішення не може задовольнити з багатьох причин.

Для мов з розмаїтістю словоформ, якою є українська, де наголоси від форми до форми часто змінюються, саме створення словників, які враховують позицію наголосу є проблематичним, не кажучи вже про обсяги пам'яті, потрібні для такого словника, та час доступу до певного слова.

Потреба у використанні величезних обсягів пам'яті обмежує використання усномовних технологій як на сучасних комп'ютерах, так і в портативних виробках особливо, коли обмеження на обсяги пам'яті та потужності процесора суттєві. Вихід з цього закуту вбачається у винайденні правил, які б дозволяли розставляти наголоси в довільному слові для певної мови.

Спроби формування таких правил вперше були зроблені для англійської мови. Як було зазначено, велика кількість англійських слів містить основний наголос на першому складі. Водночас були виявлені суфікси, що "перетягають" наголос. Це "-ion", "-ize", "-ate" та деякі інші. Послабленими в багатьох випадках залишаються також префікси. Так на прикладі слів "expect" і "expectation" видно, що префікс "-ex" залишається ненаголошеним, а додавання суфіксів "-ate", "-ion" веде до зсуву наголосу в напрямку до суфікса.

В українській мові на роль найбільш поширеного наголосу претендують як перший, так і другий і третій від останнього. Остаточо визначити це можливо лише проаналізувавши ужиток слів, керуючись текстовим корпусом, а також з міркувань більш економного задання правил. Також виокремлюються нечисленні суфікси, за якими можна визначити наголос ("-еньк", "-есеньк"). Крім того, спостерігається чітка тенденція до послаблення закінчень, за винятком прикметників.

Складені та довгі слова видаються найбільш проблематичними для розставлення наголосу, особливо в англійській мові.

### 4 Зсув наголосу при творенні слів і форм

Зміна наголосу широко використовується при словотворенні. Так англійські "record" і "record" ілюструють перехід іменника у дієслово, під час якого відбувається зсув наголосу.

Українська мова більш широко використовує зміну позиції наголосу і не лише при творенні нових слів, але і при творенні форм слова. Так в іменниках наголос може зсунутись при переході у множину (рука-руки, поле-поля), рідше при звичайному відмінюванні під впливом певних суфіксів.

Дієслова досить вільно поводяться з наголосами, хоча простежуються деякі закономірності з урахуванням сукупного впливу на наголос усіх частин слова та кількості складів у слові.

Прикметники найбільш консервативно зберігають наголошену голосну, при тому, що утворені від них прислівники уникають наголосів на останньому складі (значний-значно), інколи – на передостанньому (веселий-весело).

### 5 Проблема неоднозначності наголосів

Зміна наголосу в слові призводить, як було сказано, найчастіше до помилки у вимові. Випадки, коли зміна наголосу веде до переведення слова в іншу словоформу, або змінює його значення на інше, часом безпосередньо не пов'язане з початковим семантично, призводять до помилки у смисловій інтерпретації як окремого слова, так і цілого вислову.

У випадках, коли наголос змінює форму слова слід витримати непорушність узгодження слів у реченні. Для цього достатньо визначити керування в якому перебуває це слово за допомогою слів-орієнтирів з найближчого контексту. Такими орієнтирами є прийменники або їх відсутність, прикметники, як найбільш консервативні у наголошенні, та слова з інших частин мови, які вимагають певного узгодження, і чие значення не викликає сумніву.

Так, у словосполученні "знімати з руки" прийменник "з" вимагає після себе родовий або твірний відмінок, тому й наголос має відповідати одному з цих відмінків. У вислові "потисли один одному руки" слово "руки" керується дієсловом "потисли", яке вимагає знахідного відмінку.

У випадках, коли можливі варіанти наголосів, що відповідають різним частинам мови, постає проблема визначення частини мови цього слова в контексті. Варіанти наголосів, які не змінюють частину мови, змінюючи при цьому значення слова (плачу, замок), слід розглядати окремо.

### 6 Підсумок

Проведений аналіз дозволяє виробити евристичні та статистичні підходи до формування правил розставлення наголосів не лише в контексті слова, а і словосполучення та цілої фрази, розв'язуючи у багатьох випадках проблему неоднозначності інтерпретації слів при усномовному синтезі та розпізнаванні. Для багатомовних систем важливим напрямком подальших досліджень вбачається використання знань про наголоси зі споріднених мов.