Інтерпретація злитого українського мовлення для усного словника-перекладача

Тарас Вінцюк, Микола Сажок, Валентина Яценко,
Міжнародний науково-навчальний центр інформаційних технологій та систем
40 просп. Академіка Глушкова, Київ 03680

Abstract

A problem of continuous speech interpretation within a subject domain is considered. A way to specify allowable sequences of words in phrases by means of LISP-structures is considered in frames of the generative model for speech understanding for highly inflective languages with relatively free word order. Experimental research presents promising results, which allowed for creating a spoken vocabulary-interpreter prototype.

1. Вступ

Актуальною є проблема підвищення надійності розпізнавання та розуміння мовленнєвих сигналів. Один із можливих шляхів досягнення цієї мети полягає в економному заданні різноманітних обмежень, зокрема синтаксичних та семантичних, на допустимі послідовності слів у фразах та їх врахуванні при автоматичному розумінні та розпізнаванні мовленнєвих сигналів.

Для чого це потрібно і де використовується? Для різних систем усного діалогу: фразника-перекладача, довідкових систем тощо.

Для слов'янських мов розроблені автоматичні модулі, які забезпечують досить високу точність розпізнавання та аналізу речень, але при цьому вимагають значних обчислювальних ресурсів, не працюють з багатоваріантними випадками. Це пов'язано з тим, що слов'янські мови мають певні відмінності від, наприклад, англійської мови. Основними відмінностями є істотно більша кількість словоформ для кожного слова та вільний порядок слідування слів у фразах. Врахування цих особливостей для слов'янських мов значною мірою ускладнює розв'язування задач розпізнавання та змістовної інтерпретації мовленнєвих сигналів.

Окремою є проблема задання всіх можливих речень мови діалогу, що виражають один і той самий зміст, генерації та пошуку найбільш правдоподібних сигналів та розроблення обмежень на допустимі послідовності слів згідно структур, якими можна представити речення.

Для дослідження та розроблення обмежень на допустимі послідовності слів у фразах було запропоновано розглянути LISP-структури [1, 2]. На основі цих структур генерується величезна кількість речень, що мають один і той самий зміст. Для підвищення ефективності алгоритмів розпізнавання було запропоновано автоматизувати побудову LISP-структур.

Далі, у 1 розділі ми дамо загальну характеристику задач розпізнавання та змістовної інтерпретації злитого мовлення, у 2 розділі – постановку задачі, що стосується задання (специфікації) речень, з врахуванням обмежень на допустимі послідовності слів, 3 – експериментальні результати.

Для вирішення поставленої задачі було запропоновано розглянути алгоритм розпізнавання мовленнєвих сигналів, що використовує речення, наперед задані за допомогою LISP-структур, ґрунтується на словнику, утвореному на основі цих речень, та враховує різні граматики, які описують порядок слідування слів у фразах.

2. Загальна характеристика задач розпізнавання та інтерпретації злитого мовлення

Розглянемо в чому полягають і як взаємозв'язані задачі розпізнавання та інтерпретації злитого мовлення [1, 2]. Розпізнавання мови – це процес автоматичної обробки сигналу з метою визначення послідовності слів, які передаються цим сигналом.

Змістовна інтерпретація мови – це процес автоматичної обробки мовленнєвого сигналу з метою виявлення змісту, що передається сигналом, та представлення цього змісту в певній канонічній формі, зручній для подальшого використання. Очевидно, що змістовна інтерпретація мови є більш високим ступенем узагальнення інформації, ніж розпізнавання, оскільки одну і ту саму думку можна виразити різними послідовностями слів. Для отримання кращих результатів розпізнавання та змістовної інтерпретації злитого мовлення ці задачі повинні виконуватися в єдиному взаємопов'язаному процесі. Кінцевою метою цього процесу є зміст повідомлення, який передається послідовністю слів.

Оскільки кожну думку можна висловити різними реченнями в мові діалогу, але при цьому зміст не зміниться, то слід визначити певні обмеження на допустимі послідовності слідування слів у реченнях. Тому, при інтерпретації змісту мови різні речення, що передають одну і ту саму думку, повинні відображатися в один і той же результат, тобто відповідь розпізнавання (послідовність слів) не повинна суперечити синтаксису, семантиці та прагматиці предметної області. Зважаючи на це пропонується розглянути моделі розпізнавання мовленнєвих сигналів, які враховують синтаксис та семантику мови [1, 2].

Задача змістовної інтерпретації злитого мовлення значно складніша за задачу розпізнавання, оскільки для її розв'язання необхідно додатково враховувати апріорну інформацію. Тому, перш за все слід навчитися економно задавати всі можливі допустимі речення в мові діалогу. Для вирішення цього питання є декілька шляхів. Один з них – побудова LISP-подібних структур та визначення за їх допомогою обмежень на допустимі послідовності слів.

Цей підхід до розпізнавання та змістовної інтерпретації злитого мовлення може бути реалізовано у вигляді генеративної моделі розуміння (змістовної інтерпретації) злитого мовлення [1].

Згідно цієї моделі при змістовній інтерпретації мови основною є задача розпізнавання змістовного висловлювання із заданої множини змістовних висловлювань. Необхідно вказати, яке змістовне висловлювання із заданої скінченої множини змістовних висловлювань міститься в пред'явленому мовленнєвому сигналі.

Кожне змістовне висловлювання задамо в канонічній формі, що записана на певній семантичній мові (формальній математичній мові, що виражає поняття та відношення між ними). Далі за допомогою генератора семантично еквівалентних речень (ГСЕР) задамо перетворення канонічної форми, що не порушує зміст висловлювання. Таким чином ГСЕР породжує всі можливі речення з однаковим змістом, що визначаються канонічною формою. Далі введемо перетворення, що породжують всі можливі еталонні сигнали злитого мовлення для кожного речення, згенерованого ГСЕР. Ці еталонні сигнали відрізняються один від одного темпом, що нелінійно змінюється, та інтенсивністю вимовляння.

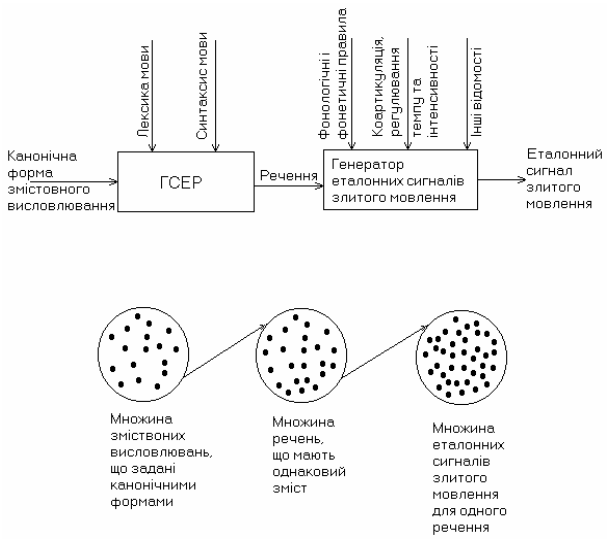

Рис.1. Модель синтезу еталонних сигналів злитого мовлення для змістовної інтерпретації.

Автоматичне розуміння (змістовна інтерпретація) пред'явленого мовленнєвого сигналу за допомогою пропонованої генеративної моделі буде полягати в тому, щоб спочатку для сигналу, що аналізується, знайти найбільш правдоподібний еталонний сигнал мови серед всіх сигналів, що породжені генеративною моделлю, а потім визначити канонічну форму того змістовного висловлювання, речення якого відповідає найбільш правдоподібному еталонному сигналу.

Далі буде розглянуто способи задання всіх можливих речень мови діалогу, що виражають один і той самий зміст, генерації та пошуку найбільш правдоподібних сигналів та розроблення обмежень на допустимі послідовності слів згідно структур, якими можна представити речення.

Покладається, що задачі розпізнавання та змістовної інтерпретації злитого мовлення повинні розв'язуватися в єдиному взаємопов'язаному процесі, в якому розпізнавання керується з боку семантико-синтаксичного рівня так, що досягається сама висока надійність як розпізнавання, так і інтерпретації змісту.

3. Врахування обмежень на допустимі послідовності слів

В рамках генеративної моделі для розпізнавання мовленнєвих сигналів запропоновано розглянути певну ієрархію розташування речень. Мається на увазі, що всі мислимі речення мови діалогу розіб'ємо на предметні області (ПО) по типу розмовника для іноземних мов. Кожна предметна область складається із скінченої множини типів змістів (ТЗ). Наприклад, стосовно предметної області щодо відвідання ресторану типи змістів виражаються питаннями про бронювання столику, меню, замовлення і тощо. Кожній предметній області відповідає не так вже і багато типів змістів. В кожен тип змісту входить множина еквівалентно змістовних типів речень (ТР), які описуються LISP-структурами [1]. Тип речення – це конструкція, що економно задає множину речень, отриманих з одного речення незалежними допустимою заміною та допустимою перестановкою слів та словосполучень.

Розглянемо приклад ТР для ПО «Повсякденні фрази», що стосується прохання про допомогу у вирішенні проблеми (ТЗ – прохання про допомогу).

$$\begin{bmatrix} Чи \\ * \end{bmatrix}\left(\left(\begin{bmatrix} не \\ * \end{bmatrix} допоможете\right)\left(\begin{bmatrix} Ви \\ * \end{bmatrix}\begin{bmatrix} мені \\ * \end{bmatrix}\right)\right)$$

$$\left(\begin{pmatrix} вирішити \\ розв`язати \end{pmatrix}(цю\ проблему)\right)$$

В дужках () вказані підсловники, які можна переставляти місцями, а в [] – які не можна переставляти. Переставляти підсловники можна лише всередині старших дужок. Символ * означає порожнє слово.

Неважко переконатися, що наведений тип речення задає 2·4!·2·4·2·1=768 різних речень, допустимих в мові діалогу та таких, що виражають один і той самий зміст прохання про допомогу. Серед цих речень є, наприклад, і такі:

Чи цю проблему не допоможете вирішити Ви мені.
Чи Ви мені цю проблему вирішити не допоможете.

Тобто ми бачимо, що в даний ТР включено багато синтаксично допустимих речень розмовної мови.

Всі речення мови діалогу можна задавати за допомогою ТЗ і відповідних їм ТР, використовуючи структуру, наведену у прикладі. За допомогою LISP-структур генерується величезна кількість речень, що мають один і той самий зміст. Оскільки побудова LISP-структур є досить громіздкою, потребує багато ручної роботи, то було запропоновано автоматизувати цю побудову.

Для побудови всіх можливих речень мови усного діалогу будемо використовувати так звану орієнтовану семантичну мережу (ОСМ) [1]. Оскільки терміни синтаксис та прагматика значно менше впливають на порядок слідування слів у фразах, ніж семантика, то будемо поки що опускати ці терміни у назві мережі.

Для побудови орієнтованої семантичної мережі будемо використовувати згадані раніше поняття ТЗ та ТР. Основним елементом ТР є підсловник. Підсловники іменуються в залежності від їх належності до предметної області.

Орієнтована семантична мережа (ОСМ) має стани, які будемо позначати y : серед них – один початковий $y_{поч}$ та один кінцевий $y_{кінц}$. Стани $y = \mu$ та $y = \nu$ поєднані стрілочками. Кожній стрілочці приписаний

підсловник $z_{\mu\nu}$. Самі підсловники описані (пойменовані) згідно семантики їх предметної області або пронумеровані відповідно до місця цього підсловника в ОСМ. Одне і те саме слово може належати різним підсловникам. Певні підсловники можуть бути виділені як базові. Рухаючись за стрілкою $\mu\nu$, що поєднує стани μ та ν, будемо вибирати лише одне із слів підсловника $k: k \in z_{\mu\nu}$.

Будемо будувати ОСМ таким чином, щоб, при переході з $y_{поч}$ в $y_{кінц}$ утворювалися лише допустимі в мові діалогу речення, тобто такі, що задовольняють синтаксису, семантиці та прагматиці ПО. ОСМ бажано будувати таким чином, щоб в ній було як можна менше станів.

Побудуємо ОСМ для речення, наведеного в прикладі.

Будуючи ОСМ, краще відштовхуватись від ТЗ та ТР. Кожному ТР може відповідати своя ОСМ. Тоді, поєднавши ОСМ типів речень, отримаємо ОСМ для даного типу змісту, а поєднавши мережі відповідних типів змістів, отримаємо ОСМ даної предметної області.

Легко переконатися, що, якщо ТР економно вказує лише спосіб утворення різних речень, що передають один і той самий зміст, то ОСМ для даного типу речення економно задає всі ці речення. ОСМ більш громіздка, але і більш зручна при змістовній інтерпретації злитого мовлення.

4. Експериментальні результати

В якості експериментальних даних було розглянуто англійсько-український розмовник. Розмовник складається з 3800 речень, які назвемо основними. Ці фрази розділені на 15 предметних областей, кожна з яких має свої типи змістів та типи речень. Для прикладу було розглянуто одну з 15 ПО, а саме «Повсякденні фрази». Ця ПО містить 47 ТЗ та 201 основних речень, в середньому по 5 основних речень на тип змісту. Щоб задати всю множину речень, яка породжується основним реченням, кожне основне речення було розмічено у відповідності до описаних LISP-структур. Таким чином, для кожного основного речення було побудовано тип речення у вигляді LISP-структури.

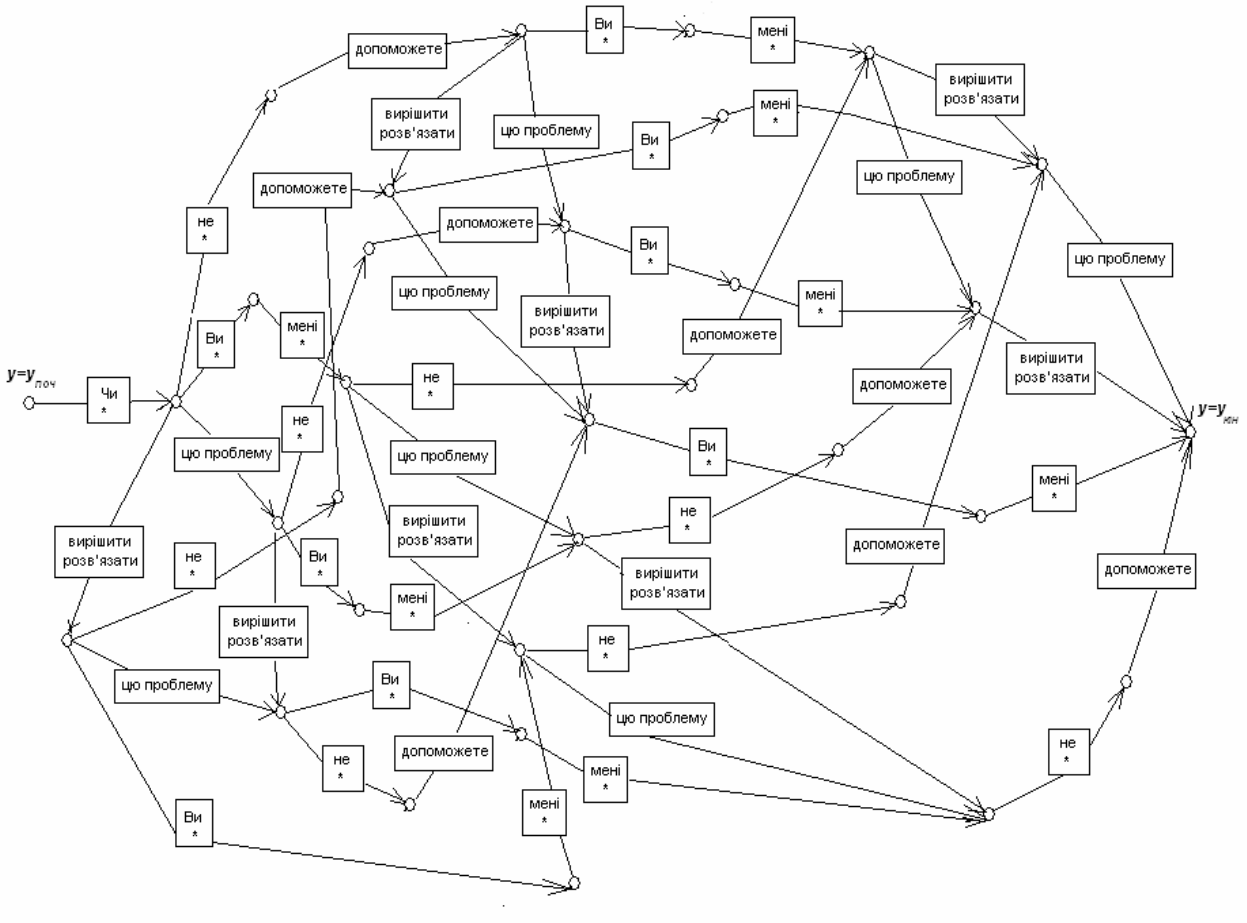

Рис.2. Структура ОСМ для речення, наведеного в прикладі

Так, наприклад, для типу змісту прохання про допомогу, побудовані такі ТР:

((будь ласка | *) (допоможіть) (мені | *) ((вирішити | розв'язати) (цю проблему)))

((будь ласка | *) (допоможіть) (мені | *) (у [цій | *] (справі)))

((не могли б) (Ви мені) (допомогти) ((вирішити | розв'язати) (цю проблему)))

[чи | *] (([не | *] [допоможете]) (Ви мені) (вирішити | розв'язати) (цю проблему))

((мені | *) (потрібна) ([Ваша | *] (допомога)) (у [цій | *] (справі)))

((у цьому питанні) (мені | *) ((буде | *) (потрібна)) ([Ваша | *] (допомога)))

У наведеному прикладі підсловники містять здебільшого по одному слову, але загалом потужність окремо взятого підсловника може бути більшою в залежності від кількості синонімів.

Було розроблено програмне забезпечення, яке з одного заданого таким тином ТР, дає змогу будувати множину всіх речень, шляхом відповідних перестановок чи заміни слів та словосполучень. В результаті застосування цієї програми до згаданих вище 201 ТР, було отримано 1045 фраз, не враховуючи змінні, якщо врахувати змінні, то фраз буде 4337. У словнику нараховується 290 слів.

Для ТР, взятого з наведеного прикладу, – ([чи | *] (([не | *] [допоможете]) (Ви мені) (вирішити | розв'язати) (цю проблему))) було згенеровано 24 фрази, без урахування змінних, серед них:

$p289 = $w11 $w24 dopomozhete tsyu problemu $w19 Vy meni ;

$p295 = $w11 tsyu problemu $w24 dopomozhete $w19 Vy meni ;

$p297 = $w11 tsyu problemu $w19 $w24 dopomozhete Vy meni ;

$p301 = $w11 $w19 $w24 dopomozhete tsyu problemu Vy meni ;

$p304 = $w11 $w19 tsyu problemu Vy meni $w24 dopomozhete ;

$p307 = $w11 Vy meni $w24 dopomozhete tsyu problemu $w19 ; та інші.

Тут змінними виступають $w11=[chy]; $w19 = vyrishyty|rozvjazaty; $w24 = [ne]. Враховуючи, що кожна змінна може приймати 2 значення, кожна фраза буде мати 8 варіантів. Отже, для даного ТР отримаємо 8·24=192 фрази, з урахуванням змінних.

Для вирішення поставленої задачі, а саме отримати правильний переклад української фрази, вимовленої у вільному стилі, на англійську, ми повинні визначити, до якого типу речень, а відповідно і типу змісту відноситься розпізнана фраза, відібрати відповідну англійську фразу та озвучити її.

Розпізнавання фраз (речень) проводилося на основі пофонемного розпізнавача за умов обмеженої та вільної послівних граматик. Обмежена граматика була задана для кожного ТР за допомогою LISP-структур.

Для експерименту довільним чином було вибрано 100 фраз серед згенерованих 4337, до яких було застосовано алгоритми пофонемного розпізнавання мовленнєвих сигналів в умовах обмеженої та вільної граматик відносно слів [1, 3]. При аналізі результатів розпізнавання враховувалися ті речення, які відрізнялися від вхідного речення не більше ніж на 2 словесних вставки/випадання або на думку експерта не відрізнялися за змістом. Підрахунок результатів наведено в таблиці.

Тип граматики	Відсоток правильно розпізнаних речень з точністю до:			
	вставок/випадань			типу змісту
	0	1	2	
Обмежена	95	97	99	98
Вільна послівна	51	70	85	95

При використанні обмеженої граматики було отримано в середньому 97% результат розпізнавання за 30 хвилин. На вільній граматиці було отримано гірший результат – в середньому 68% правильно розпізнаних фраз з точністю до не більш ніж двох випадань. Хоча в цьому випадку алгоритм працює значно швидше: результат отримано за півтори хвилини.

Враховуючи ці результати, було розроблено демонстраційне програмне забезпечення для перекладу фрази, вимовленої українською мовою, на англійську мову. При цьому слідування слів в українській фразі може бути будь-яким із допустимих. Фразі, вимовленій українською мовою, за допомогою евристичного алгоритму, заснованого на аналізі ключових слів, ставиться у відповідність англомовний тип змісту або речення, а перше речення цього типу змісту оголошується результатом перекладу.

5. Висновки

В роботі були розглянуті питання смислової інтерпретації усномовного сигналу з урахуванням специфіки слов'янських мов: відносно вільний порядок слідування слів і їх змінюваність.

Опрацьовано спосіб задання множини допустимих речень, що відповідають одному і тому ж змістові, шляхом побудови ТР засобами LISP-структур. Розроблено програмні засоби побудови орієнтованих семантичних мереж за ТР і ТЗ при розпізнаванні та породження речень при синтезі відповіді розпізнавання.

Експериментальні дослідження показали високі результати розпізнавання та інтерпретації злитого мовлення в умовах обмеженої граматики, побудованої відповідно до ОСМ, та обнадійливі результати в умовах вільної граматики. Розпізнавання в умовах вільної граматики відбувається в реальному часі. Справедливим вбачається припущення, що часткове обмеження початково вільної граматики дасть змогу підвищити результати розпізнавання та інтерпретації змісту зі збереженням роботи алгоритму в реальному часі.

Запропоновано евристичний алгоритм віднесення результатів розпізнавання злитого мовлення до типу речення і відповідно змісту. Його використання дає прийнятні результати.

На основі експериментальної моделі розроблено демонстраційну модель усного перекладу з української мови на англійську в межах предметної області.

При генеруванні речень за LISP-структурами отримуються в тому числі і речення, які є менш типовими у мовленні. На майбутнє це слід дослідити. Корисним вбачається розроблення алгоритму автоматичної побудови LISP-структур за заданими реченнями.

Одні і ті ж самі тексти з різною інтонацією можуть виражати як питальне речення, так і розповідне. Отже, в подальшій роботі слід дослідити можливість розпізнавання інтонації (просодики) з метою автоматичного розставляння розділових знаків у розпізнаних фразах.

Література

1. Т.К. Винцюк. *Анализ, распознавание и смысловая интерпретация речевых сигналов*. – Киев. Наукова думка, 1987.
2. Т.К. Винцюк. *Учет синтаксиса языка при распознавании слитной речи*. – Киев. Институт кибернетики, 1975.
3. Young S.J. et al., *HTK Book, version 3.1*, Cambridge University, 2002.