

Аналіз зразків машинного перекладу у світлі конфігуративної моделі мовлення

Мирон Федоришин

Кафедра іноземних мов, Національний університет «Львівська політехніка», вул. С. Бандери, 12, Львів, 79013, УКРАЇНА,
E-mail: myfedo@polynet.lviv.ua

Examples of Japanese-English machine translations are analyzed in the light of a configurative linguistic communication model that takes into account both linguistic and extra-linguistic factors of communication to reveal the peculiarities of predication.

Ключові слова: машинний переклад, мовлення, модель, предикація, комунікація, екстралінгвістичний фактор, конфігуратема, конфігурація.

Вступ

Машинний переклад досяг значних успіхів і має практичне застосування, проте в ньому ще багато місця для вдосконалення. На попередній конференції „Людина. Комп'ютер. Комунікація. 2013” доповідь „Конфігуративна модель мовлення: інструмент для аналізу мовних явищ” ми закінчили такими словами: „Придатна ця модель і для аналізу такого важливого явища як машинний, комп'ютерний переклад, на чому маємо намір зосередитися у майбутньому дослідженні” [6: 38]. Тож у цій праці пропонуємо спробу такого аналізу. Але спочатку коротко згадаємо основні елементи запропонованої моделі.

I. Основні засади конфігуративної моделі мовлення

Ідею конфігуративної моделі мовлення пропонуємо й розвиваємо на матеріалі англійської та японської мов у публікаціях [2 – 11], суть яких узагальнимо кількома постулатами та формулами і графічним синтезом на Рис. 1 і 2.

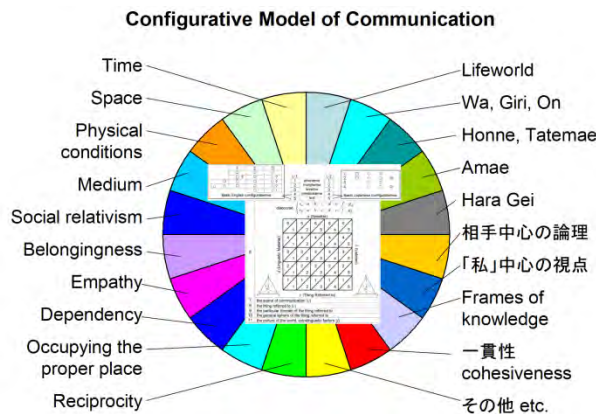


Рис. 1 Конфігуративна модель мовлення.

Основним постулатом нашого підходу є твердження про те, що ВСЕ у нашому світі – і матеріальне, і ідеальне – має форму та будову, що актуалізуються у часі та просторі. А те, що має форму і будову, зазвичай має складові частини. Мовленнєву дію у найпростішому вигляді описуємо формулою:

$$r = a \times b \times c \times d, \quad (1)$$

де a – мовець, b – слухач, c – предмет мовлення, d – мовний матеріал, r – результат мовленнєвого акту або дискурс, а для унаочнення взаємодії чотирьох базових чинників впроваджуємо поняття комунікаційного квадрата та подаємо його зображення з умовною шкалою рівня знання учасниками мовленнєвої дії предмета мовлення та мови (мовного матеріалу) (див. Рис. 1).

Долучивши місце дії e до Рівн. (1) отримуємо повнішу формулу мовлення:

$$r = a \times b \times c \times d \times e. \quad (2)$$

Стосовно мовного матеріалу також використовуємо поняття конфігуративи речення, яке визначаємо як розташування членів речення відносно присудка згідно з керуванням, і за допомогою табличок схематично показуємо конфігуративи для англійської та японської мов; а далі об'єднуємо мовний матеріал та предмет мовлення в ієрархічну систему з умовних 11 рівнів, що має картину світу в основі (див. Рис. 1: 1 – фонема, 2 – морфема, 3 – лексема, 4 – предикатема або речення, 6 – текст, 7 – місце дії, 8 – предмет мовлення, 9 – окрема галузь знань предмета мовлення, 10 – загальна галузь знань предмета мовлення, 11 – картина світу з екстралінгвістичними факторами).

Оскільки на характер висловлювання мовця та на сприйняття цього висловлювання слухачем-адресатом впливає багато позамовних чинників, то вводимо поняття екстралінгвістичного поля, в якому реалізується мовленнєвий акт, і умовно окресливши його (поле) колом отримуємо універсальну конфігуративну модель мовлення, в якій екстралінгвістичне поле розбито на конкретні позамовні чинники, що визначають характер висловлювання мовця та сприйняття такого висловлювання слухачем-адресатом, і як приклад ілюструємо загальну структуру екстралінгвістичного поля для японської мови. Теоретично кількість секторів/чинників необмежена, і деякі з них можна розглядати як постійні величини, інші як змінні, а також як відомі та невідомі. Позначивши релевантні екстралінгвістичні фактори символом f (а у дійсності ж $f = f_1 + f_2 + f_3 + \dots + f_n$), дописуємо їх до Рівн. (2) й отримуємо ширшу формулу мовлення:

$$r = a \times b \times c \times d \times e \times f. \quad (3)$$

Зважаючи на фундаментальність запозиченого у Е. Гуссерля [1] поняття життєсвіту, вважаємо за

доцільне у нашій моделі виділити його з-поміж інших екстралінгвістичних чинників і, позначивши його символом g , долучити до Рівн. (3), щоб отримати формулу:

$$r = a \times b \times c \times d \times e \times f \times g, \quad (4)$$

яка описує явище мовлення у загальнотеоретичній абстракції.

І нарешті, конкретизувавши життєсвіт кожного з комунікантів та результат мовлення для кожного з них за допомогою нижнього індекса до символів g та r , відповідно: g_a – життєсвіт мовця, g_b – життєсвіт слухача, r_a – дискурс для мовця, r_b – дискурс для слухача-адресата, мовленнєвий акт у процесі комунікації записуємо системою рівнянь:

$$\begin{cases} r_a = a \times b \times c \times d \times e \times f \times g_a, \\ r_b = a \times b \times c \times d \times e \times f \times g_b. \end{cases} \quad (5)$$

У реальному процесі комунікації всі чинники мовлення діють інтегрально у динамічній взаємодії. Тож якщо з Рис. 1 провести фізичний експеримент, а саме – обертати багатокольоровий круг навколо осі з достатньою швидкістю, то за відомим законом змішання кольорів виходить „ахроматичний” колір – білий на Рис. 2. Отриманий результат назвемо абсолютною моделлю мовлення, яка також придатна для теоретичних міркувань та практичної ілюстрації природи мовлення.

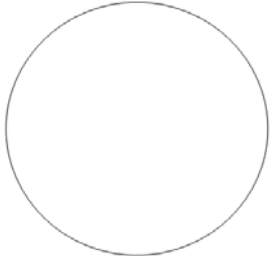


Рис. 2 Абсолютна модель мовлення.

Отже, запропонована конфігуративна модель мовлення охоплює усі чинники комунікації, лінгвістичні та екстралінгвістичні, занурені у життєсвіт. Тепер, опираючись на цю конфігуративну модель мовлення, пропонуємо таке визначення мовлення (лінгвістичної комунікації). Мовлення – це мовна взаємодія між комунікантами, мовцем та слухачем-адресатом, що відбувається у певному місці у певний час і виникає в екстралінгвістичному полі з екстралінгвістичної необхідності-потреби як явище життєсвіту.

II. Аналіз зразків машинного перекладу

Аналіз зразків машинного перекладу проведемо з точки зору предикації на прикладі японсько-англійського перекладу. На основі запропонованої конфігуративної моделі мовлення предикацію визначаємо як співвідношення висловлювання до

дійсності через речення, побудоване на конфігуратемі, з урахуванням екстралінгвістичних факторів на ґрунті життєсвіту. Отже, у такому підході до трактування мовних явищ поняття конфігурації охоплює всі чинники комунікації – лінгвістичні та екстралінгвістичні – занурені у життєсвіт. Виокремлюємо два типи предикації: експліцитну та імпліцитну. Експліцитна предикація – це речення, що має у своєму складі підмет, присудок та усі релевантні другорядні члени речення. У реченні з імпліцитною предикацією деякі члени речення „фізично відсутні” з огляду на особливості конкретної мови. У цьому контексті побіжно зазначимо, що імпліцитна предикація та еліipsis поняття не тотожні.

Для аналізу ми взяли уривок статті з японської газети The Asahi Shimbun, що виходить двома мовами – японською [14] та англійською [15]. Нижче подаємо оригінал японською мовою, його транскрипцію латинкою, цей же уривок з англійської версії газети і два зразки машинного перекладу, виконані сервісами автоматичного перекладу Google Translate [16] та Weblib Eigo Honyaku [17], що є у вільному доступі в Інтернеті (нумерація речень наша – М. Ф.).

Японський оригінал:

1. カンヌの映画祭で賞を受けた「そして父になる」は、親子の絆というものを考えさせる。
2. 病院で赤ちゃんを取り違えられた2組の夫婦が登場する。
3. 子が6歳になって事実を知らされ、葛藤する。
4. 血のつながりか、一緒に過ごした時間か。
5. あつさり割り切れる話ではない。
6. 46年前に実際にあった例では、4歳のときに取り違えがわかった。
7. これ以上成長すると難しくなると判断され、生みの親に返した。
8. 双方の両親の煩悶は如何ばかりだったか。 [14]

Транскрипція оригіналу (Латинізація тексту сервісом Google Translate [16]):

1. Kannu no eigamatsuri de shō o uketa “soshite chichi ni naru” wa, oyako no kizuna to iu mono o kangaesaseru.
2. Byōin de akachan o torichigaerareta 2-kumi no fūfu ga tōjō suru.
3. Ko ga 6-sai ni natte jijitsu o shirasare, kattō suru.
4. Chinotsunagari ka, issho ni sugoshita jikan ka.
5. Assari warikireru hanashi dewanai.
6. 46-Nen mae ni jissai ni atta rei de wa, 4-sai no toki ni torichigae ga wakatta.
7. Kore ijō seichō suru to muzukashiku naru to handan sare, umi no oya ni kaeshita.
8. Sōhō no ryōshin no hanmon wa ikabakari datta ka.

Цей же уривок з англійської версії газети (підкреслено елементи предикації, словесно відсутні в японському оригіналі; підкреслення наше – М. Ф.):

1. The movie “Soshite Chichi ni Naru” (Like Father, Like Son), which won the Jury prize at the 2013 Cannes Film Festival, offers a keen insight into the bonds between parents and their children.
2. The story revolves around two couples whose babies were switched at birth at the hospital where they were delivered.
- 3 – 4. When their children turn 6, the parents are told of the mishap and are now faced with choosing between blood ties and the bonds they formed with the children they raised as their own.
5. It is not a simple question to which there is a

cut-and-dry answer. **6.** In an actual case that occurred 46 years ago, the mix-up was discovered when the children were 4. **7.** Based on the belief that it would be difficult to return the children to their biological parents when they got older, they were switched back. **8.** I can only imagine the emotional conflict the parents on both sides went through. [15]

Переклад сервісом Google Translate [16]:

1. "Become a father and" received the award at film festivals of Cannes, I make you think what's called parent-child bond. **2.** Couple of two sets of misinterpret the baby in the hospital appeared. **3.** Informed of the fact that the child has turned six years old, I will conflict. **4.** Or ties of blood, or time spent together. **5.** It is not a story to be easily divisible. **6.** In the example that was in fact 46 years ago, mix-up has been found in at the age of four. **7.** It is judged to be difficult and to grow further, it returns to the biological parent. **8.** Agony of both parents or had just how.

Переклад сервісом Weblio Eigo Honyaku [17]:

1. I let you think about (and which "and become father" which got a prize for a film festival of Cannes, what's called bond of parent and child. **2.** Two sets of couples that a baby was misunderstood at a hospital appear. **3.** A child becomes 6 years old and is informed of a fact and conflicts it. **4.** A connection of the blood or the time that I spent together. **5.** It is not a story divisible plainly. **6.** I really understood misunderstanding at the age of 4 years old in the example which there was 46 years ago. **7.** It was judged to become difficult when I grew up further and returned it to a biological parent. **8.** How would you like your agony of both parents?

Для зручності аналізу пронумеровані речення позначаємо латинськими літерами J – японський першотвір, E – переклад в англійській версії газети The Asahi Shimbun, G – машинний переклад Google Translate, W – машинний переклад Weblio Eigo Honyaku. Порівнюючи японський оригінал з текстом англійської версії газети, бачимо, що англійські речення завдяки доданому підкресленим словам, що ніби пропущені в тексті оригіналу, мають експліцитну предикацію. Речення E-1 доповнено 2013 роком, який у нашій моделі комунікації (Рис. 1 і 2) перейшов з екстралінгвістичного поля у мовний матеріал. Речення G-1 та W-1 незграбні й невиразні, бо на нашу думку комп'ютерні програми цих перекладачів не можуть чітко визначити конфігуратему речення, хоча й, здається, враховують такі особливості предикації в японській мові як 「私」中心の視点 watashi chuushinno shiten (self/speaker-centered point of view, висловлювання з точки зору першої особи) [13: 23] та 相手中心の論理 aite chuushinno ronri (partner/listener-centered logic, висловлювання на основі логіки, що в центрі спілкування ставить співбесідника-адресата) [12: 8]. Цим, мабуть, і можна пояснити появу у цих двох реченнях машинного перекладу відсутніх в оригіналі займенників на позначення першої та другої особи однини. У реченні J-2 дієслово 登場する tōjō suru – виходити на сцену є імпліцитним присудком, що зв'язує це речення з попереднім. Предикатема E-2

виражена експліцитним реченням, створеним людиною. Речення G-2 та W-2 недосконалі, мабуть, тому що застосовані тут програми машинного перекладача працюють на рівні речення, а не тексту. Два імпліцитні речення J-3 та J-4 трансформовані людиною в одне хороше експліцитне E-3 – 4. Речення G-3 і 4 та W-3 і 4 претендують на експліцитність, але вийшли неокорисні. У J-5 підмет/тему речення пропущено з огляду на імпліцитний характер предикації в японській мові. E-5 – експліцитне речення, що характерно для англійської мови. G-5 та W-5 – експліцитні конфігуратемами, але неадекватні переклади. У J-6 слово 4歳 yonsai – імпліцитний іменник на позначення віку людини. Знов-таки англійське E-6 – експліцитне речення. Варіант G-6 прийнятний з огляду на те, що комп'ютерна програма перекладу застосувала правильні конфігуратемами. А речення W-6 – неадекватний переклад, бо перекладач, здається, запрограмований на таку вже згадану суттєву характеристику японської предикації як висловлювання з точки зору першої особи [13: 23], що не відповідає цьому випадку. В реченні J-7 предикація імпліцитна: тут відсутні підмети до присудків seichō suru – рости і muzukashiku naru – ускладнюватися; а ще в цьому японському реченні немає підмета і прямого додатка до присудка kaeshita – повернули. І знову ж, E-7 – експліцитне речення. Варіанти G-7 та W-7 – неадекватні переклади, оскільки машинні перекладачі не можуть правильно перекладати речення з імпліцитною предикацією. Останнє зауваження стосується і речення 8, яке нижче розглянемо окремо.

Цікаве спостереження ми зробили при повторному застосуванні використаних тут онлайн-перекладачів до того самого тексту. Вищенаведені машинні переклади було зроблено майже рік тому для праці [11]. Коли ж цього разу ми спробували перевірити переклади, то у Google Translate для цілого уривка без нумерації окремих речень отримали такий варіант:

1. Received the award at the Cannes Film Festival "consisting and father", I make you think a thing called parent-child bond. **2.** Two sets of couples that have been mistaken the baby in the hospital appeared. **3.** Child is informed of the fact turned six years old, I will be conflict. **4.** Or connection of blood, or spent time together. **5.** It is not the easily divisible story. **6.** In fact there were cases in '46 before, was found to mix-up in the age of four. **7.** It is determined that it becomes difficult to grow and further, it returns to the creator. **8.** Or agony of both parents was just how.

Потім ми провели експеримент з введенням тексту до перекладача у різних формах – цілий абзац без нумерації речень, абзац з пронумерованими кількома способами (після цифри ставили японську крапку, „європейську” крапку, круглу дужку) реченнями, окремі речення з номером та без номера. Сервіс Google Translate, на приклад, для речення 6 (4 6年前に実際にあった例では、4歳のときに取り違えがわかった。46-Nen mae ni jissai ni atta rei de wa,

4-sai no toki ni torichigae ga wakatta.) видав такі варіанти (з урахуванням першого минулорічного перекладу):

1) In the example that was in fact 46 years ago, mix-up has been found in at the age of four. 2) In fact there were cases in '46 before, was found to mix-up in the age of four. 3) In fact there were cases in 6.46 years ago, was found to mix-up in the age of four. 4) In fact, there were cases of 46 years ago, was found to mix-up in the age of four. 5) In fact there was Example 6) '46 ago, was found to mix-up in the age of four. 6) In fact met examples to 6.46 years ago, was found confusion to the age of four.

У світлі конфігуративної моделі мовлення „нові” варіанти перекладу за великим рахунком майже не відрізняються від першого стосовно предикації та конфігуративи, проте у кожному реченні є певні відмінності. Це свідчить про те, що з одного боку програма машинного перекладу перебуває у стадії розвитку, а з другого про те, що без урахування позамовних чинників, мабуть, неможливо правильно запрограмувати предикацію та конфігуративи.

Для сервісу Weblio Eigo Honyaku у всіх варіантах вводу тексту ми отримали переклади ідентичні вищенаведеному першому. Це може бути свідченням того, що програми Google Translate та Weblio Eigo Honyaku засновані на різних засадах.

А ще на прикладі орфографічної трансформації речення 8 виявляємо роль графіки, запису слів графічними знаками, у машинному перекладі. Японська система письма має назву „кандзі кана мадзірі бун”, що дослівно означає „змішане ієрогліфічно-абеткове письмо”, в якому слова записують ієрогліфами кандзі та складовою абеткою кана. Результат машинного перекладу може бути різний залежно від того, чи слово записане ієрогліфом чи канюю. Так у нашому прикладі речення 8 в оригіналі [14] має таку графічну форму: 双方の両親の煩悶 (はんもん) はいかばかりだったか。 Sōhō no ryōshin no hanmon (hanmon) wa ika bakari datta ka. У цьому реченні в дужках літерами хірагани подано читання слова з рідковживаним ієрогліфом – hanmon, а питальне слово іка (як? яким чином?) записано канюю, хоча має ієрогліфічний варіант 如何, який можна віднести до книжного письмового стилю. У праці [11], де питання машинного перекладу відіграло допоміжну роль, ми свідомо змінили написання слова іка з буквеного на ієрогліфічне, щоб отримати максимально наближений до реального змісту результат. Тепер же виконуємо машинний переклад речення 8 зі збереженням правопису оригіналу й отримуємо такі результати: „Of both parents agony (mottle) or was Ika bakari.” – від Google Translate та „Was the agony (はんもん) of both parents only a cuttlefish?” – від Weblio Eigo Honyaku. Порівняймо ці переклади з варіантами G-8 та W-8 і побачимо, як по-різному Google та Weblio трактують слово іка залежно від його графічної форми. У випадку запису літерами кани Google з прислівників іка та bakari (тільки) утворив іменник на позначення якоїсь формальної власної назви Ika bakari, а Weblio

прийняв прислівник іка за омонімічний йому іменник іка – кальмар, каракатиця. Не знаємо, чи програми застосованих тут машинних перекладачів мають в своєму складі такий елемент як конфігуратива, але можемо вважати, що у цих програмах графічна форма слова має перевагу над конфігуративою речення, тоді як у мові людини власне через конфігуративу реалізується предикація, тобто визначається значення слів з урахуванням позамовних чинників.

Висновки

З проведеного аналізу можна зробити висновок, що в японській мові переважає імпліцитна предикація, а англійській мові властива експліцитна предикація. Тип предикації тісно пов'язаний з конфігуративною будовою мови та з екстралінгвістичними факторами мовлення на ґрунті життєсвіту. Імпліцитна предикація може свідчити про те, що комуніканти глибоко вкорінені у екстралінгвістичне поле комунікації з акцентом на суб'єктивність світобачення, а експліцитна предикація може бути доказом налаштованості мовця на об'єктивність.

Стосовно ж машинного перекладу бачимо, що використані нами перекладачі досить прийнятно справляються з реченнями, зокрема простими, з експліцитною предикацією, а переклад речень з імпліцитною предикацією та складнопідрядних виходить неокоривний і тяжкозрозумілий. Також з наведених прикладів робимо висновок, що застосовані для цього аналізу комп'ютерні програми перекладають на рівні речення і не „бачать” тексту, а в самих реченнях можуть віддавати перевагу графічній формі слова над конфігуративою речення; машинний переклад не виконує предикації, а лише комп'ютерно опрацьовує матеріальну форму речення. Тож, мабуть, необхідно впровадити в теорію й інкорпорувати в програму комп'ютерного перекладу поняття дискурсного поля як посередника узгодження та взаємодії як членів речення, так і окремих речень як елементів цілого тексту, що вимагає врахування екстралінгвістичних факторів та життєсвіту. А порівняння варіантів перекладу, виконаних різними машинними перекладачами, вказує на існування різних шляхів вирішення задач комп'ютерного опрацювання живої мови людини.

Запропонована конфігуративна модель мовлення добрий інструмент для аналізу мовних явищ, що може сприяти вдосконаленню якості машинного перекладу. Вона також стала дієвим засобом унаочнення в нашій практиці викладання японської та англійської мов.

Література

1. Гуссерль Э. Кризис европейских наук и трансцендентальная феноменология / Эдмунд Гуссерль. – Спб.: Владимир Даль, 2004. – 400 с.
2. Федоришин М. A Configurative Model Approach to Linguistic Communication Study: English and Japanese / М. Федоришин // Мовні і концептуальні картини

- світу. Вип. 35. – К.: ВПЦ „Київський університет”, 2011. – С. 92 – 99.
3. Федоришин М. A Configurative Model of Linguistic Communication: Extralinguistic Factors with Reference to Japanese / М. Федоришин // Мовні і концептуальні картини світу. Вип. 40. – К.: ВПЦ „Київський університет”, 2012. – С. 138 – 143.
4. Федоришин М. Місце термінології в конфігуративній моделі мовлення / М. Федоришин // Збірник наукових праць учасників XII Міжнародної наукової конференції „Проблеми української термінології СловоСвіт”, м. Львів, 27-29 вересня 2012 р. – С. 25 – 29.
5. Федоришин М. С. Конфігуративна модель мовлення у контексті вивчення іноземних мов / М. С. Федоришин // Наукові записки. Серія „Філологічна”. – Острог: Видавництво Національного університету „Острозька академія”. – Вип. 33. – 2013. С. 229 – 233.
6. Федоришин М. Конфігуративна модель мовлення: інструмент для аналізу мовних явищ / Мирон Федоришин // Людина. Комп'ютер. Комунікація: Збірник наукових праць / за ред. О. П. Левченко. – Львів: Видавництво Львівської політехніки, 2013. С. 33 – 38.
7. Федоришин М. A Configurative Model of Linguistic Communication: Structure of the Speaker and of the Listener / М. Федоришин // Мовні і концептуальні картини світу. Вип. 45. – К.: ВПЦ „Київський університет”, 2013. – С. 95 – 101.
8. Федоришин М. С. Конфігуративна модель мовлення у контексті вивчення іноземних мов: структура комунікантів / М. С. Федоришин // Наукові записки. Серія „Філологічна”. – Острог: Вид-во Національного університету „Острозька академія”. – Вип. 42. – 2014. С. 238 – 242.
9. Федоришин М. Конфігуративна модель мовлення: термінологія у контексті структури комунікантів / М. Федоришин // Збірник наукових праць учасників XIII Міжнародної наукової конференції „Проблеми української термінології СловоСвіт”, м. Львів, 25-27 вересня 2014 р. – С. 5 – 9.
10. Fedoryshyn, M. A Configurative Model of Linguistic Communication: the Implication for Teaching/Learning Foreign Languages / М. Fedoryshyn // Людина. Комп'ютер. Комунікація: збірник наукових праць. Вид-во НУ „Львівська політехніка”. – Львів, 2010. – С. 192 – 196.
11. Fedoryshyn, M. A Configurative Model of Linguistic Communication: Implicit vs. Explicit Predication / М. Федоришин // Мовні і концептуальні картини світу. Вип. 49. – К.: ВПЦ „Київський університет”, 2014. – С. 93 – 98.
12. 堀川直義. 話しことばにおける日本人の論理 // Itasaka G. Modern Japanese. An Advanced Reader. Volume I Text / Gen Itasaka, Seiichi Makino and Kikuko Yamashita. – Tokyo: Kodansha International, 1974. – p. 8 – 17.
13. 森田良行. 日本人の発想、日本語の表現 – 東京 中公新書 1998. – 242 p.
14. 天声人語: 朝日新聞デジタル。2013年11月28日。 [Електронний ресурс]. Режим доступу: <http://www.asahi.com/paper/column>.
15. VOX POPULI: The consequences of being switched at birth // The Asahi Shimbun, Nov. 28, 2013. [Електронний ресурс]. Режим доступу: <http://ajw.asahi.com/article/views/vox>.
16. Google Translate [Електронний ресурс]. Режим доступу: <https://translate.google.com.ua/#zh-CN/ja/>.
17. Weblio Eigo Honyaku [Електронний ресурс]. Режим доступу : <http://translate.weblio.jp/>.