

Monika Szczygielska

Dostępni.eu

monika@dostepni.eu

Łukasz Stanisław Dutka

Uniwersytet Warszawski

lukasz.dutka@uw.edu.pl

Historia napisów na żywo tworzonych metodą respakingu w Polsce

1. Napisy na żywo

Napisy na żywo (*live subtitling*) to jeden z rodzajów przekładu audiowizualnego [Díaz Cintas, Remael 2007], a jednocześnie jedna z usług dostępowych (*access services*) umożliwiających lub ułatwiających osobom z niepełnosprawnościami korzystanie z treści audiowizualnych [Díaz Cintas 2005; Jankowska, w druku]. Jest to tekstowy zapis wypowiedzi i innych, ważnych dla zrozumienia treści elementów dźwiękowych. W odróżnieniu od napisów przygotowywanych wcześniej w całości (*pre-recorded*) lub częściowo (*semi-live*), napisy na żywo są tworzone w czasie rzeczywistym i zwykle ukazują się na ekranie z kilkusekundowym opóźnieniem¹.

Napisy na żywo są wykorzystywane podczas wydarzeń, np. konferencji, lub w programach telewizyjnych nadawanych na żywo [Romero Fresco 2012: 138]. W zależności od wyglądu wyróżnia się zasadniczo dwa rodzaje tekstowego zapisu na żywo: napisy, złożone z maksymalnie 2–3 linii, wyświetlane najczęściej na dole obrazu audycji telewizyjnej lub transmisji internetowej, oraz symultaniczny przekaz tekstowy (SPT),

¹ Nadawca telewizyjny lub dostawca transmisji na żywo w streamingu może zdecydować się na opóźnienie transmisji o określoną liczbę sekund, co pozwala zmniejszyć lub całkowicie wyeliminować opóźnienie napisów.

wyświetlany w miejscu wydarzenia jako blok tekstu (6–10 wierszy) zajmujący cały ekran, widoczny dla bezpośrednich uczestników wydarzenia. Ponieważ napisy na żywo i SPT są tworzone zasadniczo w ten sam sposób, a różnią się sposobem prezentacji, organizatorzy wydarzeń często stosują te terminy wymiennie. W niniejszym artykule termin „napisy na żywo” stosowany jest do tekstu na żywo w transmisji online i w telewizji, a termin SPT do tekstu wyświetlanego w miejscu wydarzenia na żywo.

W odróżnieniu od napisów przygotowywanych z wyprzedzeniem, naturalne zjawiska przy tworzeniu napisów na żywo to opóźnienie i redukcja, które powinny być możliwie jak najmniejsze. Te dwa parametry stanowią główny przedmiot badań nad jakością napisów na żywo [Romero Fresco 2012; 2018].

Z punktu widzenia odbiorców za napisy na żywo mogą zostać uznane napisy częściowo przygotowywane z wyprzedzeniem, ale nadawane na żywo (*semi-live*). Tego rodzaju napisy towarzyszą programom na żywo, które nie zawierają elementu improwizacji i są realizowane ściśle według scenariusza lub zawierają fragmenty nagrane wcześniej [European Broadcasting Union 2004: 10]. Przykładem takiej audycji są programy informacyjne, które składają się z materiałów reporterskich nagranych i zmontowanych z wyprzedzeniem oraz zapowiedzi i komentarzy prezentera, nagrywanych na żywo, ale zwykle zredagowanych wcześniej i odczytywanych z promptera. Napisy do takich programów lub ich fragmentów mogą zostać przygotowane z wyprzedzeniem. Natomiast ich emisja jest sterowana ręcznie (w trakcie programu na żywo osoba „wypuszczająca” napisy decyduje o ukazaniu się, a następnie zniknięciu napisu poprzez naciśnięcie przycisku).

W praktyce większość audycji nadawanych na żywo łączy w sobie fragmenty nagrywane na żywo i przygotowane z wyprzedzeniem, wypowiedzi odczytywane ze scenariusza oraz improwizowane. W rezultacie napisy nadawane na żywo najczęściej są hybrydą napisów przygotowanych z wyprzedzeniem i tworzonych na żywo. Napisy na żywo są najczęściej opracowywane w ramach tego samego języka, a więc są formą przekładu wewnątrzjęzykowego (zob. pojęcie *intralingual translation*

[Jakobson 1959]). Mogą też być tworzone przy wykorzystaniu przekładu z języka obcego [projekt ILSA]².

Jako usługa dostępowa (*access service*), czyli jedna z form udostępniania treści w ramach dostępności (*accessibility*) dla osób z niepełnosprawnościami, napisy na żywo umożliwiają lub ułatwiają dostęp do treści audiowizualnych osobom słabosłyszącym i niesłyszącym [Eugeni 2007; 2008], a także osobom uczącym się języka, obcokrajowcom i innym [Díaz Cintas, Remael 2007]. Oficjalnie blisko milion³ Polaków boryka się z poważnymi problemami ze słuchem. Z analizy Europejskiej Federacji Osób Słabosłyszących (*European Federation of Hard of Hearing People*, EFHOH) wynika, że co najmniej 51 mln osób, czyli 10% mieszkańców Europy, ma problemy ze słuchem. Według Światowej Organizacji Zdrowia 15% dorosłych na świecie i co czwarta osoba po 65. roku życia ma niedosłuch. W tym kontekście polskie dane wydają się więc zaniżone⁴.

W przeciwieństwie do osób Głuchych, słabosłyszący i późno ogłuchli Polacy nie są rodzimymi użytkownikami języka migowego. W zdecydowanej większości posługują się językiem polskim. Coraz częściej odbiorcami napisów na żywo stają się także użytkownicy słyszący, np. w sytuacjach, gdy nie można użyć dźwięku – podczas oglądania transmisji z konferencji w miejscu pracy lub wiadomości telewizyjnych w głośnych miejscach: na dworcu, w pubie, w pociągu, na siłowni. Jeśli napisy na żywo są formą tłumaczenia międzyjęzykowego, skorzystają z nich odbiorcy nieznający języka obcego [Robert, Remael 2017].

Choć napisy na żywo są stosowane na świecie od kilkudziesięciu lat (audycje telewizyjne z napisami na żywo zaczęto nadawać po raz pierwszy w 1982 r. w USA) [Romero Fresco 2012: 22–40], w Polsce znalazły zastosowanie stosunkowo niedawno – [por. Szarkowska 2008; 2016]. Rozwój polskich napisów na żywo był wstrzymywany przede wszystkim przez dwa główne problemy: stan prawny i technologię. W wielu krajach na świecie rozwój oprogramowania nastąpił w związku z wprowadzeniem

² <http://www.ilsaproject.eu/project/>, dostęp dn. 14 listopada 2018.

³ Wg danych GUS 14% osób niepełnosprawnych w wieku 15–70 lat ma wadę słuchu. *Stan zdrowia ludności w Polsce 2009*, http://stat.gov.pl/cps/rde/xbcr/gus/ZO_stan_zdrowia_2009.pdf, dostęp dn. 14 listopada 2018.

⁴ *Hearing Loss: The Statistics 2015*, EFHOH 2015, <https://efhoh.org/wp-content/uploads/2017/04/Hearing-Loss-Statistics-AGM-2015.pdf>, dostęp dn. 14 listopada 2018.

wymogu prawnego udostępniania wysokiej kwoty audycji telewizyjnych z napisami, co wytworzyło konieczność przygotowywania napisów także do programów na żywo⁵.

2. Regulacje prawne

Do 2011 r. polskie prawo medialne nie wymagało od nadawców telewizyjnych zapewniania żadnych napisów. W latach 2012–2018 nadawcy telewizyjni byli zobowiązani ustawowo do zapewniania minimum 10% kwartalnego czasu nadawania z napisami, audiodeskrypcją i tłumaczeniem migowym łącznie⁶. Realizacja tak niskiego progu nie wymagała od nadawców stosowania napisów na żywo. Zmiany wprowadzone w prawie w 2018 r.⁷ zobowiązują nadawców do nadawania do 2024 r. minimum 50% audycji z wszystkimi usługami dostępu łącznie. W przypadku programów informacyjnych minimum 44% audycji musi zostać opatrzonych napisami.

Nowelizacja ustawy o radiofonii i telewizji z 25 marca 2011 r. w żaden sposób nie odniosła się do napisów na żywo. Po wejściu w życie ustawy z inicjatywy Krajowej Rady Radiofonii i Telewizji (KRRIT) nadawcy zdecydowali się na samoregulację. W 2013 r. podpisali porozumienie⁸, w którym napisy na żywo zostały zdefiniowane jako jedna z form napisów dla niesłyszących. W dokumencie nie określono „wymiaru czasowego usług dedykowanych dla osób z dysfunkcją słuchu”. Zaznaczono jedynie, że ilość napisów powinna dopełniać określony minimalny

⁵ *State of subtitling access in EU 2015 Report*, EFHOH 2015 <https://efhoh.org/wp-content/uploads/2017/01/EFHOH-State-of-Subtitling-2015-English.pdf>, dostęp dn. 30 listopada 2018.

⁶ Ustawa z dnia 25 marca 2011 r. o zmianie ustawy o radiofonii i telewizji oraz niektórych innych ustaw, Dz.U. 2011 nr 85 poz. 459.

⁷ Rozporządzenie Krajowej Rady Radiofonii i Telewizji z dnia 15 listopada 2018 r. w sprawie ugodnień dla osób niepełnosprawnych z powodu dysfunkcji narządu wzroku i osób niepełnosprawnych z powodu dysfunkcji narządu słuchu w programach telewizyjnych, Dz.U. 2018 poz. 2261.

⁸ Porozumienie nadawców w sprawie sposobu realizacji obowiązków wynikających z art. 18a ustawy o radiofonii i telewizji dotyczących ugodnień w programach telewizyjnych dla osób z niepełnosprawnością wzroku i dla osób z niepełnosprawnością słuchu w programach telewizyjnych podpisane przez TVP SA, Telewizja POLSAT, Telewizja TVN, Telewizja PULS, Polskie Media (TV4), ATM Grupa SA, Stavka Sp. z o.o. z 5 czerwca 2013 roku, <http://www.krrit.gov.pl/krrit/aktualnosci/news,1208,porozumienie-nadawcow--dotyczace-ugodnien-dla-niepelnosprawnych.html>, dostęp dn. 20 listopada 2018.

poziom audiodeskrypcji i języka migowego do ustawowego 10% czasu nadawania audycji z „udogodnieniami”.

Napisy na żywo zostały zdefiniowane w tym porozumieniu jako „usługa umożliwiająca osobom z dysfunkcją narządu słuchu dostęp do treści dźwiękowych towarzyszących obrazowi w czasie prawie rzeczywistym (to jest z niewielkim przesunięciem czasowym), stosowana w audycjach informacyjnych i publicystycznych”. Definicja ta została powtórzona w Stanowisku KRRIT z 2016 r. w sprawie jakości i sposobu realizacji napisów dla niesłyszących w utworach audiowizualnych⁹. W stanowisku tym KRRIT wyraziła opinię, że audycje na żywo „dotychczas opatrzone były ograniczoną ilością napisów dla niesłyszących, co wynikało z trudności techniczno-organizacyjnych występujących podczas ich tworzenia w czasie rzeczywistym”. Podczas konsultacji społecznych tego stanowiska w 2015 r. eksperci zwracali uwagę, że definicja ta nie jest wyczerpująca. Podkreślali również, że brak napisów na żywo nie wynika z problemów technicznych, a jest następstwem niskiego wymogu prawnego:

Niewłaściwym wydaje się więc utrwalanie stereotypu o trudnościach techniczno-organizacyjnych związanych z tworzeniem napisów na żywo. Przy dzisiejszym rozwoju technologicznym, wiedzy na temat procesu *respeakingu* oraz w świetle prób podejmowanych przez telewizje taki zapis jest usprawiedliwianiem obecnego stanu i może działać demotywująco¹⁰.

Opinia KRRIT miała wpływ na polityków, wśród których nadal panuje przekonanie, że zapewnienie dostępności telewizji na poziomie wyższym niż 50% jest „niemożliwe ze względów technologicznych”. Ten argument powtarzał się m.in. w sejmowej odpowiedzi na petycję w sprawie

⁹ Stanowisko Krajowej Rady Radiofonii i Telewizji z 26 stycznia 2016 roku w sprawie jakości i sposobu realizacji napisów dla niesłyszących w utworach audiowizualnych, http://www.krrit.gov.pl/Data/Files/_public/Portals/0/konsultacje/2016/skan_stanowisko-krrit-z-26.01.2016.pdf, dostęp dn. 20 listopada 2018.

¹⁰ Stanowisko Fundacji Widzialni popierane przez: Fundację Promocji Kultury Głuchych „Kokon”, stowarzyszenie Instytut Polskiego Języka Migowego oraz Inicjatywę „Napisy plus” w sprawie projektu Stanowiska rekomendacyjnego KRRIT w zakresie jakości i sposobu realizacji napisów dla niesłyszących w utworach audiowizualnych z 17.12.2015.

zwiększenia dostępności audycji do 80% z 2016 r.¹¹ oraz w uzasadnieniu senackiego projektu nowelizacji ustawy o radiofonii i telewizji z 2017 r.¹².

Po raz pierwszy w akcie prawnym w Polsce napisy na żywo zostały zdefiniowane w Rozporządzeniu KRRIT z 2018 r. w sprawie udogodnień dla osób niepełnosprawnych z powodu dysfunkcji narządu wzroku i osób niepełnosprawnych z powodu dysfunkcji narządu słuchu w programach telewizyjnych¹³. Definicja napisów na żywo została wprowadzona obok definicji audiodeskrypcji, napisów i tłumaczenia na język migowy, i według niej to: „tekst w audycjach na żywo, pojawiający się na ekranie z nieznacznym przesunięciem czasowym, będący odzwierciedleniem dialogów lub wypowiedzi”.

Rozporządzenie KRRIT zostało wydane po nowelizacji ustawy o radiofonii i telewizji z dnia 22 marca 2018 r.¹⁴. Ustawodawca w art. 18a pkt 2 zobowiązał KRRIT do określenia w drodze aktu wykonawczego m.in. rodzajów „udogodnień” oraz ich udziału w łącznym czasie nadawania, w zależności od pory emisji, charakteru i rodzaju programu, uwzględniając jednocześnie potrzeby odbiorców, możliwości nadawców oraz rozwój technologii i samych „udogodnień”. Znowelizowana ustawa o radiofonii i telewizji stopniowo podnosi kwartalny czas nadawania audycji z „udogodnieniami” do co najmniej 50% w 2024 r. Minimalne progi procentowego kwartalnego czasu nadawania dostępnych audycji w okresie przejściowym wynoszą co najmniej 15% w 2019 r., 25% w 2020 r. i 2021 r.; 35% w 2022 r. i 2023 r. W reżimie ustawy o radiofonii i telewizji z 2011 r. i z 2018 r. KRRIT ma prawo ustalania niższego progu i zwalniania nadawców z obowiązku zapewniania dostępności. Dotyczy to głównie nadawców małych i specjalistycznych.

¹¹ Opinia nr 7 Komisji Kultury i Środków Przekazu dla Komisji do Spraw Petycji dotycząca petycji Pana Mariusza Nowaka w sprawie zmiany ustawy z dnia 29 grudnia 1992 r. o radiofonii i telewizji w zakresie art. 18a ust. 1, poprzez wprowadzenie udogodnień zapewniających dostępność audycji dla osób niepełnosprawnych na poziomie co najmniej 80% kwartalnego czasu nadawania programu z wyłączeniem reklam i telesprzedaży (BKSP-145-61/16) z dnia 20 lipca 2016, [http://orka.sejm.gov.pl/opinie8.nsf/nazwa/ksp_o7/\\$file/ksp_o7.pdf](http://orka.sejm.gov.pl/opinie8.nsf/nazwa/ksp_o7/$file/ksp_o7.pdf), dostęp dn. 20 listopada 2018.

¹² Uzasadnienie do uchwały Senatu RP z dnia 17 maja 2017 w sprawie wniesienia do Sejmu projektu ustawy o zmianie ustawy o radiofonii i telewizji (druk 366), <https://www.senat.gov.pl/download/gfx/senat/pl/senatuchwaly/3138/plik/366uch.pdf>, dostęp dn. 20 listopada 2018.

¹³ Rozporządzenie KRRIT z dn. 15 listopada 2018 w sprawie udogodnień dla osób niepełnosprawnych z powodu dysfunkcji narządu wzroku i osób niepełnosprawnych z powodu dysfunkcji narządu słuchu w programach telewizyjnych, Dz.U. 2018 poz. 2261.

¹⁴ Ustawa z dnia 22 marca 2018 r. o zmianie ustawy o radiofonii i telewizji, Dz.U. 2018 poz. 915.

Krajowa Rada ustaliła udział napisów w programach telewizyjnych w docelowym 2024 r. na poziomie 40% w programach uniwersalnych, oraz np. 42% w programach społeczno-religijnych, 44% w audycjach informacyjnych i publicystycznych. Stopniowy wzrost kwoty programów udostępnianych z napisami zasadniczo powinien wiązać się z koniecznością wykorzystania w szerszym zakresie napisów na żywo w audycjach telewizyjnych, chociaż to, jaka część z tych napisów będzie napisami na żywo, pozostaje w gestii nadawców, gdyż samodzielnie decydują oni, jakie treści i w jakich godzinach udostępniają. Pierwsze pełne sprawozdanie¹⁵ z realizacji obowiązków wynikających z ustawy Przewodniczący KRRiT przedstawi Sejmowi, Senatowi oraz Ministrowi Kultury i Dziedzictwa Narodowego w 2025 r. Omówi wówczas możliwości zwiększenia poziomu udogodnień dla osób niepełnosprawnych¹⁶. Sprawozdanie to raczej będzie traktowało ogólnie o napisach. Napisy na żywo nie funkcjonują jako oddzielna pozycja w statystyce udostępnianej przez KRRiT na stronie program.krrit.gov.pl. W programach telewizyjnych są komunikowane łącznie z innymi typami napisów: dialogowych, dla niesłyszących oraz na żywo, i oznaczane symbolem N.

Ani w Unii Europejskiej, ani w Polsce nie ma wymogu prawnego napisów na żywo podczas transmisji online. Choć przewidują je międzynarodowe standardy (WCAG 2.0. AA), ze względów technicznych w Polsce zostały pominięte w rozporządzeniu definiującym zakres dostępności stron internetowych podmiotów realizujących zadania publiczne¹⁷. Prace nad rozporządzeniem trwały w latach 2010–2011. Wówczas faktycznie nie było w Polsce wystarczająco dopracowanego programu rozpoznawania mowy polskiej, dzięki któremu byłoby to wykonalne. Ze względu na nadmierne obciążenie podmiotów publicznych obowiązek ten pominięto także w dyrektywie o dostępności publicznych stron www i aplikacji mobilnych¹⁸. Polska jest pewnie jednym z nielicznych krajów, gdzie

¹⁵ Co roku KRRiT przedstawia parlamentowi sprawozdanie z działalności i podaje ogólną statystykę z tytułu wykonania obowiązków wynikających z Art. 18 a Ustawy o radiofonii i telewizji.

¹⁶ Art. 4 Ustawy o radiofonii i telewizji.

¹⁷ Rozporządzenie Rady Ministrów z dnia 12 kwietnia 2012 r. w sprawie Krajowych Ram Interoperacyjności, minimalnych wymagań dla rejestrów publicznych i wymiany informacji w postaci elektronicznej oraz minimalnych wymagań dla systemów teleinformatycznych, Dz.U. 2012 poz. 526.

¹⁸ Dyrektywa Parlamentu Europejskiego i Rady (UE) 2016/2102, Dz.U. L 327/1 z 2.12.2016.

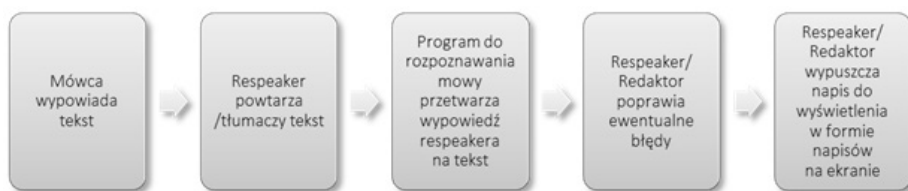
napisy na żywo od 2018 r. są wymagane w transmisjach online finansowanych z funduszy unijnych¹⁹.

2. Respeaking jako metoda tworzenia napisów na żywo

2.1. Proces tworzenia

Napisy na żywo tworzone są obecnie najczęściej z wykorzystaniem metody *respeakingu* [Romero Fresco 2018: 200]. *Respeaking* jest techniką tworzenia napisów na żywo z wykorzystaniem oprogramowania do rozpoznawania i przekładu mowy na tekst (*speech-to-text*). *Respeaker* słucha oryginalnych wypowiedzi wypowiedzianych na żywo i powtarza (*re-speaks*) je do specjalistycznego mikrofonu, jednocześnie dzieli tekst na krótsze zdania, dodaje niezbędne znaki interpunkcyjne i – na potrzeby osób niesłyszących – np. etykiety identyfikujące mówców oraz dokonuje niezbędnych zmian edycyjnych. Program do rozpoznawania mowy automatycznie rozpoznaje słowa *respeakera* i przekształca je na tekst. Po korekcie tekst jest wyświetlany na ekranie lub przekazywany do oprogramowania do tworzenia napisów telewizyjnych z możliwie jak najkrótszym opóźnieniem [Romero Fresco 2011: 1].

W zależności od modelu korektę może wykonywać sam *respeaker* lub moderator. Przykładowo we Francji, we Włoszech i w Polsce *respeakery* pracują w zespole z moderatorami. W Wielkiej Brytanii *respeaker* sam wykonuje korektę.



Rys. 1. Schemat procesu tworzenia napisów na żywo z wykorzystaniem metody *respeakingu* [Szarkowska 2015]

¹⁹ Załączniki nr 2 Standardy dostępności dla polityki spójności 2014–2020 do Wytyczne w zakresie realizacji zasady równości szans i niedyskryminacji, w tym dostępności dla osób z niepełnosprawnościami oraz zasady równości szans kobiet i mężczyzn w ramach funduszy unijnych na lata 2014–2020, https://www.funduszeuropejskie.gov.pl/media/55001/Zalacznik_nr_2_do_Wytycznych_w_zakresie_rownosc_zatwierdzone_050418.pdf, dostęp dn. 20 listopada 2018.

Respeaking został zastosowany po raz pierwszy w 2001 r. przez nadawców publicznych w Wielkiej Brytanii (BBC) i Belgii (flamandzkojęzyczna telewizja VRT) [Romero Fresco 2018: 200; Remael, Van Der Veer 2006]. W kontekście telewizyjnym jest także stosowany w Hiszpanii, Szwajcarii, Danii, Francji, Włoszech, Holandii i Kanadzie, a także w USA, gdzie ta metoda nazywana jest *voice writing* [Romero Fresco 2012: 22–40].

2.2. Metody alternatywne

Obok *respeakingu* najpopularniejsze na świecie metody to stenotypia oraz Velotype. Stosunkowo rzadko tekst zapisywany jest na zwykłej klawiaturze QWERTY [Lambourne 2006; Romero Fresco 2018].

Stenotypia w wariacie stosowanym w tworzeniu napisów na żywo wykorzystuje stenografię komputerową [Robson 2004]. Stenotypista pisze na klawiaturze akordowej (*chord keyboard*). Możliwość naciśnięcia wielu klawiszy jednocześnie oraz stosowane skróty pozwalają na zapisywanie wypowiedzi w tempie 200 słów na minutę (dla porównania *respeaking* pozwala na zapisywanie wypowiedzi w języku angielskim w tempie 140–160 słów na minutę) [Lambourne 2006]. Metoda ta jest szeroko stosowana w USA, a w mniejszym stopniu w Kanadzie, Australii i Wielkiej Brytanii [Downey 2008; Romero Fresco, Pöchhacker 2017]. Szkolenie stenotypisty trwa około dwóch lat, a brak przeszkolonych specjalistów jest głównym ograniczeniem w szerszym zastosowaniu tej metody w innych krajach. Wreszcie każdy język wymaga indywidualnego komputerowego systemu zapisu stenotypicznego. W przypadku wielu języków, w tym języka polskiego, takiego systemu dotychczas nie opracowano.

Velotype, podobnie jak stenografia komputerowa, opiera się na wykorzystaniu klawiatury akordowej oraz oprogramowania komputerowego. W odróżnieniu od stenografii komputerowej Velotype opiera się na zapisie sylab lub całych słów. Klawiatura Velotype została wynaleziona w Holandii w latach 80. i to właśnie tam ta metoda jest najpopularniejsza [De Korte 2006]. Stosowana jest też w Szwecji (pod nazwą handlową Veyboard), Belgii i Wielkiej Brytanii. Zmodyfikowany wariant klawiatury Velotype jest też znany pod nazwą handlową Palantype.

Zaletą Velotype w porównaniu ze stenotypią jest krótszy okres szkolenia (według deklaracji producenta – 6 miesięcy), natomiast wadą jest prędkość zapisu, która u przeszkolonych użytkowników waha się od 90 do 120 słów na minutę [Lambourne et al. 2004]. Według Romero Fresco [2011: 13] prezenterzy telewizyjni w języku angielskim mówią w średnim tempie 180 słów na minutę. Rozwiązaniem może być praca dwóch osób piszących równocześnie w skoordynowany sposób na klawiaturach Velotype.

Producent Velotype opracował warianty oprogramowania pozwalające na pracę w różnych językach, w tym w języku polskim, przy zachowaniu identycznego układu klawiszy. Jak dotąd Velotype nie został jednak zastosowany w Polsce do tworzenia napisów na żywo²⁰.

Podjęmowano próby tworzenia napisów na żywo z wykorzystaniem klawiatury QWERTY, ale nawet szybko piszące osoby nie są w stanie nadążyć za przeciętnym tempem mówienia. Lepsze rezultaty uzyskano z metodą *dual keyboard*, ale nawet przy zatrudnieniu dwóch osób piszących jednocześnie nie osiągnięto oczekiwanych efektów. W Polsce z powodu braku sprzętu, oprogramowania i wyszkolonej kadry ta metoda nie występuje, natomiast okazjonalnie stosuje się szybkie pisanie bezwzrokowe na klawiaturze QWERTY²¹.

Możliwe jest tworzenie napisów na żywo z wykorzystaniem automatycznego rozpoznawania mowy niezależnie od respeakera (tzw. *ASR – Automatic speech recognition*). Przeważnie rozpoznany tekst poddawany jest korekcie na żywo.

3. Technologia rozpoznawania i zamiany mowy na tekst (*speech to text*)

Badania nad rozwojem technologii rozpoznawania mowy na świecie trwają od połowy XX w. Początkowo efekty wykorzystywane były do wyszukiwania prostych informacji, wydawania krótkich i nielicznych poleceń głosowych, alternatywnych form komunikacji, następnie przepisywania mowy dyktowanej, a docelowo pełnej transkrypcji ludzkiej mowy niezależnie od mówcy i warunków otoczenia [Romero Fresco 2011].

²⁰ Bezpośrednia komunikacja z Wimem Gerbecksem, pracownikiem firmy Velotype.

²¹ Tę metodę stosuje Polska Fundacja Osób Słabosłyszących podczas swoich spotkań.

Każde oprogramowanie do rozpoznawania mowy działa w oparciu o podobny schemat, wykorzystując model akustyczny, słownik i model językowy [Benesty et al. 2007; Ziółko, Ziółko 2010]. Do stworzenia modelu akustycznego wykorzystuje się duże ilości nagrań oraz ich transkrypcje. Model ten wskazuje zależność między sygnałem dźwiękowym a zapisem dźwięków i jest wykorzystywany przy analizie sygnału dźwiękowego. System poprzez usunięcie szumów wyodrębnia sygnał mowy, a następnie dokonuje segmentacji. Pierwszym etapem segmentacji mowy jest podział na fonemy. Po zidentyfikowaniu ciągu fonemów system próbuje przyporządkować mu wyrazy, korzystając z wbudowanego słownika. Słownik jest listą słów zapisanych ortograficznie i fonemicznie, przy czym każda forma wyrazowa jest osobną jednostką, której przypisany jest jeden lub więcej wariantów wymowy. System może rozpoznać tylko te wyrazy, które znajdują się w słowniku. Jeśli w słowniku nie ma w pełni pasującego słowa, system wybierze inne słowo o zbliżonej wymowie. Jeśli do analizowanego ciągu fonemów (głosek) pasuje wiele różnych form wyrazowych, wybór wyrazu, który zostanie wykorzystany, dokonywany jest na podstawie modelu językowego. Opiera się on na statystycznym prawdopodobieństwie wystąpienia kolejnych elementów sekwencji, czyli określa prawdopodobieństwo wystąpienia poszczególnych wyrazów w danym kontekście. Spośród rozważanych wyrazów system na podstawie modelu językowego wybiera ten wyraz, który najczęściej jest używany w danym kontekście. Modele językowe tworzy się poprzez analizę korpusów tekstów. Im większa ilość przeanalizowanego tekstu, tym wyższa jakość modelu językowego. Przy czym model językowy będzie tym bardziej skuteczny, im bardziej rozpoznawany tekst będzie podobny składniowo i tematycznie do tekstów, na podstawie których utworzono model. Na przykład model stworzony na podstawie tekstów prawnych będzie lepiej radził sobie z tekstami z dziedziny prawa niż z tekstami medycznymi i *vice versa*.

W przeciwieństwie do innych technologii informacyjnych, które można stosunkowo łatwo przenieść na nowe rynki, technologia rozpoznawania mowy wymaga pracochłonnej adaptacji. Ze względu na specyfikę każdego języka, przeniesienie osiągnięć technologicznych z jednego języka do innego jest trudne – por. [Ziółko, Ziółko 2011: 327].

Rozpoznawanie mowy w językach słowiańskich, w tym polskim, z zasady jest trudniejsze niż w języku angielskim i daje gorsze efekty ze względu na charakter fleksyjny i dużą swobodę składniową – por. [Ziółko, Ziółko 2011; Brocki et al. 2012].

Skuteczność rozpoznawania mowy, często przez producentów nazywaną dokładnością (ang. *accuracy*), mierzy się najczęściej miarą WER (ang. *word error rate*), znaną też jako stopa błędów, którą wyraża się jako procent poprawnie rozpoznanych wyrazów [Dumouchel et al. 2011]. Wynik WER 98% oznacza, że na każde 100 wyrazów w tekście dwa zostały rozpoznane błędnie. Liczba błędnie rozpoznanych słów zależy od modelu akustycznego (który analizuje sygnał dźwiękowy i rozpoznaje fonemy), wbudowanego słownika (system może rozpoznać tylko te wyrazy, które znajdują się w jego słowniku), modelu językowego (w sytuacji, gdy rozpoznanemu ciągowi fonemów odpowiada wiele różnych wyrazów, model językowy określa prawdopodobieństwo wystąpienia poszczególnych wyrazów w danym kontekście), sprzętu (mikrofonu, karty dźwiękowej), zachowania użytkownika (usytuowania mówiącego względem mikrofonu, głośności mowy, poprawności wymowy) oraz otoczenia (szumu w tle) – por. [Ziółko 2011: 299].

Marasek [2012] zwraca uwagę, że oprogramowanie do rozpoznawania mowy działa tym lepiej, im mniejszy jest słownik (a więc zbiór wszystkich form wyrazowych, które program może rozpoznać). Dlatego język angielski jest rozpoznawalny lepiej niż język polski czy inne języki słowiańskie (liczba form wyrazowych jest mniejsza). Co istotne, „stopa błędów rozpoznawania mowy spontanicznej jest co najmniej dwukrotnie większa niż w przypadku dyktowania tekstu, stopa błędów jest wysoka dla konwersacji wielu mówców w trudnym akustycznie środowisku”²². Ponadto dotychczas automatyczne rozpoznawanie mowy działało najlepiej dla ściśle określonego mówcy. Są to wyniki badań sprzed wprowadzenia w oprogramowaniu mechanizmów uczenia się opartych o sieci neuronowe²³.

²² http://cesar.nytud.hu/events/polish_roadshow/krzysztof-marasek.pdf, dostęp dn. 30 listopada 2018.

²³ <https://elektronikab2b.pl/biznes/50090-rozpoznawanie-mowy-techniki-i-zastosowania>, dostęp dn. 30 listopada 2018.

Zasadniczy rozwój oprogramowania do rozpoznawania mowy polskiej nastąpił w drugiej dekadzie XXI w. Aktualnie oprogramowanie to ma praktyczne zastosowanie m.in. w branżach takich jak telekomunikacja i bankowość, monitoring mediów elektronicznych, archiwa cyfrowe, tworzenie dokumentacji na potrzeby administracji, medycyny czy sądownictwa.

4. Oprogramowanie do rozpoznawania mowy w języku polskim

Najpopularniejszym programem do rozpoznawania mowy na świecie jest Dragon amerykańskiej firmy Nuance. Pierwsza wersja programu Dragon Dictate pojawiła się w 1990 r. W 1997 r. do sprzedaży wszedł Dragon NaturallySpeaking; a najnowsza wersja do zastosowań komercyjnych sprzedawana jest pod nazwą Dragon Professional 15²⁴. Program ma wersje w 8 językach, ale nie jest dostępny w języku polskim. Od 2011 r. firma Nuance udostępnia w języku polskim aplikacje Dragon Dictation i Dragon Search, działające na urządzeniach mobilnych z systemem iOS, połączonych z Internetem. Pozwalają one jednak tylko na wyszukiwanie głosowe oraz dyktowanie krótkich wiadomości tekstowych i notatek.

Od 2018 r. na rynku jest co najmniej pięć komercyjnych programów do rozpoznawania mowy dyktowanej w języku polskim: Skrybot, MagicScribe, Newton Dictate, ARM, VoiceLab (ten ostatni jest dostępny tylko dla klientów korporacyjnych).

Dotychczas przy tworzeniu napisów na żywo stosowano programy MagicScribe i Newton Dictate 5, przy czym Telewizja Polska nie podaje do publicznej wiadomości, z jakiego oprogramowania korzysta.

MagicScribe przez producenta sprzedawany jest jako program do transkrypcji nagrań, edycji rozpoznanego tekstu, tworzenia napisów na żywo oraz do materiałów wideo. Producent deklaruje skuteczność rozpoznawania mowy na poziomie 95%. Jako przewagę konkurencyjną wyróżnia bezpieczeństwo danych, ponieważ oprogramowanie działa wyłącznie

²⁴ <https://www.nuance.com/about-us/newsroom/press-releases/dragon-new-releases-powered-by-deep-learning.html>, dostęp dn. 30 listopada 2018.

na urządzeniach klienta. Posiada słowniki dla branż: medycyna, psychologia, prawo, administracja²⁵.

Producent oprogramowania Newton Dictate jako pierwszy na polskim rynku deklaruje działanie rozpoznawania mowy w oparciu o sieci neuronowe. Zastosowanie nowej technologii ma przyspieszać rozpoznawanie mowy w obecnie dostępnej wersji 5 oprogramowania o 25% (w stosunku do wersji 4)²⁶. W wersji profesjonalnej program zawiera ponad 500 tys. najczęściej używanych słów i wyrażeń oraz wybrany słownik specjalistyczny z dziedzin takich jak prawo, histopatologia, interna, radiologia, administracja publiczna, mowa spontaniczna. Według deklaracji producenta skuteczność programu to „niemal 100%”²⁷.

5. Początki respeakingu w Polsce

Premiera respeakingu w Polsce miała miejsce w 2013 r. podczas konferencji w Sejmie RP. Od tamtej pory zespół Dostępni.eu korzystał z metody respeakingu m.in. podczas wydarzeń w Sejmie RP, Kancelarii Prezydenta RP, Biurze Parlamentu Europejskiego w Polsce. Podczas mszy kanonizacyjnej Jana Pawła II w 2014 r. po raz pierwszy tworzył napisy na odległość oraz z przekładu z języków obcych (tzw. *interlingual live subtitling*) [den Boer 2011; Neves 2009]. Pozytywnie zakończyły się także próby zastosowania tej metody w napisach nadawanych na żywo w transmisji online. Opisane poniżej przykłady pochodzą z praktyki zespołu Dostępni.eu, który w latach 2013–2017 pracował na polskim programie MagicScribe, a od 2018 r. z wykorzystaniem Newton Dictate.

5.1. Pierwsze testy respeakingu

Pierwsze konferencyjne zastosowanie napisów odbyło się 7 czerwca 2013 r. w Sejmie RP podczas konferencji „E-dostępność sektora publicznego w Polsce – osiągnięcia i wyzwania”, organizowanej przez Fundację

²⁵ <http://www.magicscribeshop.pl/czym-magicscribe-rozni-sie-od-innych-cz1>, dostęp dn. 30 listopada 2018.

²⁶ <http://www.dyktowanie.pl/newton-dictate-5-rozpoznawanie-mowy-napedzane-przez-neuronowe-serce/>, dostęp dn. 30 listopada 2018.

²⁷ Tamże.

Widzialni²⁸. We współpracy z Ośrodkiem Informatyki Kancelarii Sejmu RP testowane były wówczas napisy na żywo dwiema metodami: szybkiego pisania na klawiaturze oraz *respeakingu*. Odbyła się również pierwsza emisja napisów na żywo połączona z transmisją wideo online. W 2013 r. w środowiskach osób słabosłyszących, specjalistów ds. dostępności oraz w jednostce odpowiedzialnej w telewizji publicznej panowało przekonanie, że realizacja napisów na żywo dla języka polskiego jest niemożliwa z powodu braku właściwego oprogramowania. W polskiej wersji Wikipedii pod hasłem „rozpoznawanie mowy” widniała opinia: „Ogólnie należy przyjąć, że rozpoznawanie mowy polskiej działa poprawnie tylko dla pojedynczych słów lub dla ustalonych zbiorów scenariuszy dialogów” (stan na marzec 2014 r.)²⁹. Pierwszy test uznany został przez Ośrodek Informatyki Kancelarii Sejmu RP za ryzykowny i dlatego zdecydowano o dopuszczeniu do transmisji online transkrypcji z klawiatury. Pracownikom Sejmu wydawało się wówczas, że większą szansę mają napisy stworzone przez człowieka niż z udziałem oprogramowania do rozpoznawania i przekładu mowy na tekst, obsługiwanego wówczas przez dwie osoby z przygotowaniem aktorskim.

Równoległe do transmisji online napisy zostały pokazane podczas wydarzenia w postaci symultanicznego przekazu tekstowego (SPT). Napisy były wyświetlane na dwóch ekranach. Nikt z uczestników nie wiedział, na którym ekranie wyświetlany był tekst zapisywany na klawiaturze, a na którym ten tworzony metodą *respeakingu*. W grupie odbiorców tak silnie zakorzeniony był stereotyp, że tworzenie napisów z wykorzystaniem oprogramowania do rozpoznawania mowy jest niemożliwe, że wskazywali na ekran z napisami z *respeakingu*, twierdząc, że na pewno stworzyła je szybko pisząca osoba. Okazało się, że napisy tworzone przez *respeakerów* były dokładniejsze i ukazywały się ze znacznie mniejszym opóźnieniem. Na ekranach oba warianty napisów wyświetlane były blokowo (fragmentami tekstu, a nie wyraz po wyrazie) w wysokim kontraście (żółte litery na czarnym tle), drukowanymi literami, fontem bezszeryfowym. Ze względów technicznych nie była stosowana interpunkcja.

²⁸ Film *Dostępne wydarzenia*, Monika Szczygielska, Materiał przygotowany na spotkanie Rady ds. Organizacji Pozarządowych w Ministerstwie Kultury i Dziedzictwa Narodowego, Październik 2013, <https://youtu.be/Y6cclCx27rc>, dostęp dn. 20 listopada 2018.

²⁹ https://pl.wikipedia.org/wiki/Rozpoznawanie_mowy, dostęp dn. 18 listopada 2018.

Ekran z napisami umieszczone były obok ekranów z prezentacjami, w pobliżu tłumacza migowego, w tej samej części sali, w której rozłożona była pętla indukcyjna.



Rys. 2. Pierwsze testy napisów na żywo podczas konferencji w Sejmie RP w 2013 roku, źródło: Dostępni.eu/ M. Pawluczuk

Tekst tworzony metodą respeakingu powstał z wykorzystaniem jednej z pierwszych wersji programu Magic Scribe. Producent zaznaczał wówczas, że program został stworzony specjalnie dla języka polskiego, w oparciu o japońskie biblioteki danych³⁰. Podkreślał, że jakość rozpoznawania mowy zależy od treningu głosu, sposobu mówienia użytkownika oraz od zaimplementowanego modelu językowego i słownika. Deklarował skuteczność rozpoznawania mowy dla terminologii ogólnej na poziomie około 90% (97% medycyna, 93% prawo). System nie rozpoznawał: imion, nazwisk, adresów zamieszkania, nazw instytucji. Wśród najczęściej pojawiających się błędów wymieniał: odmianę wyrazów, zjadanie łączników typu: „a”, „i”, „z” itd., wstawianie niewypowiedzianych głosek (interpretowanie szumów w otoczeniu jako fonemów wymagających rozpoznania).

Pierwsi polscy respeakerzy byli aktorami. Przeszli podstawowy kurs dyktowania i stworzono im profile głosowe. Pracowali w wyizolowanym akustycznie pomieszczeniu, gdzie doprowadzony został dźwięk i podgląd z konferencji, na sprzęcie dostarczonym przez producenta oprogramowania.

³⁰ Oferta firmy Unikkon Integral z 2013 r.

Podczas tego wydarzenia wśród odbiorców przeprowadzona została krótka ankieta dotycząca oceny napisów. W ankiecie uczestniczyło 16 osób, po połowie kobiet i mężczyzn, w zróżnicowanym wieku (7 osób w wieku powyżej 60 lat, 9 osób w wieku 20–32 lat). W grupie tej byli zarówno uczestnicy konferencji, jak i osoby oglądające transmisję online z napisami. Zdecydowana większość ankietowanych nigdy wcześniej nie uczestniczyła w żadnym wydarzeniu, podczas którego były zastosowane napisy na żywo. Odbiorcy odebrali pierwszy test napisów pozytywnie. Do poprawy na przyszłość wskazali następujące elementy, które przeszkadzały im w odbiorze napisów: brak interpunkcji, nierówne tempo wyświetlania się napisów, literówki, pomijanie niektórych zdań i słów oraz ustawienie ekranów – światło ze świetlówek spod sufitu padało na monitory i utrudniało czytanie. W krótkim czasie po wydarzeniu odbyło się spotkanie konsultacyjne z osobami, które badają i tworzą napisy. Organizatorzy napisów na żywo zapoznali się z aktualnymi badaniami na temat percepcji napisów w Polsce [Szarkowska 2010; 2011] i napisów na żywo w Anglii [Romero Fresco 2013] oraz odbyli wizytę studyjną w Red Bee Media, firmie tworzącej napisy na żywo w Wielkiej Brytanii. Ta wiedza i doświadczenia zostały wykorzystane podczas kolejnych wydarzeń.

Drugie konferencyjne zastosowanie napisów na żywo miało miejsce 5 września 2013 r., w rocznicę podpisania przez Polskę Konwencji o prawach osób niepełnosprawnych w Pałacu Prezydenckim, podczas konferencji „Prawa i bariery. Edukacja, informacja, praca, przestrzeń publiczna a problemy osób niepełnosprawnych”. Tak jak za pierwszym razem, napisy zostały przygotowane z wykorzystaniem oprogramowania Magic Scribe, tym razem z udziałem *respeaker*a i moderatora, którego zadaniem była edycja tekstu. Po raz pierwszy zastosowana została interpunkcja, małe i wielkie litery, a system został poszerzony o opcję korekty. *Respeaker*zy przeszli dodatkowe szkolenie. Zmieniony został też sposób wyświetlania napisów. Utrzymano wysoki kontrast oraz blokowe wyświetlanie napisów. Dodatkowo wprowadzono dwa kolory napisów. Najnowszy fragment tekstu był wyświetlany w kolorze żółtym, a reszta tekstu w kolorze białym. Ekran z napisami stał obok ekranu z tłumaczem języka migowego i prezentacjami. Podczas tego wydarzenia problemem

okazało się przesyłanie napisów lokalną siecią wewnętrzną. Aby umożliwić realizację usługi, konieczne było stworzenie własnej sieci bezprzewodowej, co wymagało ustawienia stanowiska respeakingu w bezpośrednim sąsiedztwie sali konferencyjnej, a w konsekwencji wymusiło na respeakerach pracę w niesprzyjających warunkach akustycznych. Doświadczenia te przyczyniły się do rozpoczęcia prac nad standardem obsługi wydarzeń na żywo.



Rys. 3. Tekst tworzony metodą respeakingu podczas konferencji w Pałacu Prezydenckim w 2013 r., źródło: Dostępni.eu

Kolejne wydarzenie to msza kanonizacyjna Jana Pawła II 27 kwietnia 2014 r.³¹. Transmisja realizowana była we współpracy z Watykańskim Ośrodkiem Telewizyjnym i Międzynarodowym Stowarzyszeniem Respeakerów. Było to pierwsze w Polsce zastosowanie napisów tworzonych metodą respeakingu na odległość i po raz pierwszy z przekładu z języków obcych.

Dwa poprzednie testy wymagały obecności respeakerów w miejscu wydarzenia. Twórcy napisów pracowali w pomieszczeniu, do którego

³¹ Film *Respeaking podczas mszy kanonizacyjnej*, Dostępni.eu, maj 2015, <https://youtu.be/xAljPi-7rM-Y>, dostęp dn. 20 listopada 2018.

doprowadzony był obraz i dźwięk, a napisy wyświetlane były z wykorzystaniem sieci wewnętrznej. Tym razem wydarzenie odbywało się na Placu Św. Piotra, a zespół pracował pod Warszawą. Inne podmioty odpowiadały za obraz, napisy i ich wyświetlanie. Wszystko to możliwe było dzięki połączeniu technologii telewizyjnych i internetu.

Podczas transmisji mszy kanonizacyjnej powstawały dwa typy napisów na żywo: moderowane napisy na podstawie uprzednio przygotowanego tekstu oraz – podczas wystąpień papieskich – z przekładu mowy na tekst metodą respeakingu.

Przy transmisji w Polsce pracowało dwóch respeakerów, z których jeden dodatkowo pełnił rolę moderatora i korektora, oraz tłumacz symultaniczny. Tłumacz symultanicznie tłumaczył homilię i podziękowania papieskie, w pozostałych częściach pomagał w synchronizacji treści z obrazem. Alternatywnie zespół przygotowany był na tłumaczenie z respeakingu angielskiego lub – w razie utraty odsłuchu z transmisji telewizyjnej – w języku polskim.

Napisy przesyłane były na stronę internetową, z której były pobierane i wyświetlane pod playerem tv. Za emisję napisów odpowiadała strona włoska. Transmisja internetowa była opóźniona o 30 sekund wobec sygnału telewizyjnego Watykańskiego Ośrodka Telewizyjnego, nadawanego we Włoszech. Msza kanonizacyjna została wyemitowana z napisami w językach: włoskim, hiszpańskim, angielskim i polskim. Przedsięwzięcie realizowane było w międzynarodowym gronie, pracowało nad nim 50 osób z Polski, Włoch, Hiszpanii i USA. Polskie napisy charakteryzowała wysoka jakość i dokładność. Tylko polscy respeakerzy stosowali identyfikację osób oraz opisy dźwięków dla osób niesłyszących.

Duże znaczenie dla powodzenia przedsięwzięcia miały wcześniejsze przygotowania. Oficjalny tekst liturgii, bez homilii i wystąpienia papieża, dostępny był wcześniej po łacinie, włosku i angielsku, w tym część hymnów jedynie w języku łacińskim. Przygotowanie do transmisji wymagało opracowania polskiego tekstu liturgii. Zespół zdecydował o przetłumaczeniu liturgii wraz z hymnami według oficjalnie obowiązujących ksiąg liturgicznych. Przygotowanie merytoryczne zapewnili współpracujący księża. Pomogło to w przygotowaniu słownika, który przed wydarzeniem należało wprowadzić do programu do rozpoznawania

mowy. Obejmował on m.in. nazwiska, tytuły pieśni, słownictwo charakterystyczne dla liturgii.

Mimo bardzo dobrego przygotowania zespołu polskiego odbiór napisów został zakłócony przez problem z wyświetlaniem polskich znaków. Za wyświetlanie tekstu odpowiadała strona włoska. Mimo wielu testów nie udało im się poprawnie wdrożyć polskich znaków diakrytycznych. Włoscy partnerzy byli przekonani, że udało im się rozwiązać problem i tak to zakomunikowali polskiemu zespołowi. Faktycznie program wyświetlał polskie znaki, ale w sposób przypadkowy, np. „ą” wyświetlane było jako „ż”, co uniemożliwiało komfortowe czytanie tekstu. Dlatego w pierwszych minutach transmisji polski zespół zdecydował się na usunięcie wszystkich polskich znaków.

5.2. Wykorzystanie doświadczeń zagranicznych

Od 2013 r. nasiliły się kontakty z praktykami i badaczami napisów na żywo z zagranicy, przede wszystkim dzięki działaniom badawczo-rozwojowym Dostepni.eu oraz badaniom prowadzonym na Uniwersytecie Warszawskim.

Zespół Dostepni.eu nawiązał kontakty z jednym największych dostawców usług dostępowych na świecie — brytyjską firmą Red Bee Media (która początkowo była częścią telewizji BBC, by następnie się usamodzielić, a w 2014 r. została przejęta przez koncern Ericsson). Firma ta obsługuje m.in. takich nadawców jak BBC, ITV, BSkyB, TV4, TV5 Monde, Canal+ i Channel 4 oraz większość usług wideo na żądanie (VOD) w Wielkiej Brytanii, w tym BBC (BBC iPlayer) i Channel 4 (4oD).

W ramach dwóch wizyt studyjnych w Red Bee Media (2013, 2014) realizatorzy napisów na żywo mieli okazję zapoznać się z metodami tworzenia napisów na żywo, wykorzystywanym oprogramowaniem, sposobem pracy respeakerów i warunkami, w jakich ją wykonują. Dyskutowali także o procesie rekrutacji i szkolenia respeakerów, roli słownika, typologiach błędów rozpoznawania i strategiach radzenia sobie z nimi.

Odwrotnie niż w Wielkiej Brytanii, w Polsce napisy na żywo rozwinęły się najpierw w kontekście wydarzeń na żywo, a dopiero potem znalazły zastosowanie w telewizji. Jako że pierwsze transmisje online z napisami na żywo z międzynarodowych wydarzeń w ONZ oraz Komisji Europejskiej miały miejsce w tym samym czasie, co w Polsce³², w tamtym okresie brakowało zagranicznych rozwiązań i praktyk, na których można by było się wzorować. Z tego względu, choć Red Bee Media świadczyła wówczas usługi dla telewizji, a nie podczas wydarzeń na żywo, możliwość obserwacji, w jaki sposób pracuje największy na świecie dostawca telewizyjnych usług dostępowych, pomogła w wypracowywaniu standardu obsługi wydarzeń na żywo w Polsce. Zespół miał także możliwość obserwować transmisje na żywo z napisami i tłumaczeniem migowym z niemieckiego parlamentu, realizowane przez Verba Voice [Szczygielska 2016: 52–53]. W Anglii *respeakerzy* telewizyjni sami starają się poprawiać błędy w napisach, co jest związane przede wszystkim ze stosunkową niską stopą błędów w rozpoznawaniu mowy w języku angielskim oraz dodatkową konsolą ułatwiającą wprowadzanie znaków interpunkcyjnych. Już po pierwszej wizycie w Londynie i konsultacjach z fachowcami w Polsce i Wielkiej Brytanii zapadła decyzja o tworzeniu w Polsce napisów w tandemie *respeaker* – moderator/korektor, co było związane z wyższą stopą błędów w rozpoznawaniu mowy w języku polskim.

Rozpoczęty w 2014 r. na Uniwersytecie Warszawskim projekt badawczy „*Respeaking – proces, kompetencje, jakość*”, kierowany przez dr hab. Agnieszkę Szarkowską, dostarczył ważnej wiedzy dotyczącej kompetencji *respeakerów*. Towarzyszył mu także cykl warsztatów przeprowadzonych w 2015 r. w Instytucie Lingwistyki Stosowanej Uniwersytetu Warszawskiego (ILS UW), które prowadzili czołowi specjaliści w zakresie *respeakingu*: dr Carlo Eugeni (Uniwersytet w Parmie/Międzynarodowe Stowarzyszenie *Respeakerów Respeaking on Air*), dr Pablo Romero Fresco (Uniwersytet w Roehampton) oraz Juan Martínez Perez (Swiss TXT,

³² Szczygielska M., Dutka Ł. (2016) *Live subtitling in online streaming in Poland: between theory and practice*, prezentacja z konferencji Intermedia, 16 kwietnia 2016 r., http://avt.ils.uw.edu.pl/files/2016/04/Live-subtitling-in-online-streaming_SZCZYGIELSKA-Monika-DUTKA-%C5%81ukasz-Stanis%C5%82aw.pdf, dostęp dn. 20 listopada 2018.

dostawca usług dostępowych dla szwajcarskiej telewizji publicznej), twórca modelu NER, stworzonego w celu badania jakości napisów telewizyjnych w Wielkiej Brytanii. Warsztaty umożliwiły dotychczasowym respeakerom podniesienie swoich kompetencji, a także pozwoliły na wyszkolenie nowych specjalistów. Była to też cenna okazja do wymiany doświadczeń w zakresie kształcenia respeakerów, która zaowocowała później uruchomieniem pierwszego polskiego kursu uniwersyteckiego w zakresie napisów na żywo w ILS UW w roku akademickim 2017/2018.

6. Rozwój respeakingu w Polsce

6.1. Napisy w transmisji online

Pierwsze transmisje online z napisami na żywo odbywały się na stronie internetowej www.sejm.gov.pl. Napisy wyświetlane były w ramce pod ekranem z filmem wideo, dzięki czemu nie zasłaniały ekranu. Ramkę z napisami można było włączyć lub wyłączyć. Nie wiadomo, czy poprawnie wyświetlała się na wszystkich urządzeniach mobilnych. Odtwarzacz wideo na stronie Sejmu był dostępny dla osób niewidomych i słabowidzących (dzięki kompatybilności z oprogramowaniem czytającym).

Napisy wyświetlane były na czarnym tle. Aby tekst był czytelny, zastosowano biały bezszeryfowy krój pisma o dużym rozmiarze. Napisy pojawiały się blokowo, po 3 wiersze – w każdym do 40 znaków. Tekst przewijał się do góry. Synchronizacja napisów z wideo odbywała się za pomocą systemu znaczników czasowych, opracowanych przez dostawcę oprogramowania do rozpoznawania mowy. Pozwalało to na żywo na zmniejszenie opóźnienia, a w transmisji archiwalnej na jego wyeliminowanie. Oprogramowanie to nie dawało jednak technicznej możliwości podziału tekstu na wiersze zgodnie z zasadami tworzenia napisów dla niesłyszących [Neves 2005; Künstler, Butkiewicz 2012].

W napisach z 2014 r.³³ zauważalne są brakujące znaki przestankowe.

iTV Sejm



10:46:05

2 IX 2014, godz. 10:06 - 14:26

Odtwórz Pauza Stop Pełny ekran Do tyłu 20% Do przodu 20% Ciszej Głośniejsze Wycisz

Pokaż tekst

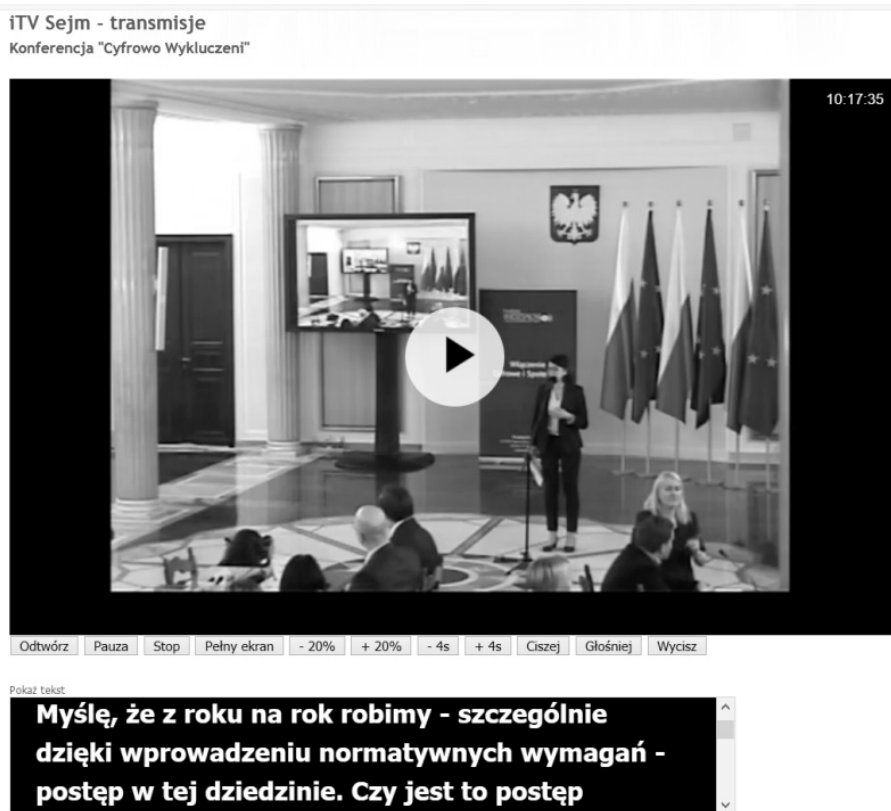
**dawno temu dziś jest to dla nich naprawdę tanie .
Za chwilę spróbujemy połączyć się z telewizją
czeską Spróbujemy jeszcze Słyszemy bardzo**

"Telewizja dostępna dla wszystkich"
Sala Kolumnowa/bud. C
Forum Dostępnej Cyberprzestrzeni

Rys. 4. Napisy na żywo z konferencji w Sejmie RP w 2014 r., źródło: iTV Sejm

³³ Archiwalny zapis transmisji na żywo z konferencji „Telewizja dostępna dla wszystkich” Forum Dostępnej Cyberprzestrzeni z 2014 r., http://www.sejm.gov.pl/Sejm7.nsf/transmisje_arch.xsp#72853F3DC3876C41C1257D3F002AF6C3, dostęp dn. 20 listopada 2018.

W napisach z 2016 r.³⁴ ten problem już nie występuje.



Rys. 5. Napisy na żywo z konferencji w Sejmie RP w 2016 r., źródło: iTV Sejm

We współpracy z komercyjnymi nadawcami transmisji online napisy na żywo wyświetlane były jako napisy otwarte, czyli wyświetlane jako złożony z pikseli obraz, nałożony na obraz wideo z kamery. Wynikało to z braku możliwości technicznych zastosowania napisów zamkniętych (przesyłanych i wyświetlanych jako tekst, a nie obraz).

Warto zwrócić uwagę, że napisy otwarte spełniają swoją rolę dla niesłyszących odbiorców napisów oraz osób, które oglądają wideo bez włączonego dźwięku. Nie jest to jednak rozwiązanie idealne, ponieważ napisy te zasłaniają część ekranu, nie da się ich wyłączyć i nie można

³⁴ Archiwalny zapis transmisji na żywo z konferencji „Cyfrowo Wykluczeni” Fundacji Widzialni z 2016 r., <http://www.sejm.gov.pl/Sejm8.nsf/transmisje.xsp?unid=AD468A69C4783B26C1257F9B00228070#>, dostęp dn. 20 listopada 2018.

skorzystać z nich za pomocą urządzeń asystujących wykorzystywanych np. przez osoby niewidome czy głuchoniewidome, ponieważ tak wyświetlane napisy są przesyłane jako część obrazu, oprogramowanie czytające nie jest w stanie ich odczytać. Rozwiązanie to uniemożliwia też pełną synchronizację napisów z obrazem, jest jednak formą pewnego kompromisu między kosztem a jakością. Natomiast dobrą praktyką powinno być wyświetlanie napisów na żywo jako napisów zamkniętych.



Rys. 6. Napisy na żywo w transmisji online ze Wspólnej Inauguracji Roku Akademickiego 2018/2019 na Uniwersytecie im. Adama Mickiewicza w Poznaniu 1 października 2018 r., źródło: Dostępni.eu

6.2. Napisy na żywo w przekładzie z języka obcego

We wrześniu 2014 r. zespół Dostępni.eu tworzył napisy podczas konferencji w Sejmie RP organizowanej przez Forum Dostępnej Cyberprzestrzeni „Telewizja dostępna dla wszystkich”. W konferencji online uczestniczyli prelegenci z Czeskiej Telewizji oraz Red Bee Media, których wystąpienia odbyły się w języku angielskim.

W procesie tworzenia napisów z języka angielskiego uczestniczył tłumacz, respeaker i moderator. Spowodowało to wydłużenie łańcucha o jedno ogniwo, co teoretycznie powinno zwiększać opóźnienie w pojawianiu się tekstu na ekranie. Jednocześnie jednak poprawie uległo

dzielenie tekstu na mniejsze fragmenty, dzięki czemu pojawiające się napisy były lepiej zredagowane i łatwiej przyswajalne.

Możliwe jest także tłumaczenie symultaniczne i jednoczesne dyktowanie tekstu przez tego samego tłumacza-respeakera. Wówczas uczestnicy wydarzenia mogą korzystać z tłumaczenia w formie napisów (zamiast tłumaczenia symultanicznego na słuchawkach). Należy jednak pamiętać, że wśród uczestników mogą być osoby niewidome, które nie będą w stanie odczytać napisów, o ile nie zostaną one udźwiękowione przez syntezytor mowy. W związku z tym, że w wydarzeniu uczestniczyły osoby z różnymi niepełnosprawnościami, zdecydowano na zastosowanie – obok respeakera – tłumacza symultanicznego z języka obcego na polski. Ułatwiło to udział w wydarzeniu osobom niewidomym oraz zapewniło płynność tłumaczenia na język migowy.

Dokładnie taki sam schemat tłumacz-respeaker-moderator zastosowany został 12 maja 2015 r. podczas konferencji „Europejski Rynek Cyfrowy – umiejętności, gospodarka, praca” z udziałem wiceprzewodniczącego Komisji Europejskiej Andrusa Ansipa, organizowanej przez Instytut Spraw Publicznych w Centrum Nauki Kopernik [Szczygielska et al. 2015: 37]. Zagraniczni goście biorący udział w wydarzeniu potrzebowali stałego tłumaczenia z języka polskiego na język angielski.

Nie bez znaczenia jest pewnie, że w obydwu opisanych przypadkach tłumacze symultaniczni współpracują na stałe z zespołem Dostępni.eu i znają specyfikę respeakingu. W sytuacjach, kiedy tłumaczenie symultaniczne zapewnia inny zespół, do decyzji respeakera zależy, czy będzie tłumaczył i dyktował tekst jednocześnie, czy skorzysta z tłumaczenia zewnętrznego. Tak było podczas konferencji inauguracyjnej programu „Dostępność Plus”, organizowanej przez Ministerstwo Inwestycji i Rozwoju 23 kwietnia 2018 r. w Centrum Nauki Kopernik. Aby zmniejszyć opóźnienie, respeaker zdecydował, że będzie jednocześnie tłumaczył i dyktował. Rozwiązanie to było korzystne dla niesłyszących uczestników wydarzenia, ale sprawiało pewien problem w transmisji online. Widzowie przed ekranami nie mieli możliwości wyboru języka. Nadawca zapewniał wyłącznie dźwięk polski, co oznaczało, że podczas wystąpień w języku angielskim widzowie słyszeli tłumaczenie na język

polski. Jeśli jednocześnie mieli włączone napisy z innego tłumaczenia, mogli mieć poczucie dysonansu.



Rys. 7. Komisarz UE Andrus Ansip i polskie napisy na żywo do jego wystąpienia, źródło: Dostępni.eu

6.3. Napisy na żywo w telewizji

W 2017 r. zespół Dostępni.eu rozpoczął współpracę z telewizją Polsat przy programie *Dancing with the Stars. Taniec z Gwiazdami*³⁵. Projekt wymagał zmian bezpośrednio w modelu językowym programu do rozpoznawania mowy. Język *Tańca z Gwiazdami* to mieszanka różnych odmian języka: mowy spontanicznej, słownictwa specjalistycznego związanego z tańcem, wyrazów obcojęzycznych. Był to projekt realizowany w oparciu o oprogramowanie Magic Scribe. Oprogramowanie zostało zintegrowane z programem do tworzenia napisów FAB Live, dzięki czemu praca respeakingera i moderatora odbywała się na jednym stanowisku. Oznacza to, że respeaker dyktował bezpośrednio do edytora w oprogramowaniu FAB Live. Podczas emisji programu testowane były różne sposoby wyświetlania napisów, różne ich pozycje i sposoby oznaczania mówców. Ostatecznie napisy wyświetlane były na górze ekranu (by nie kolidowały

³⁵ Film *Jak powstały napisy na żywo do Tańca z Gwiazdami*, Dostępni.eu, Wrzesień 2017, <https://youtu.be/DDMhQiakhtY>, dostęp dn. 20 listopada 2018.

z elementami graficznymi i nie zasłaniały nóg tancerzy), słowo po słowie (by zminimalizować opóźnienie). Prowadzący oznaczeni zostali kolorami, a członkowie jury otrzymali wizytówki.



Rys. 8. Napisy na żywo w *Tańcu z Gwiazdami*, Polsat 2017, źródło: Dostępni.eu

Telewizja Polska nie podaje do publicznej wiadomości informacji o ilości napisów na żywo w ramówce³⁶ ani o metodach ich tworzenia³⁷. Zgodnie z udostępnianym przez stację programem telewizyjnym³⁸ (stan na listopad 2018 r.) z napisami ukazują się następujące programy nadawane na żywo: Teleexpress, Teleexpress Extra, Wiadomości, Gość Wiadomości oraz Sport, do których część napisów jest potencjalnie tworzona w czasie rzeczywistym. TVP najprawdopodobniej testowała respeaking podczas

³⁶ Zbigniew Kunecki, *Realizacja usług dostępowych w telewizji publicznej w Polsce*, <https://www.google.com/url?q=http://www.fdc.org.pl/gallery/Zbigniew-Kunecki-Telewizja-Polska-Realizacja-us%25C5%2582ug-dost%25C4%2599powych-w-telewizji-publicznej-w-Polsce-.pdf&sa=D&ust=1543585425318000&usg=AFQjCNE06Kh8agp8quRzr0PyDIOALDcYRA>, dostęp dn. 20 listopada 2018.

³⁷ Bezpośrednia komunikacja z p. Zbigniewem Kuneckim, specjalistą ds. dostępności w TVP SA.

³⁸ <https://www.tvp.pl/program-telewizyjny>, dostęp dn. 24 listopada 2018.

Światowych Dni Młodzi w 2016 r., a także wieczoru wyborczego w 2018 r³⁹.

7. Model respeakingu stosowany w Polsce

Na podstawie wyżej opisanych doświadczeń wypracowany został model respeakingu stosowany w Polsce, opierający się na współpracy respeakera i moderatora. Podobnie jak tłumacze symultaniczni, respeakerzy i moderatorzy pracują w kabinach tłumaczeniowych lub miejscach akustycznie wydzielonych i zmieniają się w trakcie pracy nie rzadziej niż co 30 minut.

7.1. Respeaking

Zadaniem respeakera jest powtarzanie po mówcy tekstu do mikrofonu służącego do rozpoznawania mowy. Respeaker dodaje znaki interpunkcyjne, oznaczenia mówców oraz dzieli tekst na mniejsze fragmenty. W zespole Dostępni.eu każdy z respeakerów pracuje na wcześniej przygotowanym profilu głosowym. Profil głosowy zapamiętuje indywidualne parametry głosu respeakera. Buduje się go jednorazowo poprzez dyktowanie różnych rodzajów tekstu w ilości zależnej od producenta oprogramowania. Wraz z rozwojem oprogramowania do rozpoznawania mowy i zastosowaniem sieci neuronowych, budowanie profili głosowych może przestać być konieczne. Przed każdym wydarzeniem respeaker testuje słownik. Sprawdza, czy wszystkie nowo wprowadzone wyrazy i zwroty prawidłowo się rozpoznają. W trakcie pracy kontroluje proces rozpoznawania mowy. Jeśli zauważy, że jakiś wyraz konsekwentnie źle się rozpoznaje, używa jego synonimu. W zależności od konfiguracji oprogramowania respeaker może pracować na tym samym lub odrębnym stanowisku co moderator. Minimalne wyposażenie stanowiska respeakera to mikrofon do rozpoznawania mowy, słuchawki i komputer z oprogramowaniem do rozpoznawania mowy.

³⁹ Współpracujący z TVP SA redaktorzy napisów dla niesłyszących, z którymi skontaktowali się autorzy artykułu, powołali się na umowy o zachowaniu poufności, które zabraniają im udostępniania informacji dotyczących ich pracy.

7.2. Moderacja

Zadaniem moderatora jest korekta tekstu powstałego w wyniku rozpoznawania mowy. Usuwa niewłaściwie rozpoznane słowa, poprawia błędy powstałe w wyniku działania oprogramowania i ewentualne pomyłki respeakera. Odbywa się to w edytorze, do którego trafia rozpoznany tekst. Po korekcie tekst wysyłany jest na ekran, do transmisji telewizyjnej lub internetowej. Podstawowe wyposażenie stanowiska moderatora to komputer z oprogramowaniem do edycji tekstu i tworzenia napisów, słuchawki oraz klawiatura. Jeśli respeaker i moderator pracują na dwóch stanowiskach komputerowych, komputery te połączone są w sieć wewnętrzną. Praca respeakera i moderatora na jednym stanowisku komputerowym przyspiesza proces, ale wymaga integracji oprogramowania rozpoznawania mowy z programem do tworzenia napisów na żywo.

8. Wyzwania na przyszłość

Dalszy rozwój napisów na żywo jako usługi dostępowej w Polsce wiąże się z szeregiem wyzwań. Potrzebne jest doskonalenie istniejących i tworzenie nowych, kompleksowych rozwiązań technicznych, przeprowadzenie pogłębionych badań naukowych, wypracowanie i upowszechnienie standardów, wyszkolenie większej liczby specjalistów, zwiększanie świadomości decydentów, organizatorów wydarzeń i potencjalnych odbiorców usługi.

Proces upowszechniania się napisów na żywo mogą przyspieszyć zmiany prawne. Projekt Ustawy o dostępności⁴⁰ zakłada, że w 2028 r. co najmniej 95% programów będzie miało napisy dla niesłyszących i audiodeskrypcję, co wymusi na nadawcach udostępnianie z napisami także programów na żywo i w konsekwencji zwiększy ilość tworzonych napisów na żywo i zapotrzebowania rynkowe na nowy typ specjalistów – tłumaczy-respeakerów.

By odpowiedzieć na to zapotrzebowanie, konieczna staje się ściślejsza współpraca środowiska akademickiego z twórcami treści

⁴⁰ Art. 22 projekt Ustawy o dostępności z dnia 23 listopada 2018 r., <https://legislacja.rcl.gov.pl/projekt/12319302>, dostęp dn. 25 stycznia 2019.

audiowizualnych i dostawcami usług dostępowych, a także organizatorami wydarzeń. Pozytywnym przykładem jest kurs napisów na żywo, włączony w 2017 r. do oferty dydaktycznej ILS UW w ramach kierunku lingwistyka stosowana. Kurs pozwala zdobyć kompetencje z zakresu respeakingu adeptom przekładu ustnego i audiowizualnego, dzięki czemu powoli zwiększa się liczba specjalistów gotowych do pracy jako respeakerzy w języku polskim. Wyzwaniem pozostaje włączenie respeakingu w programy kształcenia tłumaczy w Polsce na szerszą skalę, a także stworzenie kursu kształcenia respeakerów międzyjęzykowych.

W Unii Europejskiej podejmowane są też pierwsze próby normalizacji napisów na żywo. W 2018 r. Międzynarodowy Związek Telekomunikacyjny (International Telecommunication Union – ITU) opublikował pierwszy dokument techniczny w tej sprawie⁴¹. Aktualnie badacze z Uniwersytetu Warszawskiego wraz z zespołem Dostępni.eu pracują nad wytycznymi dla organizatorów i nadawców transmisji online z napisami na żywo⁴². W latach 2018–2020 trwa międzynarodowy projekt ILSA (Interlingual Live Subtitling Access), który ma na celu opracowanie standardów świadczenia usług napisów na żywo, a także stworzenie materiałów dydaktycznych do kształcenia respeakerów międzyjęzykowych. Powstanie kurs tworzenia napisów na żywo w trzech kontekstach: w telewizji, podczas wydarzeń społecznych i politycznych oraz na potrzeby edukacji. Jednocześnie podczas innego projektu międzynarodowego, LTA (Live Text Access), w latach 2018–2021, mają zostać wypracowane zasady certyfikacji respeakerów. Dużym wyzwaniem w Polsce będzie wpisanie kwalifikacji respeakera do Zintegrowanego Rejestru Kwalifikacji⁴³.

Wraz ze wzrostem liczby kształconych respeakerów i upowszechnieniem się napisów na żywo, coraz bardziej paląca staje się potrzeba prowadzenia badań nad jakością napisów, tak by zwiększona oferta napisów na żywo realnie przyczyniła się do zwiększenia dostępności mediów i wydarzeń na żywo dla wszystkich odbiorców.

⁴¹ *Overview of remote captioning services. ITU-T Technical Paper*, ITU 2018.

⁴² Szczygielska M., Dutka Ł., *Proposal of guidelines on universally accessible online streaming with live subtitles*, prezentacja wygłoszona podczas konferencji 6th International Symposium on Accessibility and Live Subtitling, Mediolan 14 września 2018, http://www.respeaking.net/milan%202018/abstract_monikaszczygielska_lukaszdutka.htm, dostęp dn. 30 listopada 2018.

⁴³ <https://rejestr.kwalifikacje.gov.pl/>, dostęp dn. 20 listopada 2018.

Podziękowania

Autorzy artykułu dziękują dr hab. Agnieszce Szarkowskiej za wnikliwą lekturę tekstu i cenne uwagi.

BIBLIOGRAFIA

- Arumí Ribas, M., Romero Fresco, P. (2008), „A Practical Proposal for the Training of Respeakers”. *The Journal of Specialised Translation*. 10 s. 106–127.
- Benesty, J., Sondhi, M. M., Huang, Y. (2007), *Springer handbook of speech processing*. Springer Science & Business Media, Berlin/Heidelberg.
- Brocki, Ł., Marasek, K., Koržinek, D. (2012), „Multiple model text normalization for the Polish language”, [w:] L. Chen, A. Felfernig, J. Liu, Z. W. Raś, red. *Foundations of Intelligent Systems*, Springer, Berlin/Heidelberg, s. 143–148.
- den Boer, C. (2001), „Live interlingual subtitling”, [w:] Y. Gambier, H. Gottlieb, red. *(Multi)media translation: concepts, practices and research*, Benjamins, Amsterdam/Philadelphia, s. 167–172.
- De Korte, T. (2006), „Live Inter-Lingual Subtitling in the Netherlands. Historical Background and Current Practice”. *Intralinea*. http://www.intralinea.org/specials/article/Live_inter-lingual_subtitling_in_the_Netherlands, dostęp dn. 20 listopada 2018.
- Díaz Cintas, J. (2005), „Accessibility for All”, *Translating Today*. 4, s. 3–5.
- Díaz Cintas, J., Remael, A. (2007), *Audiovisual translation: subtitling*, St Jerome, Manchester.
- Downey, J. G. (2008), *Closed Captioning: Subtitling, Stenography, and the Digital Convergence of Text with Television*, John Hopkins University Press, Baltimore.

- Dumouchel, P., Boulianne, G., Brousseau, J. (2011), „Measures for quality of closed captioning”, [w:] A. Şerban, A. Matamala, J.-M. Lavaur, red. *Audiovisual translation in close-up: Practical and theoretical approaches*, Peter Lang, Bern, s. 161–172.
- Eugeni, C. (2007), „Il rispeakeraggio televisivo per sordi. Per una sottotitolazione mirata del TG”, *inTRAlinea*. 9 http://www.intralinea.it/volumes/eng_open.php?id=P513, dostęp dn. 20 listopada 2018.
- Eugeni, C. (2008), „Respeaking the TV for the Deaf: For a Real Special Needs-Oriented Subtitling”, *Studies in English Language and Literature*. 21, s. 37–47.
- Eugeni, C. (2009), „Respeaking the BBC News. A strategic analysis of Respeaking on the BBC”, *The Sign Language Translator and Interpreter*. 3(1), s. 29–68.
- Jankowska, A. (w druku), „Audiovisual media accessibility”, [w:] E. Angellone, M. Ehrensberger-Dow, G. Massey, red. *The Bloomsbury companion to language industry studies*, Bloomsbury Academic Publishing, London.
- Künstler, I., Butkiewicz, U. (2012), *Napisy dla osób niesłyszących i słabosłyszących – zasady tworzenia*, Fundacja Kultury bez Barrier.
- Lambourne, A. (2006), „Subtitle Respeaking. A New Skill for A New Age”, *Intralinea*, http://www.intralinea.org/specials/article/Subtitle_respeaking, dostęp dn. 20 listopada 2018.
- Neves, J. (2005), *Audiovisual Translation: Subtitling for the Deaf and Hard-of-Hearing*. Nieopublikowana rozprawa doktorska. Roehampton University, <http://roehampton.openrepository.com/roehampton/handle/10142/12580>, dostęp dn. 20 listopada 2018.
- Neves, J. (2009), „Interlingual subtitling for deaf and hard-of-hearing”, [w:] *Audiovisual translation. Language transfer on screen*, Palgrave Macmillan UK, London, s. 151–169.
- Remael, A., Van Der Veer, B. (2006), „Real-Time Subtitling in Flanders: Needs and Teaching”, *Intralinea, Special Issue: Respeaking*, <http://www.intralinea.org/specials/respeaking>, dostęp dn. 20 listopada 2018..

- Robert, I., Remael, A. (2017), „Assessing Quality in Live Interlingual Subtitling: A New Challenge”, *Linguistica Antverpiensia New Series: Themes in Translation Studies*. 14, s. 168–195.
- Robson, D. G. (2004), *The Closed Captioning Handbook*, Focal Press, Waltham.
- Romero Fresco, P. (2011), *Subtitling through Speech Recognition: Respeaking*, St Jerome, Manchester.
- Romero Fresco, P. (2012), „Respeaking in Translator Training Curricula. Present and Future Prospects”, *The Interpreter and Translator Trainer*. 6(1), s. 91–112.
- Romero Fresco, P. (2018), „Reception studies in live and pre-recorded subtitles for the deaf and hard of hearing”, [w:] E. Di Giovanni, Y. Gambier, red. *Reception Studies and Audiovisual Translation*, Benjamins, Amsterdam.
- Romero Fresco, P., Pöchhacker, F. (2017), „Quality assessment in interlingual live subtitling: The NTR Model”, *Linguistica Antverpiensia, New Series – Themes in Translation Studies*. 16, s. 149–167.
- Szarkowska, A. (2008), „Przekład audiowizualny w Polsce – perspektywy i wyzwania”, *Przekładaniec*. 20, s. 8–25.
- Szarkowska, A., Krejtz, K., Dutka, Ł., Pilipczuk, O. (2016), „Cognitive load in intralingual and interlingual respeaking – a preliminary study”, *Poznań Studies in Contemporary Linguistics*. 52(2), s. 209–233.
- Szczygielska, M. (2016), *Dostępne multimedia*, <http://www.widzialni.org/container/Dostepne-multimedia.pdf>, dostęp dn. 20 listopada 2018.
- Szczygielska, M., Trzeciakiewicz, M., Kasperkowiak, M. (2015), *Dostępne wydarzenia w praktyce*, <http://www.fdc.org.pl/gallery/DOST%C4%98PNE-WYDARZENIA-W-PRAKTYCE.pdf>, dostęp dn. 20 listopada 2018.
- Ziółko, M., Ziółko, B. (2011), *Przetwarzanie mowy*. Wydawnictwo AGH, Kraków.

STRESZCZENIE

Respeaking jest techniką tworzenia napisów na żywo z wykorzystaniem oprogramowania do rozpoznawania i przekładu mowy na tekst. Napisy tworzone tą metodą stosuje się np. podczas konferencji lub w programach telewizyjnych nadawanych na żywo, a ich odbiorcami są m.in. osoby niesłyszące. Premiera respeakingu w Polsce miała miejsce w 2013 r. podczas konferencji w Sejmie RP. W latach 2013–2018 respeaking wykorzystywano także podczas wydarzeń rządowych, samorządowych i akademickich. Podczas mszy kanonizacyjnej Jana Pawła II w 2014 r. po raz pierwszy tworzone były napisy na odległość oraz z przekładu z języków obcych. Od 2014 r. respeaking stosowany jest w napisach nadawanych na żywo w transmisji online, a od 2016 r. w telewizji. Artykuł opisuje metodę respeakingu i związane z nią aspekty techniczne komputerowego rozpoznawania mowy oraz omawia oprogramowanie dostępne w języku polskim. Przedstawiona została historia respeakingu w Polsce, aktualna dostępność napisów na żywo w telewizji i podczas wydarzeń oraz scharakteryzowany został model świadczenia usługi napisów na żywo funkcjonujący w Polsce. Artykuł omawia też wyzwania związane z rozwojem napisów na żywo w języku polskim.

Słowa kluczowe: Przekład audiowizualny, napisy, napisy dla niesłyszących, napisy na żywo, dostępność.

SUMMARY

The History of Live Subtitling with Respeaking in Poland

Respeaking is a technique of producing live subtitles by means of speech recognition software. Live subtitles are used during conferences or on live TV shows, and their recipients are, among others, the deaf and the hard of hearing. Respeaking was first used in Poland in 2013 during a conference

at the Polish Parliament. Since then respeaking has also been used during government, local government and academic events. During the canonization mass of John Paul II in 2014, for the first time in Poland, respeaking was done remotely and interlingually. Since 2014, respeaking has been used to produce subtitles for live online streaming, and since 2016, live subtitles through respeaking have been available on Polish television. The article describes respeaking, related technological aspects of speech recognition and discusses software available in Polish. The article presents the history of respeaking in Poland, the current availability of live subtitles on television and during live events and the Polish model of providing live subtitling service. The final section discusses challenges for further development of live subtitling in Polish.

Keywords: Audiovisual translation, subtitles, SDH, live subtitles, accessibility.