

ANNA MAJEWSKA-TWOREK

ZASTOSOWANIE NOWYCH TECHNOLOGII
W ANALIZIE NIEPŁYNNOCI WYPOWIEDZI
NAJSTARSZEGO POKOLENIA UŻYTKOWNIKÓW POLSZCZYZNY
(NA PRZYKŁADZIE AUTOKOREKT)

WPROWADZENIE

Niewątpliwie rozwój technologii cyfrowych otworzył nowe, interesujące perspektywy przed badaczami mówionej odmiany języka. Wcześniejszy brak narzędzi do utrwalania i analizy wypowiedzi mówionych, przetwarzania ich oraz archiwizowania hamował badania tej odmiany języka. W przeszłości mówiona odmiana polszczyzny często więc była opisywana przez kontrast do odmiany pisanej oraz zastosowanie norm odnoszących się do lepiej zbadanej pisanej odmiany języka. Niniejszy tekst ma za zadanie wskazać możliwości analizy pewnego aspektu wypowiedzi mówionej z wykorzystaniem nowych technologii. Przedmiotem zainteresowania są symptomy wahania mówcy z psycholingwistycznej perspektywy wskazującej na wieloetapowość powstawania wypowiedzi ustnej. Przeprowadzona w tekście analiza autokorekt stanowi wstęp do badań nad przebiegiem mówienia osób w okresie późnej dorosłości.

WPLYW NOWYCH TECHNOLOGII
NA ROZWÓJ BADAŃ POLSZCZYZNY MÓWIONEJ

Początki badań mówionej odmiany polszczyzny mają swoje źródła w tradycyjnej dialektologii oraz wczesnej socjolingwistyce podejmującej badania

Dr hab. ANNA MAJEWSKA-TWOREK, prof. UW_r – Uniwersytet Wrocławski, Wydział Filologiczny, Instytut Filologii Polskiej, Zakład Językoznawstwa Stosowanego, pl. Biskupa Nankiera 15b, 50-140 Wrocław; e-mail: anna.majewska-tworek@uwr.edu.pl; ORCID: <https://orcid.org/0000-0003-3997-5991>.

Artykuły są objęte licencją Creative Commons Uznanie autorstwa – Użycie niekomercyjne – Bez utworów zależnych 4.0 Międzynarodowe (CC BY-NC-ND 4.0)

nad polszczyzną mówioną miast na przełomie lat siedemdziesiątych i osiemdziesiątych XX wieku. Duże taśmowe magnetofony zabierane na nagrania terenowe odeszły w przeszłość, a wraz z nimi – tradycyjne kasety z interesującymi nagraniami pochodzącymi z najstarszych, wartościowych projektów. Te ocalałe oczekują na cyfryzację, ponowne opracowanie oraz zarchiwizowanie zgodnie ze współczesnymi standardami (Majewska-Tworek i in., 2020). Nie wszystko można uratować, gdy brakuje pełnych informacji o mówcach – uczestnikach badań, a jakość dźwięku ze starych nośników jest trudna do skorygowania mimo starań współczesnych informatyków.

Prawdopodobnie z technicznych ograniczeń wynikało też wykorzystywanie bardzo różnych źródeł danych mówionych w ramach jednego projektu. W dobie rozwoju nowych technologii i lingwistyki korpusowej niejednolity sposób gromadzenia danych jest odrzucony. Dane do badania mówionej odmiany języka mogły bowiem stanowić i przypadkowo zasłyszane wypowiedzi (notowane na fiszkach, które badacz zawsze miał przy sobie), i fragmenty wypisane z dialogów współczesnych powieści lub nawet przykłady samodzielnie przez badaczy wymyślone jako możliwe do wystąpienia (Majewska-Tworek, Zaśko-Zielińska, 2020b, s. 182). Lingwistyka korpusowa oraz programy do anotacji mowy zmieniły więc badania mówionej odmiany języka. Zbieranie danych można zaplanować według starannie przemyślanego klucza, podporządkowanego danemu projektowi, z troską o pełną archiwizację ze stosownymi metadanymi (Zaśko-Zielińska i in., 2020). W pierwszym etapie badań ważna jest więc zarówno koncepcja korpusu danych mówionych, który jest współcześnie rozumiany jako zbiór nagrań mowy wraz z zestawem informacji: transkrypcjami, anotacjami i metadanymi, często również z podstawowymi analizami statystycznymi danych (Leech, 2005, za: Pieniowska, 2019), jak i poprawna organizacja sesji nagranych.

Kolejnym etapem, w którym również znajdują zastosowanie nowe technologie, jest transkrypcja zgromadzonego materiału. Możliwość opracowania utrwalonej wypowiedzi w cyfrowym programie zapewnia badaczowi precyzję w prowadzeniu analizy jakościowo-ilościowej. Współcześnie badacze mają do wyboru kilka programów specjalnie przeznaczonych do anotacji danych mówionych. Technicznie możliwe jest automatyczne przeniesienie danych z pliku dźwiękowego do programu przeznaczonego do anotacji (automatyczna transkrypcja). Zabieg zawsze wymaga starannej korekty. Z tego powodu badacze czasem wybierają czasochłonną analizę słuchową oraz anotację ręczną (milisekunda po milisekundzie). Dzięki precyzyjnej linii czasu oraz skoordynowanym z nią obrazem fali dźwiękowej dokładność zapisu danych może wynosić do jednej tysięcznej sekundy. Różne narzędzia do analizy danych mówionych przywołała i w syntetyczny sposób

scharakteryzowała Pieniowska (2019). Wśród nich znalazły się m.in. programy: AnnotationPro¹, Anvil², Dolmen³, Elan⁴, WaveSurfer⁵. Jednym z najczęściej i najchętniej wykorzystywanych programów są programy ELAN i Anotation Pro. Wybór narzędzia zależy, oczywiście, od głównego celu projektu oraz preferencji samego zespołu badawczego.

Powyższą listę warto uzupełnić o program EXMARaLDA⁶. Podobnie jak powyżej przywołane, służy do transkrypcji wypowiedzi mówionych. Program ten jest bardzo funkcjonalny przy analizie wypowiedzi nie tylko monologowych (Majewska-Tworek 2014), lecz także przy badaniu dynamiki i przebiegu interakcji werbalnych (Rogozińska 2014). Jego autorzy stale go rozwijają. Dodawane są do niego kolejne funkcje, dzięki którym łatwe jest pozyskiwanie interesujących ilościowych oraz jakościowych danych w trakcie analizy starannie przetranskrybowanego pliku dźwiękowego⁷. W Polsce program EXMARaLDA został zastosowany w niemiecko-brytyjsko-polskim projekcie GeWiss⁸ do opracowania dużego korpusu danych mówionych w oficjalnej, naukowej odmianie języka. Część korpusu stanowią wypowiedzi w języku polskim. Z kolei w projekcie „Biografie językowe Dolnoślązaków – pokolenie najstarsze”⁹ dane językowe zostały przetranskrybowane w programie ELAN, lecz w następnej kolejności – na potrzeby kolejnego projektu – wyeksportowano je właśnie do programu EXMARaLDA. Jest to technicznie możliwe, podobnie jak eksportowanie danych do programu Praat, gdy trzeba dokonać analizy fonetycznej.

Należy też podkreślić, że obok dogodnego narzędzia do analizy mowy, niezmiernie istotna jest konwencja transkrypcyjna, którą zazwyczaj dodatkowo dostosowuje się do realizacji danego celu. Badacz korzysta z określonej konwencji, którą przeważnie trzeba doprecyzować (ze względu na cel analizy) przez wprowadzenie dodatkowych symboli anotacyjnych. Dobrze przemyślana anotacja ułatwia bowiem, a często po prostu umożliwia, pełne korzystanie z możliwości wyszukiwarek w obrębie programu cyfrowego w trakcie pracy z korpusem na kolejnym etapie analizy. W ramach projektu GeWiss konwencja transkrypcyjna

¹ <http://annotationpro.org/> [dostęp: 15.05.2023].

² <http://www.anvil-software.de/> [dostęp: 15.05.2023].

³ <https://dolmen.readthedocs.io/en/latest/> [dostęp: 15.05.2023].

⁴ <https://archive.mpi.nl/tla/elan> [dostęp: 15.05.2023].

⁵ <https://sourceforge.net/projects/wavesurfer/> [dostęp: 15.05.2023].

⁶ <https://exmaralda.org/en/> [dostęp: 15.05.2023].

⁷ <https://exmaralda.org/en/2023/05/06/release-candidates/> [dostęp: 15.05.2023].

⁸ https://gewiss.uni-leipzig.de/index.php?id=about_gewiss&L=2 [dostęp: 15.05.2023].

⁹ <http://ppm.uni.wroc.pl/category/badania/> [dostęp: 15.05.2023 na potrzeby Clarin PL].

GAT2 została dostosowana do potrzeb kolejnych indywidualnych projektów (Rogoziński, Zarzeczny, 2012).

PRZYKŁADY KIERUNKÓW BADAŃ Z ZASTOSOWANIEM PROGRAMU DO ANOTACJI MOWY

Praca z materiałem językowym opracowanym w programie do anotacji mowy inspiruje do poszukiwań, formułowania nowych pytań, a w efekcie – umożliwia realizację nowych celów badawczych. Są to zazwyczaj pogłębione analizy jakościowo-ilościowe cech mówienia, które wcześniej – przy tradycyjnym, ręcznym transkrypcie – na pewno nie były możliwe do przeprowadzenia. Wynikiem takich analiz są m.in. prace dotyczące opisu cech mówienia osób o wysokiej sprawności językowej (Majewska-Tworek, 2014); porównania przebiegu tych samych zjawisk (np. autokorekt czy wyrazów retardacyjnych) u mówców posługujących się różnymi językami (Rogozińska, 2014) bądź też posługujących się tym samym językiem w różnych jego wariantach (Majewska-Tworek, Zaśko-Zielińska, 2020b); charakterystyki powiązań między komunikacją werbalną i niewerbalną (Karpiński i in., 2008), także w aspekcie porównawczym – międzykulturowym (Karpiński i in., 2018); poszukiwań dyskretnych wyznaczników mowy, które prawdopodobnie zdradzają początki zaburzeń neurolingwistycznych (Jankowska, 2021).

W pracy poświęconej niepełności wypowiedzi w oficjalnej odmianie polszczyzny (Majewska-Tworek, 2014) została przeprowadzona analiza polskich danych mówionych zgromadzonych na potrzeby projektu GeWiss i opracowanych w programie EXMARaLDA. Jest to materiał pochodzący z dziesięciu godzin nagrań, w których wzięli udział uczestnicy rozmów egzaminacyjnych (uwzględniono zarówno egzaminowanych studentów, jak i ich egzaminatorów po wcześniejszym wyrażeniu na to zgody). W pracy pominięto metodologię językoznawstwa strukturalnego oraz normatywnego, natomiast zastosowano perspektywę psycholingwistyczną z modelem planowania i realizacji wypowiedzi spontanicznej w centrum. W wyniku jakościowo-ilościowego opisu materiału językowego zaproponowano typologię stwierdzonych wykładników niepełnego mówienia wraz z ich charakterystyką. Typologia została opracowana na podstawie analizy wypowiedzi zdrowych mówców o wysokiej sprawności językowej. Może być stosowana jako pewien punkt odniesienia w analogicznych badaniach przeprowadzanych w grupie mówców o odmiennej charakterystyce socjolingwistycznej.

PSYCHOLINGWISTYCZNA PERSPEKTYWA
OPISU WYPOWIEDZI MÓWIONEJ

Bardzo ważne jest, by oderwać się w badaniach wypowiedzi mówionej od dominującej przez lata perspektywy normatywistycznej, wypracowanej dla pisanej odmiany języka. W takim opisie bowiem spontaniczna wypowiedź mówiona (także powstające w sytuacji oficjalnej) jest zawsze charakteryzowana przez pryzmat błędów i usterek. Niewątpliwie z pomocą przyszły tu nie tylko nowe technologie, ale i osiągnięcia tekstologii, która wskazuje na cechy obligatoryjne danego gatunku mowy (rozmowy egzaminacyjnej, wywiadu biograficznego itd.), a także osiągnięcia psycholingwistyki, w której nurcie w wyniku wieloletnich badań różnych naukowców (por. Roelofs, 2020) został opracowany oraz wielokrotnie dyskutowany model planowania i realizacji wypowiedzi ustnej. Badacze wypowiedzi mówionych, oczywiście, mają do dyspozycji kilka wariantów modelu, co wynika z różnych założeń wstępnych czynionych przez kolejnych psycholingwistów (Fromkin, Ratner, 2005, s. 356). Do najbardziej znanych i do dziś stosowanych należy model Levelta, opierający się na założeniu, że mówca jest zdolny do kontrolowania swojej wypowiedzi na każdym etapie jej powstawania (Levelt, 1989, s. 470; por. też Fromkin, Ratner, 2005, s. 362). Mówca może monitorować wiele aspektów swojej wypowiedzi. Jednak zazwyczaj nie jest w stanie kontrolować wszystkich z nich. W zależności od konsytuacji, w której powstaje wypowiedź słowna, oraz kondycji psychicznej mówcy występuje swoista hierarchia obszarów związanych z wypowiedzią słowną, które mówca poddaje ścisłej kontroli. W różnym stopniu więc kontroluje swą wypowiedź: od poziomu motywacji – czy w ogóle coś powiedzieć; przez planowanie semantyczne na poziomie mowy wewnętrznej – co powiedzieć; kodowanie językowe – jak to powiedzieć; aż po wykonanie motoryczne – czysto fizyczny aspekt mówienia, związany z artykulacją, tempem, prozodią, akcentowaniem, natężeniem głosu. Analizując wypowiedź w programie do anotacji mowy, można zaobserwować symptomy kontrolowania wypowiedzi przez mówcę. Obraz symptomów jest niepowtarzalny, mocno wpisujący się w biolekt osoby wypowiadającej się. Analiza korpusów wyraźnie wskazuje, że u mówców występują cechy mówienia zależne od uwarunkowań idiolektalnych. Są niepowtarzalne, w sposób stały charakteryzują danego mówcę (Rogoziński, Zarzeczny, 2013). Każdy mówca w nieco inny sposób radzi sobie też z przejściowymi trudnościami w trakcie mówienia spontanicznego w sytuacji nieoficjalnej czy oficjalnej.

AUTOKOREKTA JAKO CECHA SPONTANICZNEJ WYPOWIEDZI.
PRÓBA OPISU NA PRZYKŁADZIE DWÓCH WYWIADÓW BIOGRAFICZNYCH

Jak zaznaczono wcześniej, w wyniku analizy rozmów egzaminacyjnych przetranskrybowanych w programie EXMARaLDA opracowano typologię symptomów niepełnego mówienia osób młodego i średniego pokolenia użytkowników polszczyzny, osób o wysokiej sprawności językowej. Typologia może służyć jako punkt odniesienia w badaniach kolejnych grup mówców o innych wskaźnikach socjolingwistycznych. Poszczególne typy niepełności, a w szczególności autokorekty, traktuje się jako wynik autokontroli sprawowanej przez mówcę w trakcie budowania spontanicznej wypowiedzi¹⁰. W typologii niepełności wypowiedzi znalazły się:

- a) autokorekty leksykalne;
 - leksykalne autokorekty bieżące (ukryte, z przerwaniem wyrazu, tuż za korygowanym wyrazem, opóźnione o co najmniej jeden wyraz),
 - leksykalne autokorekty z odtwarzaniem wcześniejszego fragmentu wypowiedzi (ukryte, z przerwaniem wyrazu, tuż za korygowanym wyrazem, opóźnione o co najmniej jeden wyraz);
- b) urwania;
 - restarty,
 - urwania wątku z kontynuacją wypowiedzi,
 - urwania totalne;
- c) autokorekty struktury gramatycznej;
 - autokorekta struktury syntaktycznej,
 - autokorekty morfologiczne;
- d) autokorekty fonetyczne;
 - fonetyczne autokorekty bieżące,
 - fonetyczne autokorekty z odtwarzaniem wcześniejszego fragmentu wypowiedzi;
- e) pauzy;
- f) wykładniki kontynuacji wypowiedzi;
 - pauzy wypełnione,
 - przeciąganie głosek,
 - powtarzanie,
 - wyrazy retardacyjne (Majewska-Tworek, 2014, s. 91).

¹⁰ Kazimierz Ożóg, jeden z prekursorów nowoczesnych badań nad mówioną odmianą polszczyzny, zauważył: „korekta jest zawsze świadectwem intelektualnej pracy nad tekstem i jego udoskonalania” (1988, s. 35). Po latach w swej kolejnej pracy dodał: „Z wypowiedzią spontaniczną nierozzerwalnie związana jest korekta” (2001, s. 93).

Na potrzeby niniejszego tekstu zostały przeanalizowane jedynie autokorekty pochodzące z dwóch wywiadów seniorów. Osoby te udzieliły wywiadu biograficznego. Pełne metadane zostały zarchiwizowane zgodnie z konwencją Dublin Core, a nagrany materiał badano w programie EXMARaLDA. Głównym interlokutorem pierwszego wywiadu, który trwał 26,58 minut, jest 83-letni mężczyzna; główną interlokutorką drugiego (trwał 20,25 minut) jest 82-letnia kobieta. Wkład konwersacyjny obu mówców w rozmowę jest znaczny. W trakcie wywiadu biograficznego mężczyzna mówił w sumie 20,32 minuty, a kobieta – 17,34 minuty. Wiąże się on na pewno z aktywnością werbalną mówców, z chęcią współtworzenia rozmowy, ale przede wszystkim z cechami gatunkowymi wywiadu biograficznego. Prowadzący rozmowę miał głównie zadawać pytania inspirujące do mówienia oraz podgrzewać kanał konwersacyjny przez dyskretne sygnały werbalne i niewerbalne.

Teoria dotycząca symptomów niepewności i wahań towarzyszących mówieniu jest bogata zwłaszcza w literaturze anglojęzycznej. Autokorekty są w tym kontekście szerokim i fascynującym badaczy zagadnieniem. Realizowane są wciąż nowe projekty, a w ich ramach – eksperymenty badające różnych mówców znajdujących się w odmiennych sytuacjach komunikacyjnych, w dialogu lub monologu, posługujących się językiem ojczystym lub obcym (por. m.in.: Haniah i in., 2020). Autokorekty mogą być rozumiane i definiowane na wiele sposobów. Fox i Jaspersen opisali siedem różnych typów autokorekt (1995, za: Haniah i in., 2020, s. 105), szeroko ujmując ich istotę. W klasyfikowaniu brali pod uwagę sposób wprowadzania autokorekty do wypowiedzi przez opis jej struktury uwidaczniającej się w starannej transkrypcji. Wśród typów autokorekt są też wyróżniane powtórzenia i urwania. W niniejszym tekście dla zachowania szczegółowości opisu autokorekty leksykalne i gramatyczne są ujęte osobno względem urwań oraz powtórzeń. Uzasadnieniem jest koncepcja wąskiego postrzegania istoty zjawisk niepełności mówienia. Wypowiedzi z autokorektami względem wypowiedzi urwanych różnią się chociażby cechami prozodycznymi. Mają odmienną linię intonacyjną (Majewska-Tworek, 2014, ss. 116-117). Natomiast powtórzenia mogą albo towarzyszyć autokorektom, urwanom czy wyrazom retardacyjnym, albo mogą ujawnić się jako samodzielny wykładnik wahania; powtarzane są wówczas głoski, sylaby, frazy z ewentualnym towarzyszeniem pauz (właściwych lub wypełnionych typu [yyy]).

Trzeba też zaznaczyć, że w wybranym materiale do analizy autokorekty nie są dominującymi cechami wypowiedzi obu mówców. W całej wypowiedzi mówcy (CLR3_SO_19) stwierdzono pięć autokorekt leksykalnych różnego typu, sześć gramatycznych oraz jedną fonetyczną. Natomiast w wypowiedzi mówczyni

(CLR3_SO_13) znaleziono osiem autokorekt (w tym trzy korekty gramatyczne). W sumie przebadano 20 typów autokorekt, starannie odróżniając je od innych wykładników nie płynności mówienia.

Przykłady autokorekt mówcy (CLR3_SO_19) są zróżnicowane względem siebie i charakteryzuje je różny przebieg. Wymienione poniżej (w pkt. a, b) są korektami bieżącymi z przerwaniem wyrazu. W pkt. a. mówca skorygował zaplanowany wyraz (obecny już na poziomie mowy wewnętrznej, lecz ostatecznie niewypowiedziany) na inny, według niego trafniejszy, o tym samym prefiksie *prze*. Natomiast w przykładzie kolejnym (pkt. b) widać, że mówcy zależało na podaniu precyzyjnej informacji. Poprawił istotny szczegół i kontynuował swoją opowieść z dużym zaangażowaniem. Mówił w tych fragmentach w przyspieszonym tempie. Stwierdzone korekty bieżące (bez powtarzania wcześniejszych fragmentów wypowiedzi) z przerwaniem wyrazu świadczą o wysokim poziomie autokontroli mówcy¹¹.

a) przez rzekę Bug żeśmy **prze(.)**kroczyli (.) i żeśmy się osiedlili^{°h} ((0,1s)) yyy w powiecie Chełm Lubelski

b) ^{°h} to jest tak **pół na p (.) a a nawet więcej** takie wioski ^{°h} ((0,6s)) to jest osiemdziesiąt procent to no jest zza Bugu bardzo mało tutaj

Kolejne przykłady są autokorektami opóźnionymi. Za wyrazem podlegającym korekcie mówca wypowiedział fragment kolejnego wyrazu (pkt. c) oraz cały kolejny wyraz (pkt. d). Takie korygowanie wyróżnia cofanie się mówcy w wypowiedzi. Autokorekta w pkt. c. jest bez powtarzania wcześniejszego fragmentu wypowiedzi (np. *to było*). Natomiast w przykładzie d. zostało powtórzone wyrażenie z zaimkiem wskazującym, typowym dla spontanicznej wypowiedzi mówionej¹². Trzeba też podkreślić, że obie autokorekty zostały wprowadzone bez większych komplikacji kontekstualnych, czyli bez nagromadzenia innych wykładników niepewności.

c) to było yyy **dawne** woje (.) **obecne** województwo rzeszowskie

d) ^{°h} to ja mieszkałem w tej **wsii** Lublatyn w tej (.) **kolonii** Lublatyn (.) jak by ci powiedzieć

Ostatni, stwierdzony u badanego mówcy sposób korygowania leksykalnego nie jest rzadki we współczesnych (zwłaszcza nieoficjalnych) wypowiedziach mówionych: towarzyszy mu wyraz *znaczy*, podkreślający istotną dla mówcy korektę.

e) to poszła tam **na^{°h} yyy** Persję tamtędy **znaczy na Iran**

W zakresie korekt gramatycznych odnotowano przykład zamiany (z towarzyszeniem pauzy wypełnionej) strony czynnej na bierną. Jest to o tyle ciekawe,

¹¹ Jednocześnie warto odnotować, że zarówno nazbyt wysoki, jak i nazbyt niski poziom autokontroli podwyższa poziom wykładników nie płynności w wypowiedzi.

¹² Funkcjonalna analiza zaimków wskazujących w wypowiedziach mówionych powinna być celem odrębnego opisu.

że konstrukcja bierna jest zdecydowanie rzadsza w codziennej polszczyźnie (pkt. f). Natomiast w pkt. g i h. znalazły się typowe korekty morfologiczne w obrębie kategorii czasu, a w pkt. i. – interesująca poprawka struktury syntaktycznej z towarzyszeniem powtarzania spójnika *że*. Mówca bowiem kilka razy próbuje rozpocząć nową frazę, zmieniając jej strukturę. Przykład w punkcie j. ilustruje sposób kodowania językowego wypowiedzi spontanicznej: negacja dodawana jest w planowaniu realizacji wypowiedzi w drugiej kolejności. Mówca dopiero w korekcie dodaje partykułę przeczącą *nie*.

f) granicę **ustalali** na yyy **była ustalana** na Bugu

g) °hyy w tych je (.) jak jak on tam **pisz** (.) gdzie on to tam pisał

h) **miał** wybudowany . **mieli** wybudowany dom

i) °h (.) **że ja że jak** (.) **że jak ja** °h yyy (.) przyszedłem tu do szkoły

j) °h no **wwo** (.) **nie wolno** było chodzić do szkoły

Na koniec warto dodać, że w wypowiedzi nagranej mężczyzny tylko raz odnotowano korektę fonetyczną. Wyjątkowo też jest to korekta odtwarzaniem wcześniejszego fragmentu wypowiedzi. Towarzyszy jej urwanie korygowanego wyrazu, który prawdopodobnie byłby wymówiony z metatezą (pkt. k). Poprawka była więc niezbędna.

k) wszyscy **kal** (.) wszyscy byli **katolikami**

U mówczynie (CLR3_SO_13) stwierdzono nieco inny przebieg korekt, jednak także wpisują się one w zastosowaną typologię. Zauważono m.in. bieżące autokorekty bez odtwarzania wcześniejszego kontekstu. Taką autokorektę ukazuje pkt. l., w którym mówczynie dość nagle zmodyfikowała semantyczny plan swojej wypowiedzi, wprowadzając do niej informację o szkole średniej. Interesujący, bo nieco bardziej skomplikowany jest też przykład następny (pkt. m), w którym zaimek *taki* zostaje zmodyfikowany pod względem rodzaju, następnie jeszcze powtórzony. Na dodatek ten fragment wywodu jest rozerwany istotną informacją (o bohaterce wypowiedzi: *zmarła*), która jest opatrzona odrębną, zmienioną linią intonacyjną. Jest to uczynione w sposób płynny, bez żadnych pauz, najprawdopodobniej dzięki bardzo wolnemu tempu mówienia. Przykład w pkt. n. również zawiera autokorektę bieżącą, jednak z towarzyszeniem innej istotnej cechy – z przerwaniem wyrazu. Kobieta nagle bowiem zdecydowała o zastosowaniu synonimu *mieszka* względem zaczętego już (najprawdopodobniej) czasownika *jest*. Mówczynie realizuje swoją wypowiedź nie tylko bardzo miarowym, powolnym tempem, ale także z niezmiernie małym zaangażowaniem emocjonalnym. Z tego powodu wątpliwości budzą niektóre przykłady w trakcie szczegółowej analizy. W grupie takowych jest autokorekta w pkt. o., bardzo prawdopodobna, lecz w zasadzie nie wyróżniona w szczególny sposób na poziomie suprasegmentalnym wypowiedzi.

l) no a potem yyy były **moje yyy ((0,3s)) moja szkoła średnia (.) studia (.)** w Krakowie

m) **takiego** °hmm **taką** mieliśmy kilka lat temu zmarła taką przyjaciółkę która°hyyy yyy h°(.) mieszkała w Knurowie

n) trzeba to ((2,0s)) zaakceptować jeżeli się **je_ yyy mieszka tu**

o) która była **zawsze od zawsze** yyy w naszym domu

Wolne tempo mówienia umożliwiało, jak się wydaje, bieżącą korektę (bez cofania się w mówieniu i powtarzania wcześniejszych fragmentów wypowiedzi). Bieżącej korekcie sprzyjało także częste pauzowanie między jednostkami mówienia, w tym stosowanie pauz wypełnionych. Jednak w ogólnym odbiorze powolnej wypowiedzi nie były one natarczywe (pkt. p):

p) zapamiętałam różnicę **pomędzy °h yyy Śląskiem yyy yyy językiem na Śląsku** Górnym ((0,7s)) i Śląsku ((0,6s)) Dolnym

U mówczyni (CLR3_SO_13) pojawiły się też trzy autokorekty gramatyczne: dwie syntaktyczne (pkt. q, r) i jedna morfologiczna (pkt. s). Towarzyszą im powtórzenia (pkt. q, s), retardacyjne *to znaczy* (pkt. r), liczne pauzy wypełnione (pt. q, r, s).

q) bo jak się wychodzi z domu w **którym ((0,6s)) który ((0,2s)) który** yy zapewni ((0,2s)) prawda yyy dobrobyt

r) a a każdy **myśli** °h yyy (.) o tym **żeby yyy że yyy to znaczy** yyy chciałby mieć ileś tam lat mniej

s) i i i **nie przyznawa** yyy i **nie przyznaje się** do tego ((0,6s)) yyy swojego środowiska a przecież trzeba sobie cenić korzenie

Cechą wspólną obu mówców jest raczej stosowanie korekt bieżących (bez powtarzania wcześniej zaakceptowanych fraz). Prawdopodobnie jest to warunkowane zarówno charakterystyką idiolektalną mówców, jak i całą sytuacją komunikacyjną (w tym cechami gatunkowymi wywiadu biograficznego).

PODSUMOWANIE

Analiza manifestowania się przebiegu autokorekty na płaszczyźnie segmentalnej i suprasegmentalnej wypowiedzi jest możliwa wyłącznie dzięki szczegółowemu opracowaniu w programie do anotacji mowy. Przytoczone przykłady autokorygowania mówienia wskazują też na operatywność opracowanej propozycji typologii. Stwierdzone typy niepełności wpisują się w jej schemat. Wydaje się, że warto pogłębiać jakościowo-ilościowe badania przebiegu mówienia w momentach wahania u mówcy senioralnego. W pierwszej kolejności powinny być to analizy mówców o zdrowym przebiegu starzenia. Uzyskane dane mogą posłużyć jako punkt odniesienia względem danych pozyskiwanych od mówców z trudnościami neurolingwistycznymi.

BIBLIOGRAFIA

- Fox, B., Jaspersen, R. (1995). A syntactic exploration of repair in english conversation. W: P.W. Davis (red.). *Alternative Linguistics: Descriptive and Theoretical Modes* (ss. 77–134). Houston Texas: Rice University. <https://doi.org/10.1075/cilt.102.04fox>
- Fromkin, V., Ratner, N.B. (2005). Wytwarzanie mowy. W: J.B. Gleason, N.B. Ratner (red.). *Psycholingwistyka* (ss. 333–374). Gdańskie Wydawnictwo Psychologiczne.
- Haniah, A.U., Sasongko, F.K., Fauziati, E. (2020). The Use of Repetition as Self-Repair of an Efl Learner. *Language Circle: Journal of Language and Literature*, 15 (1), 104–111. <https://doi.org/10.15294/lc.v15i1.24469>
- Jankowska, K. (2021). Analiza mowy jako narzędzie diagnozy chorób neurodegeneracyjnych. *Investigationes Linguisticae*, 45, 83–92.
- Karpiński, M., Czoska, A., Jarmołowicz-Nowikow, E., Juszczyk, K., Klessa, K. (2018). Aspects of gestural alignment in task-oriented dialogues. *Cognitive Studies | Études cognitives*, 3, 1–17. <https://doi.org/10.11649/cs.1640>
- Karpiński, M., Jarmołowicz-Nowikow, E., Malisz, Z., Szczyszek, M., Juszczyk, K. (2008). Rejestracja, transkrypcja i tagowanie mowy oraz gestów w narracji dzieci i dorosłych. *Investigationes Linguisticae*, 16, 83–98. <https://doi.org/10.14746/il.2008.16.7>
- Leech, G. (2005). Adding Linguistic Annotation. W: M. Wynne (red.). *Developing Linguistic Corpora: a Guide to Good Practice* (ss. 25–39). University of Oxford. http://icar.cnrs.fr/ecole_thematique/contact/documents/Baude/wynne.pdf
- Levelt, W.J.M. (1989). *Speaking. From Intention to Articulation*. Massachusetts Institute of Technology.
- Majewska-Tworek, A. (2014). *Niepłynność wypowiedzi w oficjalnej odmianie polszczyzny. Propozycja typologii*. Quaestio.
- Majewska-Tworek, A., Zaśko-Zielińska, M., Pęzik, P. (2020). Polszczyzna mówiona miast – kontynuacja badań z lat 80. XX wieku z wykorzystaniem narzędzi lingwistyki cyfrowej. *Forum Lingwistyczne*, 7, 71–87. <https://doi.org/10.31261/FL.2020.07.06>
- Majewska-Tworek, A., Zaśko-Zielińska, M. (2020a). Funkcje wyrazów o postaci *jakby* w nieoficjalnych i oficjalnych wypowiedziach mówionych na podstawie badań korpusowych. *LingVaria*, 15, 1 (29), 103–117. <https://doi.org/10.12797/LV.15.2020.29.07>
- Majewska-Tworek, A., Zaśko-Zielińska, M. (2020b). Opis leksykograficzny jednostki mówionej. Refleksje na kanwie badań nad ‘no’ i ‘właśnie’ z wykorzystaniem korpusów polszczyzny mówionej. W: M. Bańko, W. Decyk-Zięba, E. Rudnicka (red.). *Leksykografia w różnych kontekstach*. T. 2 (ss. 177–194). Wydawnictwo Uniwersytetu Warszawskiego.
- Ożóg, K. (1988). Autokorekta wypowiedzi mówionej. *Język Polski*, 68 (1), 29–37.
- Ożóg, K. (2001). Ustna odmiana języka. W: J. Bartmiński (red.). *Współczesny język polski* (ss. 85–98). Wydawnictwo UMCS.
- Pieniowska, K. (2019). Narzędzia do analizy mowy – przegląd dostępnych programów i bibliotek. *Językoznawstwo*, 1 (13), 216–225. https://doi.org/10.25312/2391-5137.13/2019_14kp
- Roelofs, A. (2020). Self-Monitoring in Speaking. In Defense of a Comprehension-Based Account. *Journal of Cognition*, 3 (1), 18, 1–13. <https://doi.org/10.5334/joc.61>
- Rogozińska, M. (2014). *Korrekturen und Reparaturen in der mündlichen Wissenschaftskommunikation von L1- und L2-Sprechern*. Quaestio.

- Rogoziński, Ł., Zarzeczny, G. (2012). Adapting GAT2 conventions for the transcription of spoken Polish as part of the GeWiss project. *Studia Linguistica*, 31, 69–78.
- Rogoziński, Ł., Zarzeczny, G. (2013). Mówca AD_1489 w świetle wybranych parametrów ilościowych. W: A. Burzyńska, A. Libura (red.). *Sapientia Ars Vivendi. Księga ofiarowana Profesor Annie Dąbrowskiej* (ss. 317–333). Oficyna Wydawnicza Atut.
- Zaśko-Zielińska, M., Majewska-Tworek, A., Śleziak, M., Tworek, A. (2020). *Od rozmowy do korpusu, czyli jak zbierać i archiwizować dane mówione*. Quaestio.

ZASTOSOWANIE NOWYCH TECHNOLOGII W ANALIZIE NIEPŁYNNOSCI WYPOWIEDZI
NAJSTARSZEGO POKOLENIA UŻYTKOWNIKÓW POLSZCZYZNY
(NA PRZYKŁADZIE AUTOKOREKT)

Streszczenie

Rozwój nowych technologii ma niezaprzeczalny wpływ na rozwój badań mówionej odmiany języka. W tekście wskazuje się wartość zastosowania nowych technologii w badaniu wypowiedzi mówionej. Cyfrowe programy do anotacji mowy umożliwiły stawianie nowych pytań badawczych, a także szybki rozwój odrębnej metodologii opisu danych mówionych. Rozwój językoznawstwa korpusowego umożliwił z kolei gromadzenie i archiwizację danych mówionych według standardów, które zapewniają zrównoważony i odpowiednio obszerny materiał badawczy. W tekście przedstawiono również mały fragment danych mówionych uzyskanych i opracowanych w programie cyfrowym. Jest to wstępna analiza w ramach projektu poświęconego wyznacznikom wypowiedzi osób tzw. późnej dorosłości, wyróżniających się zdrowym starzeniem.

Słowa kluczowe: mówiona odmiana języka; polszczyzna mówiona; programy do anotacji mowy; niepłynność wypowiedzi; pokolenie najstarsze; autokorekty

THE USE OF NEW TECHNOLOGIES IN THE ANALYSIS OF DISFLUENCIES
IN THE SPEECH OF THE OLDEST GENERATION OF POLISH LANGUAGE USERS
(USING AUTOCORRECT AS AN EXAMPLE)

Summary

The development of new technologies has an evidential impact on the development of research into the spoken variety of the language. The text indicates the value of using new technologies in the study of spoken utterances. Digital speech annotation programs have enabled new research questions to be posed and the rapid development of a separate methodology for describing spoken data. The development of corpus linguistics, in turn, enabled the collection and archiving of spoken data according to standards that ensure a balanced and sufficiently extensive research material. The text also presents a small fragment of spoken data obtained and processed in a digital program. This is a preliminary analysis as part of a project devoted to the determinants of the statements of so-called late adulthood, distinguished by healthy aging.

Keywords: spoken language; spoken Polish; programs for speech annotation; disfluency of speech; the oldest generation; autocorrect