

Pablo Romero Fresco
Campus Universitario
Lagoas-Marcosende
36200 Vigo, Spain
promero@uvigo.es

5 October 2022

**Review of “Live subtitling with respeaking in Polish:
technology, user expectations and quality assessment”, doctoral thesis submitted
by the Ph.D candidate Łukasz Dutka
at the Faculty of Applied Linguistics of the University of Warsaw**

The Ph.D dissertation “Live subtitling with respeaking in Polish: technology, user expectations and quality assessment”, presented by the candidate Łukasz Dutka at the Faculty of Applied Linguistics of the University of Warsaw, meets the requirements set for doctoral theses. The following lines provide a review of the work submitted by the candidate, including an assessment of its different chapters and an appraisal of both its main merits and of those aspects that may be improved with a view to a potential publication of the thesis.

This is a rigorous, well-written and clearly structured piece of work that provides a valuable contribution to the area of media accessibility and, more specifically, live subtitling in Poland. With this thesis, the candidate shows a thorough understanding of the theoretical framework that underpins his study, of the practical implications that the results have on the subtitling profession and of the requirements involved in undertaking an empirical analysis at doctoral level.

After a brief introduction that sets the scene (aptly drawing on *Alice in Wonderland* to reflect on the (im)possibility of live subtitling in Poland) the first chapter, “Live subtitling at the crossroads of Audiovisual Translation and Accessibility Studies”, allows the candidate to define live subtitling as opposed to, for instance, semi-live subtitling and pre-recorded subtitling. It also situates live subtitling within the areas of Translation Studies, Audiovisual Translation and Accessibility Studies, resorting to key canonical sources as well as others (Tomaszkiewicz 2006) that are pertinent but not easily found in similar theses. The candidate shows relevant and up-to-date knowledge of the current state of affairs in these areas, including

their overlaps and intersections, and presents this as a useful vantage point from which to look at live subtitling. As the discussion moves from technical issues to social aspects that concern us all, the candidate includes his voice/say for the first time in his thesis (“Can we not end up with services that attempt to please everyone and satisfy no one?”, page 29). This is a welcome addition that could have been more present throughout the different chapters and especially in this one. Likewise, statements such as “As long as such services are available, a Deaf person’s ability to communicate is on par with anybody else” (page 23) could be reviewed, since, as stated by the candidate himself throughout the thesis, having subtitles is no guarantee of (proper, high-quality) access.

Chapter 2 provides a thorough discussion of the history of live subtitling, from its origins to its technological developments, including the invention of respeaking. Particularly informative is the section devoted to (live) subtitling in Poland, the legal situation of accessibility in this country within the wider context of accessibility regulations in the EU and the US and the social pressure exerted by Deaf activists to require live subtitling and sign language interpreting.

The third chapter describes and analyses the main live subtitling methods, with special emphasis on respeaking in Polish: performance, technological challenges and potential for future development. This provides the reader with the necessary knowledge to follow the next chapters of the thesis and includes up-to-date information about some of the cutting-edge developments in the area of speech recognition, such as the use of neural networks (page 91). Here, when discussing the recently introduced ASR (“While advances in speech recognition technology can make it more realistic in the future to use ASR with human postediting as a method of creating live subtitles, high subtitle speeds might still be a concern”, page 78), it would have been useful to draw on (or mention, in order to discuss elsewhere) results already available about the accuracy of this technology.

Chapter 4 focuses on the users of the subtitles, mostly on their needs, their expectations and their role as one of several stakeholders involved in live subtitling. This is a short chapter, but a necessary one, as the users are still largely absent from most studies dealing with the quality of live subtitles. This is true for a great deal of the work that I have conducted in this area and largely applies to the doctoral work under scrutiny here.

The fifth chapter addresses the elusive notion of quality, aptly identifying and analysing the various dimensions included in live subtitling quality, both linguistic and technical, and then moves on to describe existing metrics of live subtitling quality such as accuracy, latency, subtitle speed and reduction rate. One of the novel contributions of the thesis is included here, namely a new metric to measure text segmentation quality. This is a useful innovation that is successfully operationalised in the analysis presented in this thesis. This chapter ends on a discussion about the differences between subjective perceptions and objective measurements and a useful summary of the current knowledge about the impact of different live subtitling workflows on the quality of live subtitles. As a minor aspect to revise, statements such as the following one could be further nuanced:

However, fast-spoken programmes and topics where speech recognition tends to produce many misrecognitions will require parallel correction as any pauses in respoken might result in loss of content (page 145).

Indeed, while this may be true in Poland, research shows that it may not necessarily apply to other countries such as the US and the UK, where self-correction seems to work fine in this context.

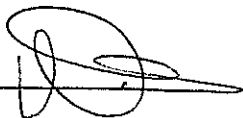
Chapter 6 constitutes the core of the thesis and it is undoubtedly its main and most valuable contribution. It offers a thorough, well thought-out and rigorously conducted analysis of the quality dimensions of live and semi-live subtitling on three Polish news channels: TVP Info, Polsat News and TVN24. The candidate discusses the method, materials and procedure of the study and critically examines the results looking at several quality dimensions (accuracy, latency, subtitle speed, reduction rate and segmentation score). The analysis includes statistical methods to ascertain the significance of the results as per the research questions posed at the beginning of the thesis, including not only the quality assessment of the individual sets of subtitles but also the major differences between live subtitles provided by the three Polish news broadcasters and the interaction between the different quality dimensions.

Along with all its merits, this chapter also reveals a series of aspects that may be subject to improvement in the thesis. Firstly, the comparison between the study presented here and the analysis commissioned by Ofcom in the UK is extremely pertinent and could have been presented and dealt with in the literature review. Secondly, while some of the most

“mechanical” aspects of the study are described at length, relatively little information is provided about how the NER analysis was performed, that is, about who was in charge, how the evaluators were trained, whether or not they were in touch with one another, etc. Finally, and perhaps most importantly, the NER model seems underused, as the candidate opts against making a distinction between edition and recognition errors and against identifying how many of the errors found were minor, standard and serious. The discussion included in this chapter and the conclusions make it clear that these data would have been extremely useful for the candidate.

Finally, the conclusions and the final recommendations contribute to providing answers to the research questions and objectives set at the beginning of the thesis. They also show how the unique position of the candidate as both researcher and professional with knowledge of the intricacies of live subtitling in Poland comes into its own and adds significant value to this thesis not only as an academic piece of work but also as an impactful study that can contribute to the improvement of live access in Poland.

In Vigo, Spain



Dr. Pablo Romero Fresco
Profesor Titular, Universidade de Vigo
Honorary Professor of Translation and Filmmaking, University of Roehampton
Director de GALMA

Pablo Romero Fresco
Campus Universitario
Lagoas-Marcosende
36200 Vigo, Hiszpania
promero@uvigo.es

5 października 2022

**Recenzja zewnętrzna rozprawy doktorskiej
pt. „Live subtitling with respeaking in Polish: technology, user
expectations and quality assessment”, złożonej przez Łukasza Dutkę
na Wydziale Lingwistyki Stosowanej Uniwersytetu Warszawskiego**

Rozprawa doktorska „Live subtitling with respeaking in Polish: technology, user expectations and quality assessment” (tytuł polski: „Napisy na żywo metodą respeakingu w języku polskim: technologia, oczekiwania użytkowników i ocena jakości”), złożona przez doktoranta Łukasza Dutkę na Wydziale Lingwistyki Stosowanej Uniwersytetu Warszawskiego spełnia wymogi stawiane pracom doktorskim. W niniejszej recenzji omówione zostaną poszczególne rozdziały pracy oraz zostanie przedstawiona ocena głównych walorów rozprawy, a także tych aspektów pracy, które mogą zostać poprawione z myślą o potencjalnej publikacji rozprawy.

Przedmiotowa praca doktorska została oparta na rygorystycznych metodologicznie badaniach; jest dobrze napisana, ma właściwą strukturę i wnosi wartościowy wkład w dziedzinę badań nad dostępnością mediów w ogólności, oraz w badania nad napisami na żywo w Polsce w szczególności.

W pracy tej doktorant wykazuje się dogłębnym zrozumieniem ram teoretycznych, w które wpisują się jego badania, jak i praktycznych implikacji, jakie niosą wyniki badań dla zawodu twórcy napisów, a także wymogów związanych z przeprowadzeniem badania empirycznego na poziomie właściwym dla rozprawy doktorskiej.

Po krótkim wstępie, stanowiącym wprowadzenie w temat i trafnie nawiązującym do „Alicji w Krainie Czarów” w ramach refleksji nad tym, czy tworzenia napisów na żywo w Polsce jest czy też nie jest możliwe, następuje pierwszy rozdział, „Napisy na żywo na pograniczu studiów nad przekładem audiowizualnym i studiów nad dostępnością”, w którym to autor definiuje pojęcie napisów na żywo na tle napisów semi-live i napisów

przygotowanych z wyprzedzeniem. Autor umiejscawia również napisy na żywo w obrębie takich dziedzin jak przekładoznawstwo, przekład audiowizualny i studia nad dostępnością, odwołując się do kluczowych pozycji z klasycznej literatury przedmiotu, jak również i innych opracowań (np. Tomaszewicz 2006), które są trafnie dobrane, a przy tym niełatwo je znaleźć w podobnych rozprawach. Doktorant wykazuje się odpowiednią znajomością aktualnego stanu wiedzy w tych dziedzinach, a także świadomością nakładania się i przenikania się tych dziedzin, wykorzystując je jako użyteczną perspektywę, z której to postrzega napisy na żywo. Gdy autor przechodzi w swoim wywodzie z dyskusji dotyczącej tematów technicznych do kwestii społecznych, po raz pierwszy sam zabiera głos („Czy konsekwencją nie jest tworzenie usług, które próbują odpowiedzieć na potrzeby wszystkich, nie zadowalając nikogo?”, str. 29). Jest to mile widziany dodatek i szkoda, że autor nie przedstawia swoich opinii częściej w pozostałych rozdziałach, szczególnie zaś w rozdziale pierwszym. Warto byłoby też zweryfikować niektóre stwierdzenia, np. „O ile takie usługi są dostępne, zdolność komunikowania się osoby Głuchej jest na równi z każdą inną osobą” (str. 23). Jak sam autor stwierdza w innych miejscach pracy, sama dostępność napisów nie stanowi gwarancji (pełnoprawnego i wysokiej jakości) dostępu.

Rozdział 2 zawiera dokładne omówienie historii napisów na żywo, od jej początków poprzez rozwój technologii, w tym wynalezienie metody respeakingu. Szczególnie pouczający jest rozdział poświęcony napisom i napisom na żywo w Polsce, sytuacji prawnej w zakresie dostępności w tym kraju, która została ukazana w szerszym kontekście regulacji dotyczących dostępności w UE i USA oraz presji społecznej wywieranej przez Głuchych aktywistów na rzecz wprowadzenia wymogu zapewniania napisów na żywo i tłumaczenia na język migowy.

Rozdział trzeci opisuje i analizuje główne metody tworzenia napisów na żywo, ze szczególnym uwzględnieniem metody respeakingu i jej zastosowania w języku polskim: efektywności, wyzwań technicznych i potencjału przyszłego rozwoju. Zapewnia to czytelnikowi wiedzę niezbędną do zrozumienia kolejnych rozdziałów, w tym aktualne informacje na temat najnowszych osiągnięć w dziedzinie rozpoznawania mowy, takich jak wykorzystanie sieci neuronowych (str. 91). Tutaj, przy omawianiu od niedawno dostępnego automatycznego rozpoznawania mowy („O ile postępy w technologii rozpoznawania mowy mogą sprawić, że w przyszłości stanie się realistyczne wykorzystanie tekstu z automatycznego rozpoznawania mowy po jego korekcie przez człowieka, to nadal problemem może być wysoka prędkość czytania napisów”, str. 78),

warto byłoby przeanalizować (lub wspomnieć, a potem szerzej omówić w innym miejscu) już dostępne wyniki badań dotyczące dokładności tej technologii.

Rozdział 4 skupia się na użytkownikach napisów, głównie na ich potrzebach, oczekiwaniach i ich roli jako jednego z kilku interesariuszy w obszarze napisów na żywo. Jest to krótki rozdział, ale konieczny, ponieważ użytkownicy są nadal w dużej mierze nieobecni w większości badań dotyczących jakości napisów na żywo. To zastrzeżenie można sformułować zarówno do wielu prac w tej dziedzinie, których ja sam jestem autorem, jak i do analizowanej rozprawy doktorskiej.

W piątym rozdziale, dotyczącym trudno uchwytnego pojęcia jakości, autor trafnie identyfikuje i analizuje różne wymiary jakości napisów na żywo, zarówno językowe, jak i techniczne, a następnie opisuje istniejące miary jakości napisów na żywo, takie jak dokładność, opóźnienie, prędkość czytania napisów i stopień redukcji tekstu. Jednym z nowatorskich elementów tej rozprawy jest to, że proponuje ona nowatorską miarę do oceny jakości segmentacji tekstu. Jest to użyteczna innowacja, która została z powodzeniem zastosowana w analizie przedstawionej w tej pracy. Rozdział ten kończy się dyskusją na temat różnic pomiędzy subiektywnym postrzeganiem jakości a jej obiektywnymi miarami oraz podsumowaniem obecnej wiedzy na temat wpływu różnych metod tworzenia napisów na żywo na jakość tychże. Można tu zgłosić drobne zastrzeżenia do niektórych stwierdzeń, które można by zniuansować; na przykład:

Jednakże, w przypadku audycji telewizyjnych, w których występuje szybkie tempo wypowiedzi oraz dotyczą one tematyki, w przypadku której wiele wyrazów jest błędnie rozpoznawanych, konieczne jest zastosowanie respeakingu z korektą wykonywaną przez moderatora, ponieważ każda sytuacja, gdy respiker musi przerwać respeaking, może spowodować utratę treści. (str. 145)

O ile może to być prawdą w przypadku Polski, to badania pokazują, że niekoniecznie musi to dotyczyć innych krajów, takich jak USA czy Wielka Brytania, gdzie korekta tekstu dokonywana przez respikera – jak się wydaje – sprawdza się dobrze w takich sytuacjach.

Rozdział 6 jest rdzeniem rozprawy i niewątpliwie stanowi jej główny i najbardziej wartościowy wkład. Rozdział ten przedstawia szczegółowe, dobrze zaprojektowane i rygorystycznie przeprowadzone badanie wielu wymiarów jakości napisów na żywo i napisów semi-live w trzech polskich kanałach informacyjnych: TVP Info, Polsat News i TVN24. Doktorant omawia metodę, materiały i procedurę badania, a także krytycznie

analizuje wyniki, dokonując przeglądu kilka wymiarów jakości (dokładność, opóźnienie, prędkość czytania napisów, stopień redukcji tekstu i ocena punktowa segmentacji tekstu). Aby upewnić się co do istotności wyników, w swojej analizie autor wykorzystał metody statystyczne, zgodnie z pytaniami badawczymi postawionymi na początku rozprawy. Analizie statystycznej poddana zostały nie tylko ocena jakości poszczególnych próbek napisów, a także główne różnice pomiędzy napisami na żywo dostarczonymi przez trzy polskie telewizje informacyjne, a także interakcje pomiędzy różnymi wymiarami jakości.

Przy wszystkich swoich walorach, rozdział ten ujawnia również pewne aspekty, które mogłyby zostać ulepszone. Po pierwsze, bardzo zasadne byłoby dokonanie porównania przedstawionego tu badania z analizą zleconą przez Ofcom w Wielkiej Brytanii, którą można było omówić w przeglądzie literatury. Po drugie, podczas gdy niektóre z najbardziej „mechanicznych” aspektów badania są obszernie opisane, rozdział zawiera stosunkowo mało informacji o tym, jak przeprowadzono analizę NER, a w szczególności, kto kierował tą analizą, w jaki sposób przeszkolono osoby oceniające, czy pozostawały one ze sobą w kontakcie, czy też nie, itp. Po trzecie, i co wydaje się najważniejsze, wydaje się, że model NER mógł zostać w pracy wykorzystany w większym zakresie, jako że doktorant zrezygnował z dokonania rozróżnienia między błędami edycji a błędami rozpoznawania mowy, a także z analizy proporcji występowania błędów drobnych, błędów średniego kalibru i błędów poważnych. Zawarte w tym rozdziale rozważania, jak i konkluzje, jasno pokazują, że uwzględnienie tego rodzaju danych byłoby niezwykle przydatne.

Wnioski i zalecenia sformułowane w pracy przyczyniają się do udzielenia odpowiedzi na pytania badawcze i realizują cele postawione na początku rozprawy. Pokazują one również, jak unikalna jest pozycja doktoranta jako badacza, ale i praktyka obeznanego w zawiłościach procesu tworzenia napisów na żywo w Polsce. Przydaje to pracy znacznej wartości jako nie tylko pracy akademickiej, ale także wpływowemu badaniu, które może przyczynić się do poprawy dostępności treści prezentowanych na żywo w Polsce.

Sporządzono w Vigo, Hiszpania

dr Pablo Romero Fresco

Profesor tytularny Uniwersytetu w Vigo

Honorowy profesor przekładoznawstwa i filmoznawstwa na Uniwersytecie w Roehampton

Dyrektor centrum GALMA