

Programy wspomagające tłumaczenie „w chmurze” – co i jak w trawie piszczy?

What’s new in the world of cloud-based CAT tools?

Maria NIKISHINA

Uniwersytet Warszawski

E-mail: m.nikishina@uw.edu.pl

Abstract: Computer-assisted translation (CAT) tools has not been immune to the radical change brought about by cloud computing. There are a number of feature-rich cloud-based CAT tools as well as the web versions of market-leading desktop-based solutions on the market today. However, some translators and translation companies are still reluctant to embrace the new technology. This paper is an overview of some of the most popular cloud-based CAT tools, such as Matecat, ABBYY Smartcat, Memsource and Lilt (an interactive translation environment which uses hybrid machine-human translation). The paper explores the possibility of those alternatives for desktop-based tools to be compatible with translator’s workflow and describes their functions and features, such as terminology management, translation memory, concordance searches, resource lookup as well as machine translation integration.

Keywords: computer-assisted translation, CAT tools, cloud-based, translation, machine-human translation, interactive translation environment

Wstęp

Programy wspomagające tłumaczenie (narzędzia CAT) „w chmurze” (*cloud-based*) obecnie coraz bardziej zyskują na popularności ze względu na większą dostępność (są albo bezpłatne, albo tanie w porównaniu z tradycyjnymi programami wspomagającymi tłumaczenie, współpracują też nie tylko z systemem Windows, ale również z innymi systemami operacyjnymi, nie wymagają aktualizacji). W środowisku tłumaczy czasami jednak uważa się, że programy „w chmurze” są dobrym rozwiązaniem tylko dla początkujących tłumaczy lub tłumaczy współpracujących z biurami na zasadzie *freelance*. Motywacją do napisania niniejszego artykułu stała się więc popularna opinia, że „poważny klient” (lub „poważny pracodawca”, tzn. renomowane biuro tłumaczeń) nie będzie sobie życzył tłumaczenia wykonanego w narzędziu „w przeglądarce” ze względu na ograniczoną funkcjonalność tego typu narzędzi. Intuicyjnie nie zgadzałam się z taką uwagą; odpowiedzią na naukowe pytanie o użyteczność o narzędzi wspomagających tłumaczenie typu *cloud-based* („w chmurze”) jest niniejszy przegląd (tzw. *state-of-the-art*) podstawowych i najbardziej ciekawych zaawansowanych funkcji niektórych najbardziej popularnych narzędzi CAT „w chmurze” – Matecat, ABBYY Smartcat, Memsource i innowacyjnego interaktywnego programu Lilt. Celem niniejszego artykułu jest zatem krytyczne przedstawienie

niektórych z najbardziej popularnych narzędzi wspomagających tłumaczenie „w chmurze” i ich funkcji, przede wszystkim pamięci tłumaczeniowej, bazy terminologii, wyszukiwania w pamięci tłumaczeniowej (tzw. *Concordance*) oraz udostępnianego przez nie tłumaczenia maszynowego.

1. Programy wspomagające tłumaczenie i ich podstawowe funkcje.

Podstawową funkcją narzędzi wspomagających tłumaczenie jest zapamiętywanie już przetłumaczonych par na poziomie segmentów (najczęściej zdań) i wyrazów (terminów): pamięć tłumaczeniowa (TM, z ang. *Translation memory*) przechowuje pary przetłumaczonych zdań, które mogą zostać później wykorzystane, natomiast baza terminologii (TB, z ang. *Termbase*) przechowuje tłumaczenia i ewentualnie definicje terminów. Narzędzia CAT oferują również dodatkowe funkcje usprawniające pracę tłumacza: zachowanie formatowania tekstu wejściowego, stworzenie pamięci tłumaczeniowej poprzez zestawienie już przetłumaczonych tekstów (tzw. *Aligning*), tworzenie glosariuszy oraz korpusów; usprawniają zarządzanie projektami i pracę w zespole (poprzez analizę projektu i możliwość delegowania zadań), proces weryfikacji i korekty. Oprócz tego w oprogramowaniu można wykonywać takie typowe czynności jak: edycja tekstu, wyszukiwanie i zamienianie oraz wyszukiwanie w pamięci tłumaczeniowej (tzw. *Concordance*) wcześniej przetłumaczonych pojedynczych słów i fraz. Programy CAT sprawdzają pisownię.

Narzędzia wspomagające tłumaczenie, w tym narzędzia typu „*cloud-based*”, pozwalają na wygenerowanie plików w uniwersalnych formatach (TMX dla pamięci tłumaczeniowej i XLIFF dla projektów) używanych przez wszystkie narzędzia wspomagające tłumaczenie, co umożliwia pracę zespołową i współpracę z biurami tłumaczeń.

Narzędzia wspomagające tłumaczenie oferują nie tylko możliwość optymalizacji pracy tłumacza poprzez korzystanie z pamięci tłumaczeniowej, ale również tłumaczenie maszynowe, w tym oparte na sieciach neuronowych, cechuje się ono większą trafnością. Tzw. rozwiązania hybrydowe dające tłumaczowi możliwość korzystania zarówno z pamięci tłumaczeniowej, jak i z tłumaczenia maszynowego, zyskują na popularności.

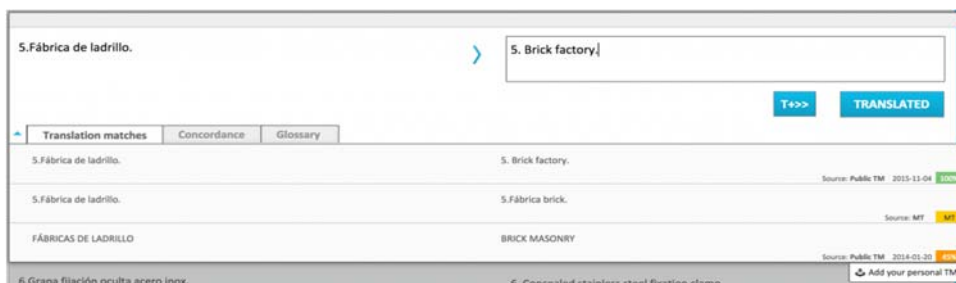
2. Pamięć tłumaczeniowa i tłumaczenie maszynowe w programie Matecat.

Innowacyjny program Matecat (MateCat) został opracowany przez zespół badaczy z europejskich uczelni przy wsparciu funduszy unijnych. Zgodnie z zapowiedzią, celem programu jest pogłębienie integracji tłumaczenia maszynowego i konwencjonalnego (tzw. *human translation*) w ramach narzędzia wspomagającego tłumaczenie¹. Pro-

¹ Federico, M. et al. (2014), *The Matecat tool*, (w:) Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations, Dublin, Ireland, 129.

gram nie wymaga rejestracji (zarządzanie projektem, prywatną pamięcią tłumaczeniową oraz aplikacjami do tłumaczenia maszynowego jest jednak dostępne tylko po zalogowaniu; można logować się przez profil użytkownika Google). Twórcy Matecat zapewniają, że to narzędzie podaje o 10–20% więcej trafień², niż inne, a to dzięki otwartej (publicznej) pamięci tłumaczeniowej MyMemory, największej światowej pamięci tłumaczeniowej³. Wykorzystywane są również korpusy równoległe i Wikipedia (do tłumaczenia terminów). Przetłumaczone segmenty trafiają do serwisu MyMemory, o ile tłumacz nie stworzył prywatnej pamięci tłumaczeniowej (tworzyć i dodawać własne bazy danych można w zakładce „Language Resources”). Ze względu na wbudowaną pamięć tłumaczeniową narzędzie proponuje dwie analizy objętości pracy (w tym czasu pracy) – standardową i tzw. *Matecat weighted words* – z uwzględnieniem trafień z MyMemory i powtórzeń (*Context Match*), na podstawie których program liczy tzw. „płatne” słowa (tzw. *payable words*).

Praca tłumacza w narzędziu Matecat wygląda w następujący sposób: po lewej stronie znajduje się tekst źródłowy, a po prawej pole tekstu wejściowego; na dole wyświetlane są propozycje z pamięci tłumaczeniowej i tłumaczenie maszynowe (o ile nie zostało wyłączone). Na przykładowym zrzucie ekranu na pierwszym miejscu po prawej stronie znajduje się stuprocentowe trafienie z pamięci tłumaczeniowej (na zielono), na drugim – tłumaczenie maszynowe (na żółto), na ostatnim (trzecim) – mniej trafna propozycja z pamięci tłumaczeniowej. Program wyświetla trafienia nawet poniżej 50%; można także wyszukać pojedyncze zwroty i wyrażenia za pomocą funkcji *Concordance*.



Zdjęcie 1. Ilustracja z *Matecat User manual and installation guide*.

Jeśli w tekście znaleziono powtórzenia, system po przetłumaczeniu pierwszego segmentu automatycznie podstawia tłumaczenie pozostałych, ale tłumacz będzie musiał je zatwierdzić klikając „*translated*”. Narzędzie wylicza prędkość tłumaczenia i pokazuje, ile procent tekstu i wyrazów zostało już przetłumaczone, a ile pozostało do przetłumaczenia wraz z przewidywanym czasem pracy.

² *Matecat User manual and installation guide*. (URL https://www.matecat.com/wp-content/uploads/2014/10/MateCatUserManualandInstallationGuide_v3.pdf, 4). [Pobrano 22.08.2017].

³ URL <http://mymemory.translated.net/doc/>. [Pobrano 22.08.2017].

Tłumaczenie maszynowe jest dostarczane przez serwis MyMemory i jest kombinacją translatorów Google Translate i Microsoft Translator. Od czerwca 2017r. w programie jest dostępne (odpłatnie) tłumaczenie maszynowe oparte na sieciach neuronowych, tzw. *Modern Machine Translation* (MMT). Decyzja o używaniu tłumaczenia maszynowego jako uzupełnienia pamięci maszynowej jest uzasadniona przez badania, które dowiodły, że niniejsze rozwiązanie jest skuteczne (zob. M. Federico et al. 2014: 126). Jak widać na rysunku, tłumaczowi jest bardzo łatwo korzystać z podpowiedzi; może on edytować tłumaczenie maszynowe, a może po prostu używać go jako narzędzia pomocniczego. Oprócz tego program Matecat pozwala na automatyczne przetłumaczenie całego tekstu w oparciu o wbudowany translator i dostępne pamięci tłumaczeniowej i następnie edytowanie tekstu w trybie „*Revise*”⁴. Zaletą programu jest także sprawnie działająca automatyczna organizacja tagów (po tłumaczeniu należy kliknąć „*Guess tags*”). Program pozwala na rozdzielanie segmentów na krótsze (tzw. funkcja *Split*) i łączenie segmentów (*Merge*), ale tylko tych, które zostały rozdzielone przez Matecat. Przy pomocy funkcji „*pre-translate 100% matches*” można już z góry zatwierdzić 100-procentowe trafienia i przyspieszyć proces tłumaczenia. W trybie offline można przetłumaczyć do 30 segmentów.

Podsumowując, można powiedzieć, że program Matecat jest łatwy w obsłudze i przyjazny użytkownikowi (program „podpowiada”, jakiego skrótu klawiaturowego trzeba użyć, wszystkie napisy są przejrzyste i zrozumiałe, program nie posiada wielu funkcji, tylko niezbędne minimum). Dużą zaletą jest bezpłatność, dostępność internetowych szkoleń i instrukcji użytkownika oraz działające dosyć szybko wsparcie techniczne. Moim zdaniem jest to naprawdę dobre rozwiązanie dla początkujących tłumaczy. Doświadczonemu tłumaczowi zapewne jednak zabraknie możliwości związanych z podstawianiem tzw. *non-translatables* oraz stworzeniem glosariuszy: glosariusz programu nie uwzględnia ani wielkości liter, ani końcówek (co jest bardzo ważne w przypadku języków słowiańskich). Jakość trafień z otwartej pamięci tłumaczeń MyMemory czasami nie jest zadowalająca.

3. Interaktywne narzędzie Lilt – przyszłość tłumaczenia?

Program Lilt to interaktywne narzędzie wspomagające tłumaczenie⁵; obsługuje ograniczoną liczbę par językowych: w listopadzie 2017 r. pozwalało wyłącznie na tłumaczenie między językiem angielskim i 12 innymi językami, w tym polskim i rosyjskim (inne pary językowe mogą zostać dostarczone klientom biznesowym na życzenie). Cena abonamentu dla tłumacza zaczyna się od 15\$ za miesiąc (tłumaczenie do 15 tys. słów; 24\$ za 150 tys. słów; abonament roczny kosztuje odpowiednio 165\$ i 264\$); dostępna jest 14-dniowa darmowa wersja próbna; można logować się za pomocą profilu Google. Lilt działa na bazie tzw. *Predictive Translation Memory* opracowanej na

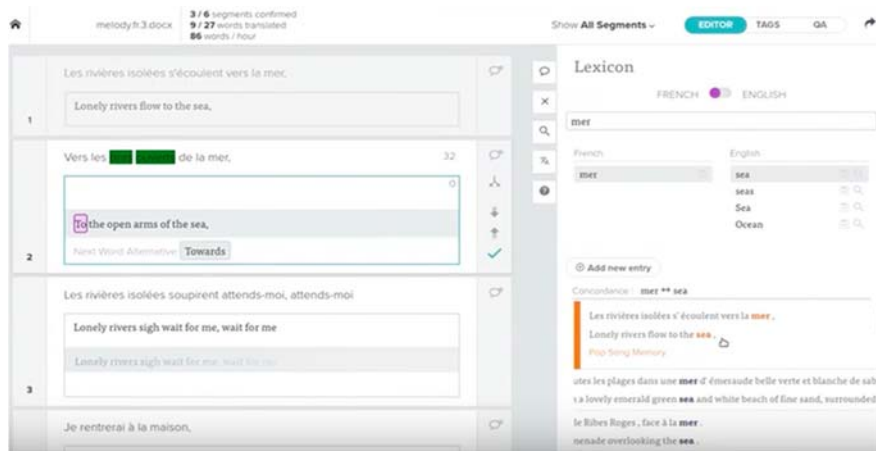
⁴ Trzeba podkreślić, że tzw. postedycja (edycja tłumaczenia maszynowego) i weryfikacja tekstu to nie jest to samo. p. norma PN-EN ISO 17100:2015-06.

⁵ Interaktywne tłumaczenie maszynowe jest również dostępne w programie Trados 2017.

Uniwersytecie w Stanford⁶ i służy *stricte* do edycji tłumaczenia maszynowego, chociaż twórcy programu podkreślają, że czasami wpisanie własnego tłumaczenia może być szybsze niż redagowanie tłumaczenia maszynowego i zalecają, aby nie skupiać się cały czas na propozycjach tłumacza⁷.

Interfejs graficzny jest podobny do narzędzi wspomagających tłumaczenie, z tym, że każdy segment pojawia się na oddzielnej „kartce”; istnieją również funkcje specyficzne dla danego narzędzia. Pamięć tłumaczeniowa wykorzystywana jest nie tylko do przechowania przetłumaczonych segmentów, ale również do uczenia się systemu; propozycje tłumacza zmieniają się w zależności od poprawek wprowadzonych przez tłumacza, w efekcie tłumaczenie maszynowe się polepsza. Warto zaznaczyć, że tłumaczenie maszynowe (*output*) zmienia się w trakcie pracy w zależności od wyrazów wprowadzanych w danym momencie przez tłumacza. Podobnie jak w innych narzędziach, można od razu wkleić tłumaczenie maszynowe. Wgrana baza terminologii służy również do uczenia się systemu; system uwzględnia tłumaczenie terminów w tekście docelowym. Aby osiągnąć wysoką jakość tłumaczenia maszynowego, tłumacz musi tworzyć różne bazy do różnych projektów i dziedzin ze względu na specyfikę systemu.

Program używa autouzupełniania wyrazów, co ma przyspieszyć pracę tłumacza. Funkcja „Lexicon” łączy w sobie wyszukiwanie w pamięci tłumaczeniowej (*Concordance*) i bazę terminologii. Aby ją uruchomić, trzeba zaznaczyć wyraz i kliknąć na lupę – pojawi się okno z glosariuszem i przykładami tłumaczenia zaznaczonego wyrazu w kontekście z pamięci tłumaczeniowej:



Zdjęcie 2. Zrzut ekranu ze szkolenia internetowego⁸.

⁶ Więcej na ten temat p. Green S. et al. (2014), *Predictive Translation Memory: A Mixed-Initiative. System for Human Language Translation*. (URL <https://idl.cs.washington.edu/files/2014-PTM-UIST.pdf>). [Pobrano 05.09.2017].

⁷ URL <https://lilt.com/kb/translators/best-practices>. [Pobrano 05.09.2017].

⁸ URL <https://www.youtube.com/watch?v=kJ8dHw815z4>. [Pobrano 05.09.2017].

Program działa na zasadzie „separation of concerns”, co znaczy, że tłumaczenie tekstu, organizacja tagów i kontrola jakości (*QA check*) są rozdzielone. Tłumacząc tekst w programie tłumacz nie widzi tagów; organizacja tagów znajduje się w oddzielnej zakładce („Tags” na górze po lewej stronie). Tagi można przeciągać myszką. W oddzielnym oknie (*QA*) są widoczne wszystkie kwestie dotyczące jakości (*quality issues*), na przykład podwójne spacje. Program Lilt ma także funkcję zarządzania projektami.

Tak samo jak program Matecat, Lilt pozwala na rozdzielanie i łączenie segmentów poprzednio rozdzielonych przez niego. Wylicza prędkość tłumaczenia i pokazuje, ile segmentów i wyrazów zostało przetłumaczonych wraz z przewidywanym czasem pracy; nie pokazuje natomiast, ile wyrazów zostało do przetłumaczenia. Lilt nie działa w trybie offline.

Moim zdaniem ten program, jak również program Matecat, jest dosyć łatwy w obsłudze i nie ma wielu funkcji; mógłby być dobrym rozwiązaniem dla początkujących tłumaczy i „wolnych strzelców”, gdyby nie konieczność tłumaczenia podobnych do siebie dokumentów z jednej branży, co jest niezbędne do uczenia systemu opartego na sieciach neuronowych. Lilt różni się od konwencjonalnych programów tym, że polega przede wszystkim na redagowaniu tłumaczenia maszynowego. Nie korzystałam z niego na tyle, żeby dostawać automatyczne tłumaczenie nie wymagające korekty więc kwestia możliwej maksymalnej prędkości tłumaczenia pozostaje otwarta.

4. Inne rozwiązania „w chmurze” – Memosource i SmartCAT9

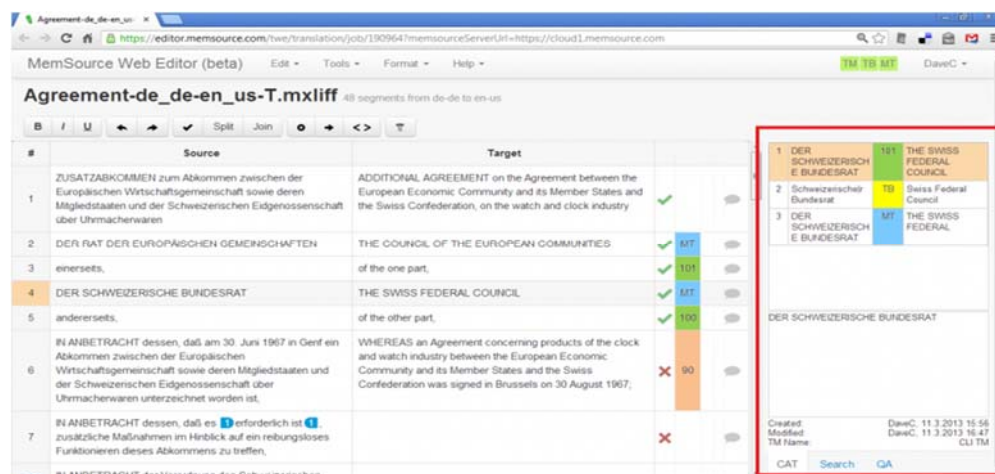
Narzędzia wspomagające tłumaczenie Memosource i SmartCAT, opracowane odpowiednio w Czechach i Rosji, mają podobny do siebie układ graficzny (p. rysunki na następnej stronie): po prawej stronie znajdują się trafienia z pamięci tłumaczeniowej oraz wyszukiwanie w pamięci tłumaczeniowej. Pod spodem w programie Memosource jest bardzo wygodny podgląd oryginału dokumentu; w SmartCAT – komentarze do dokumentu/ segmentu, historia zmian (można więc zobaczyć wszystkie wersje dokumentu) i automatyczna kontrola jakości. Program SmartCAT jest bezpłatny dla tłumaczy, natomiast w programie Memosource można przetłumaczyć w ramach jednego projektu maksymalnie 2 pliki o maksymalnej wielkości 10 Mb każdy. W obu narzędziach można używać Microsoft Translator bezpłatnie z informacją zwrotną.

Z najciekawszych funkcji, których nie mają narzędzia Matecat i Lilt, można wymienić następujące: w niniejszych programach można eksportować dokumenty równoległe (dwujęzyczne) oraz rozdzielać i łączyć wszystkie segmenty (niezależnie od tego, czy zostały one poprzednio rozdzielone przez program czy nie). Narzędzia pozwalają również na stworzenie własnej pamięci tłumaczeniowej z równoległych tekstów, system dopasuje je jednak do siebie automatycznie. Oba narzędzia mają rozbudowaną funkcję *pre-translate*. W Memosource istnieje funkcja autouzupełniania (na podstawie danych z pamięci maszynowej i tłumaczenia maszynowego). SmartCAT,

⁹ Narzędzie SmartCAT jest narzędziem „w chmurze”, ale Memosource ma także aplikację „Editor”, która współpracuje nie tylko z systemem Windows, ale również z systemami Macintosh i Linux. Wszystkie funkcje są jednak dostępne po połączeniu się z Internetem.

oprócz tego, pozwala na podstawianie liczb (trafienia z pamięci tłumaczeniowej, które różnią się od tekstu wejściowego tylko liczbami, zostaną automatycznie skorygowane i będą oznaczone jako 100-procentowe trafienia). W obu programach można wybrać próg wyświetlania trafień z pamięci tłumaczeniowej. Użytkownik programu SmartCAT może wybrać automatyczne zatwierdzenie powtórzeń, co przyspieszy proces tłumaczenia. Niniejsze narzędzia mają także bardziej funkcjonalne glosariusze: glosariusz w programie Matecat rozpoznaje termin niezależnie od wielkości liter i końcówek przypadków, co jest bardzo ważne przy tłumaczeniu w językach słowiańskich. W programie Memsorce można wprowadzić informacje o terminie i użyć specjalnego symbolu | (kreski pionowej), żeby oddzielić temat wyrazu od końcówek fleksyjnych.

Użytkownicy programu SmartCAT mogą korzystać ze słowników firmy ABBYY Lingvo poprzez wbudowaną wyszukiwarkę. Jako tłumacz angielsko-rosyjski i francusko-rosyjski, mogę z czystym sumieniem powiedzieć, że to są dobre słowniki w tych parach językowych. Według mnie dużą wadą programu SmartCAT jest płatna usługa rozpoznawania formatów graficznych (takich jak JPEG, DJVU lub PDF), bo wiem tłumacze często mają do czynienia z plikami w różnych formatach.

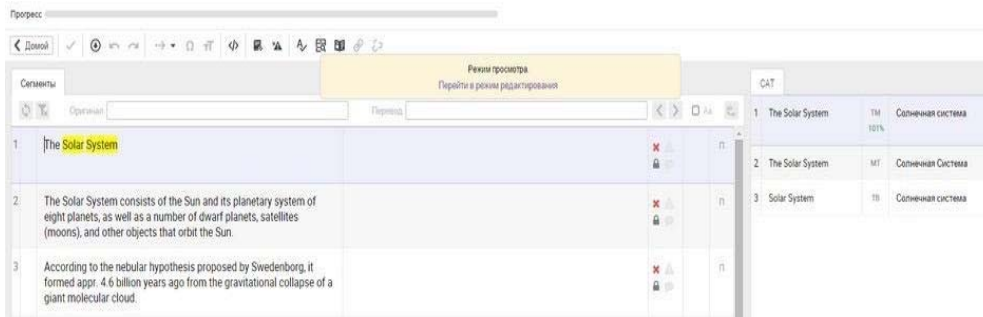


Zdjęcie 3. Ilustracja z Memsource Editor User Manual¹⁰.

Poniżej fragment ilustracji z instrukcji obsługi programu SmartCAT¹¹:

¹⁰ URL https://wiki.memsorce.com/wiki/Memsorce_Editor_User_Manual. [Pobrano 05.09.2017].

¹¹ URL https://smartcat.ai/Help/corp/ru/translating_a_document_ru_parent.htm. [Pobrano 05.09.2017].



Zdjęcie 4. Ilustracja z instrukcji obsługi programu SmartCAT.

Memsource i SmartCAT to zdecydowanie bardziej zaawansowane programy niż Matecat. Są też trochę trudniejsze w obsłudze ze względu na różne funkcje i mogą być nieco większym wyzwaniem dla początkującego tłumacza. Doświadczeni tłumacze jednak zapewne szybko sobie z nimi poradzą, ponieważ te narzędzia są bardzo podobne do popularnych konwencjonalnych programów CAT. Chociaż programy są podobne do siebie i mają również bogatą funkcjonalność, moim zdaniem program Memsource jest bardziej przyjazny użytkownikowi i dlatego bez wątpienia może zastąpić konwencjonalny program w codziennej pracy profesjonalnego tłumacza.

5. Podsumowanie i wnioski

Bez wątpienia można stwierdzić, że narzędzia wspomagające tłumaczenie „w chmurze” mają taką samą podstawową funkcjonalność (zapamiętywanie segmentów, tworzenie glosariuszy, zachowanie formatu tekstu wejściowego) i są w stanie zapewnić płynną pracę zespołową poprzez generowanie plików o uniwersalnych formatach TMX i XLIFF używanych przez inne popularne narzędzia, takie jak na przykład Trados¹².

Najprostszym w obsłudze i najbardziej przyjaznym dla użytkownika z punktu widzenia łatwości w obsłudze oraz dostępności szkoleń i innych informacji dotyczących pracy w programie jest program Matecat. Programy Smartcat i Memsource mają jednak bogatszą funkcjonalność i są zbliżone pod tym względem do konwencjonalnych narzędzi wspomagających tłumaczenie. Program Lilt jest inny niż popularne narzędzia CAT i tłumacz może potrzebować więcej czasu, żeby się do niego przyzwyczaić. Jest to dobre narzędzie tylko pod warunkiem, że tłumacz ma do czynienia z wieloma dokumentami z jednej branży i może zbudować oddzielne pamięci tłumaczeniowej dla każdej dziedziny oraz nie ma nic przeciwko redagowaniu tłumaczenia maszynowego.

Podsumowując, można stwierdzić, że zarówno początkujący, jak również profesjonalni tłumacze mogą znaleźć odpowiednie dla siebie narzędzie CAT „w chmurze”, jeśli z jakiegoś powodu ich wybór padł na tego typu narzędzia (na przykład są użytkownikami systemu MacOS, chcą mieć dostęp do tłumaczeń z różnych komputerów,

¹² Warto wspomnieć, że Trados również ma wersję „w chmurze” — tzw. *SDL Online Editor*.

komputer nie spełnia wymagań sprzętowych do instalacji konwencjonalnego oprogramowania itd.)

Korzystanie z narzędzi w chmurze może jednak budzić pewne wątpliwości co do bezpieczeństwa danych ze względu na ich przechowywanie na zewnętrznym serwerze, mimo że w niniejszych narzędziach używa się specjalnych środków ochrony. Ten temat pozostaje do dalszych badań.

Bibliografia

Federico, M. et al. (2014), *The Matecat tool*, (w:) Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations, Dublin.

Green S. et al. (2014), *Predictive Translation Memory: A Mixed-Initiative System for Human Language Translation*. (URL <https://idl.cs.washington.edu/files/2014-PTM-UIST.pdf>). [Pobrano 05.09.2017].

Matecat User manual and installation guide. (URL https://www.matecat.com/wp-content/uploads/2014/10/MateCatUserManualandInstallationGuide_v3.pdf, s. 4). [Pobrano 22.08.2017].

Norma PN-EN ISO 17100:2015-06.

Źródła internetowe

URL https://wiki.memsource.com/wiki/Memsource_Editor_User_Manual. [Pobrano 22.08.2017].

URL <https://smartcat.ai/Help/>; <https://lilt.com/kb/>. [Pobrano 23.08.2017].