

Imperializm lingwistyczny

1 marca 2015

Dominacja języka angielskiego ma wpływ na jakość usług takich jak Google translator, a także na autouzupełnianie tekstu w urządzeniach takich jak smartfony. Czy doprowadzi do wytworzenia nowego języka kreolskiego opartego na konstrukcjach podsuwanych przez algorytmy?

Jeśli na początku grudnia ub. roku zdarzyło nam się zlecić translatorowi Google przełożenie na włoski wyrażenia „Cette fille est jolie” („Ta dziewczyna jest ładna”), otrzymywaliśmy osobliwie brzmiące zdanie „Questa ragazza č abbastanza”, które należałoby przetłumaczyć jako „Ta dziewczyna jest dość”. Najwyraźniej piękno „is lost in translation”, zostało utracone w tłumaczeniu. Dlaczego jeden z najwydajniejszych automatycznych translatorów świata – potężna i unikatowa stolica lingwistyczna, zbudowana z milionów zdań – popełnił tak rażący błąd? Odpowiedź okazuje się prosta: etapem pośrednim tłumaczenia jest język angielski: „jolie” („ładna”) tłumaczy się jako „pretty”, „pretty” zaś – jako „abbastanza”.

Gdy ma się świadomość istnienia tej zasady, całkiem łatwo produkować zdania osobliwe i nierzadko zabawne. „Je pense que vous avez un président magnifique” („Myślę, że macie wspaniałego prezydenta”) staje się „Penso che tu abbia una bella sedia”, czyli „Myślę, że masz ładne krzesło”, ponieważ słowo „président” („prezydent”) Google tłumaczy jako „chair”.

Wykorzystanie angielskiego jako języka osiowego może prowadzić do nieporozumień. „Hai fatto un compito terrificante”, czyli „Zrobiłeś zadanie bardzo źle” to dla tłumacza Google „Zrobiłeś świetną robotę”. Wszystko przez pośrednictwo angielskiego „terrific”. Francuskie wyrażenie idiomatyczne „Il pleut des cordes” („Leje jak z cebra”) przemienia się w nieco poetyckie „Piove cani e gatti” – „Pada kotami i psami”. Bezpośrednie tłumaczenie „It rains cats and dogs” okazuje się zupełnie

niezrozumiałe dla Włocha.

CO WYNIKA Z JĘZYKA OSIOWEGO?

Aby opracować automatyczny translator, należy dysponować wielkim korpusem identycznych tekstów przełożonych z jednego języka na drugi. Amerykańska firma Google swoje narzędzie stworzyła na podstawie logiki: bazowano na parach tekstów i prawie zawsze stosowano angielski jako język osiowy. Przy tłumaczeniu z jakiegokolwiek języka musimy więc skorzystać z jego pośrednictwa.

Konsekwencje są poważne. O ile bowiem francuski i włoski to języki stosunkowo bliskie, o tyle angielski znacznie się od nich różni – jest językiem wyjątkowym, zwartym i bogatym w wyrażenia idiomatyczne. Złe zrozumienie kontekstu toruje drogę licznym błędom. Rzucenie wyrażenia w anglofońską przestrzeń i zabranie jej w obszar języka docelowego prowadzą do mimowolnych innowacji lingwistycznych.

Kurioza produkowane przez automatyczne translatory są wręcz anegdotyczne. Owszem, tłumaczenia w ogóle, a tłumaczenie automatyczne w szczególności, wiążą się z problemami trudnymi do rozwiązania. Skoro tak, trudno się dziwić omyłności maszyn. Ich błędy nie są zresztą bezwartościowe – zmuszają nas do zastanowienia się nad specyfiką każdego języka. Stworzenie dwujęzycznych korpusów z pominięciem pośrednictwa angielskiego i korekt dokonywanych przez internautów powinno poprawić jakość tłumaczeń. Błędy opisane w tym artykule może będą już poprawione, gdy trafi on do druku. Czy zatem to zjawisko rzeczywiście powinno nas niepokoić?

By zrozumieć realne skutki uczynienia angielskiego językiem osiowym, musimy umieścić tłumaczenie automatyczne w szerszym kontekście tekstów wykorzystujących internetowe algorytmy. Te programy informatyczne nie tylko generują lingwistyczne innowacje w sferze zrobotyzowanych tłumaczeń, lecz także są wykorzystywane do automatycznego redagowania artykułów

prasowych, korekty zawartości Wikipedii pod kątem składni i semantyki, tworzenia celowanych wiadomości o charakterze reklamowym czy też optymalizacji treści strony WWW.

KIEDY ALGORYTMY PISZĄ ZA NAS

A więc jak odróżnić pierwotne źródła lingwistyczne, stworzone przez ludzi, bez mediacji algorytmicznej (pisemne konwersacje, zawartość książek numerycznych itp.), od wtórnych źródeł językowych, które powstały w wyniku algorytmicznych transformacji źródeł pierwotnych?

W związku z rozpowszechnieniem się autouzupełniania (użytkownik wpisuje początek wyrażenia, a program kończy za niego) algorytmy stały się niemalże systemowymi łącznikami. Najczęściej już po chwili przestajemy pisać litera po literze czy słowo po słowie – zadowalamy się wyborem kontynuacji spośród wielu możliwości zaproponowanych przez algorytmy. Taka forma pisania łączy szybkość ze skutecznością, zwłaszcza gdy korzystamy ze zredukowanej klawiatury telefonu komórkowego. Ten nowy kontekst ogranicza pisanie do wyboru opcji spośród całej gamy przewidywalnych wyrażen. Za kilka lat z pewnością trudno będzie znaleźć interfejs niekorzystający z tej technologii.

Podobnie jak inne źródła wtórne, teksty produkowane algorytmicznie przez automatyczne translatory niekoniecznie są jako takie identyfikowane czy oznaczane. Dzieje się dokładnie odwrotnie: często wyglądają na źródła pierwotne, naturalne. Internauta, dla którego włoski nie jest językiem ojczystym, nie ma żadnych powodów, by zdanie „Piove cani e gatti” uznać za niepoprawne. Dotyczy to zwłaszcza algorytmów analizujących strukturę języka w celu sztucznego tworzenia nowych tekstów. Algorytm, który podczas prób optymalizacji swoich możliwości translatorskich szuka źródła pierwotnego, może przez przypadek wykorzystać tekst stworzony przez inny algorytm, a to zaowocuje fałszywym sensem i fałszywymi przyjaciółmi.

Internetowa ekspansja źródeł „skażonych” przez maszyny zagraża całości budowli technologicznej, która ogrom danych przedkłada nad skrupulatną kontrolę ich jakości. W sieci roi się od językowych osobliwości. Dowód? W internetowym sklepie Apple (AppStore) czytamy: „Aplikacje do obsługi kulki Sphero zapewnią pasjonującą przygodę z rozgrywką w trybie rzeczywistości mieszanej. Kulka Sphero przeniesie Cię w zupełnie inny wymiar rozgrywki, zwany rzeczywistością mieszaną, która płynnie łączy elementy wirtualne i realne. Dzięki wielu bezpłatnym aplikacjom do pobrania (ciągle powstają nowe) kulka Sphero zapewni Ci mnóstwo atrakcji”. Kolejny fragment, pisany w podobnym stylu: „Etui iPad Air 2 Smart Case chroni iPada Air 2 zarówno z przodu, jak i z tyłu, a jednocześnie zachowuje jego smuklejszą i lżejszą formę. Gdy otworzysz etui Smart Case, iPad Air 2 w magiczny sposób natychmiast się budzi”.

Zdania z przedziwnymi wyrażeniami autorstwa algorytmów mogą później posłużyć jako model usługom mediacji tekstowej, które oferują np. autouzupełnianie wpisywanego wyrażenia. Niewykluczone, że za jakiś czas Włoch rozpoczynający zdanie od *Piove* ujrzy następującą propozycję dalszego ciągu: „*cani e gatti*” – czyli wyrażenie, którego nikt w całej historii języka włoskiego ani nigdy nie wypowiedział, ani nie napisał.

KREOLIZACJA

Tak oto angielska oś lingwistyczna uczestniczy być może w zjawisku kreolizacji – tworzeniu nowego języka przez transformacje języków innych, starszych. Jest ono dobrze znane lingwistom. Modyfikacje wprowadzone przez mediację algorytmiczną zaowocowały stworzeniem pewnego rodzaju pidginu, języka kontaktu dla dwóch systemów lingwistycznych, którego żywot raczej nie będzie długi. Ponieważ młoda generacja styka się nieustannie z takimi wyrażeniami, innowacje mogłyby przybrać regularną formę spójnego i autonomicznego języka – języka kreolskiego.

Angielski imperializm językowy wywołuje o wiele subtelniejsze skutki, niż można by się spodziewać po podejściu nastawionym na „wojnę języków”. Fakt, że obracamy się wokół jednego tylko idiomu, przekłada się na konieczność wprowadzenia we wszystkich pozostałych własnych logik językowych, a tym samym, niepostrzeżenie – specyficznych sposobów myślenia. Zmiany te mogą doprowadzić do globalnej transformacji językowej, w której algorytmy odegrają kluczową rolę. Skoro angielski funkcjonuje jako oś dla języków europejskich, nie ma najmniejszej wątpliwości, że na odmiennych obszarach językowych tę samą pozycję zajmą inne języki (np. hindi). Sieć łańcuchów przekładu, działająca przez odwołanie do licznych idiomów pośrednich, sytuuje się tym samym na szczyblu światowym.

Czy za 5 lat ktoś jeszcze będzie produkował nieskażone źródła pierwotne, pisane bez pośrednictwa algorytmów? Ile czasu musi upłynąć, by pierwsze innowacje algorytmiczne zaczęto postrzegać jako formy naturalne? To pisanie hybrydowe bez wątpienia potrzebuje wnikliwej analizy. A może nawet rozwoju nowej lingwistyki, która by lepiej rozumieć i nadzorować efekty algorytmów, korzystałaby z pomocy... algorytmów.

Autorstwo: Frédéric Kaplan, Dana Kianfar

Tłumaczenie: Magdalena Madej

Źródło: [Le Monde diplomatique – edycja polska](#)