

# Autoreferat

## 1. Imię i nazwisko:

Jakub Szewczyk

## 2. Posiadane dyplomy, stopnie naukowe

### Doktor nauk społecznych

Radboud University, Nijmegen, Holandia, 2016

Promotor: Prof. Herbert Schriefers

Tytuł rozprawy: The mechanisms of prediction in language comprehension

### Magister psychologii

Instytut Psychologii, Wydział Filozoficzny, Uniwersytet Jagielloński w Krakowie, 2004

Promotor: Prof. Dr hab. Anna Grabowska

Tytuł: Splątanie gramatyki, morfologii i semantyki w percepcji języka: badanie primingowe; praca obroniona z wyróżnieniem

## 3. Informacja o dotychczasowym zatrudnieniu w jednostkach naukowych lub artystycznych.

- 2022-teraz      staż podoktorski: Max Planck Institute for Psycholinguistics, laboratorium *Neurobiology of Language* (kierownik: prof. Peter Hagoort) i laboratorium *Predictive Brain Lab* (kierownik: prof. Floris de Lange)
- 2021-2022      staż podoktorski, Donders Institute for Brain, Cognition and Behaviour, w ramach programu Maria Skłodowska-Curie Action, Individual Fellowship
- 2019–2021      staż podoktorski, Wydział Psychologii Uniwersytetu Illinois w Urbana-Champaign, w ramach programu Maria Skłodowska-Curie Action, Individual Fellowship
- 2017–2019      staż podoktorski, Instytut Psychologii, Wydział Filozoficzny, Uniwersytet Jagielloński, laboratorium Psychologii języka i dwujęzyczności (kierownik: prof. Zofia Wodniecka)

## 4. Omówienie osiągnięć, o których mowa w art. 219 ust. 1 pkt. 2 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2021 r. poz. 478 z późn. zm.).

W ramach osiągnięcia (rozumianego w kontekście Ustawy wskazanej w tytule sekcji) chciałbym przedstawić dwa cykle powiązanych tematycznie badań, dotyczące: 1) mechanizmów predykcji w percepcji języka i 2) wpływu podobieństwa międzyjęzykowego na nabywanie słów w języku obcym. W ramach opisu każdego cyklu przedstawiam kolejno artykuły wchodzące w jego skład, uwypuklając stopniowy rozwój rozumienia badanych procesów. Poszczególne badania starałem się przedstawić bez wchodzenia w nadmierne szczegóły, ale na tyle dokładnie, by możliwa do uchwycenia była logika łącząca badania.

#### 4.1. Pierwszy cykl publikacyjny: Predykcje w percepcji zdań

Nasz umysł jest w stanie przetwarzać język z ogromną prędkością. Mija niespełna pół sekundy od momentu, kiedy fale dźwiękowe uderzają w membrany naszych uszu do momentu, gdy zaczynamy być świadomi znaczenia słyszanego słowa w ogólniejszym kontekście zdaniowym. W czasie tego krótkiego okresu, nasz mózg wykonuje serię złożonych obliczeń, w ramach których fale dźwiękowe zamieniane są w reprezentacje o charakterze czasowo-częstotliwościowym (ang. *spectrotemporal code*), identyfikowane są poszczególne fonemy, morfemy, słowa, by wreszcie nastąpił tzw. dostęp leksykalny, czyli aktywacja znaczenia słowa przechowywana w pamięci długotrwałej. Cały ten proces przebiega bardzo sprawnie mimo faktu, że język jest inherentnie niejednoznaczny: te same dźwięki mogą odpowiadać różnym fonemom w zależności od sąsiedztwa innych fonemów. To samo słowo może mieć bardzo różne sensy, w zależności od ogólniejszego kontekstu. Mimo to, przeważnie automatycznie i bezbłędnie jesteśmy w stanie wyłowić znaczenie, które nasz rozmówca chciał zakomunikować. Wielu badaczy podejrzewa, że zarówno zadziwiająco szybkość samej percepcji języka, jak i naszą zdolność do rozwiązywania wszechobecnych niejednoznaczności zawdzięczamy predykcjom: mechanizmowi, dzięki któremu mózg wykorzystuje informacje z przeszłości (kontekst, doświadczenie, wiedza) by odpowiednio pokierować przetwarzaniem nowej informacji unikając przy tym niejednoznaczności.

W ramach przedstawionego niżej cyklu publikacyjnego skupiałem się na predykcjach rzeczowników. Zadawałem pytanie o cel predykcji (czy są nimi słowa czy raczej bardziej ogólne reprezentacje semantyczne), jaka funkcja matematyczna wiąże siłę odpowiedzi mózgowej z przewidywalnością słowa, czy (i jak szybko) predykcje są aktualizowane, w miarę jak napływają nowe informacje, i wreszcie, jaki mechanizm neuropoznawczy prowadzi do generowania nowych predykcji.

Aby móc zacząć badań mechanizmy predykcji słów musimy wpieryw dysponować oszacowaniem stopnia ich przewidywalności w zdaniu. Od kilku dekad w tym celu wykorzystuje się tzw. test luk (ang. *cloze test*; Taylor, 1953). Jest to test typu papier i ołówek. W ramach niego osoby badane proszone są o uzupełnienie niedokończonych zdań pierwszym słowem, które przychodzi na myśl, np.:

*Od razu wiedziała, że jest zdenerwowany, gdy tylko usłyszała ton jego ...*

Jeśli 94 na 100 badanych dokończy powyższe zdanie słowem "głosu", powiemy, że to słowo w tym kontekście zdaniowym ma prawdopodobieństwo równe 94%, a więc jest wysoce przewidywalne. Jednym z najbardziej klasycznych odkryć w badaniach nad przetwarzaniem języka jest to, że bardziej przewidywalne słowa (te, które uzyskują większe prawdopodobieństwo w teście luk) są łatwiej przetwarzane: w czasie czytania badani fiksują na nich wzrok krócej lub częściej je przeskakują wzrokiem; słowa te także prowadzą do mniej negatywnej amplitudy komponentu N400.

W mojej pracy skupiłem się na tym ostatnim. Jest to potencjał wywołany (ang. *Event-Related Potential*, ERP) przez każdy bodziec niosący znaczenie. Mierzy się go bezpośrednio z aktywności mózgu (za pomocą EEG) w czasie, np. gdy osoba badana czyta lub słucha zdań. Wykorzystanie metody ERP ma tę ogromną zaletę nad badaniami behawioralnymi, że nie wymaga żadnego dodatkowego zadania ponad samą recepcję języka. Dzięki temu pomiar N400 daje możliwie najczystszy wgląd w procesy leżące u podłoża recepcji języka. Najprostszym sposobem na wywołanie N400 jest użycie w zdaniu słowa, które jest niezgodne z kontekstem (np. "Zjadł kanapkę z masłem i skarpetkami", gdzie silne N400 zostanie wywołane przez ostatnie słowo zdania; Kutas & Hillyard, 1980, 1983). Amplituda N400 wywołana przez słowo jest tym mniej negatywna, im bardziej jest ono przewidywalne (dzięki poprzedzającemu je kontekstowi). Gdyby w powyższym zdaniu użyć słowa "dżemem", wywołałoby ono mniejsze (mniej negatywne) N400.

Jaki mechanizm kryje się za tym efektem? Jedna interpretacja mówi, że w powyższym przykładzie, "skarpetkami" wywołuje N400 o większej (bardziej negatywnej) amplitudzie niż "dżemem" ponieważ to drugie słowo jest bardziej przewidywalne—i co za tym idzie—zostało przez mózg w większym stopniu przewidziane, jeszcze zanim pojawiło się jako część zdania. Wg tej interpretacji, N400 odzwierciedla proces dostępu leksykalnego do słowa, czyli aktywacji jego reprezentacji semantycznych w pamięci długotrwałej. Z kolei predykcja słowa oznacza preaktywację tej reprezentacji zanim jeszcze zostanie ono napotkane w zdaniu. Wg drugiej interpretacji, słowo "skarpetkami" jest mniej zgodne z kontekstem, a wystąpienie większego N400 spowodowane jest trudnością integracji tego słowa z już przetworzoną częścią zdania. Hipoteza ta nie zakłada zatem żadnego mechanizmu predykcji, zaś we wzroście amplitudy N400 upatruje ona działanie procesu integracji słowa z kontekstem. Rozstrzygnięcie między tymi hipotezami nie jest łatwe, gdyż przewidywalne słowa są przeważnie łatwiejsze do zintegrowania z kontekstem. Aby tego dokonać, przeprowadziłem badanie opisane w artykule Szewczyk i Schriefers (2018).

#### **4.1.1. Badanie 1: Czy N400 odzwierciedla trudność integracji słowa, czy stopień w jakim zostało wzbudzone przez kontekst? (Szewczyk i Schriefers, 2018)**

W tym eksperymencie osoby badane czytały kilkuzdaniowe historyjki napisane w j. polskim. W tym samym czasie rejestrowałem aktywność ich fal mózgowych za pomocą EEG. Poniżej znajduje się przykład jednej z historyjek:

*Trzeba przyznać, że mój wujek nie ma umiaru w żartach. Gdy ostatniego lata byliśmy nad morzem, oddalił się od kąpieliska i założył na plecy trójkątną płetwę. Następnie wskoczył do wody i podpłynął do kąpieliska tak, że tylko płetwa wystawała ponad poziom wody. Narobił się straszny raban i wszyscy byli przekonani, że ujrzeli **rekina/lekarza** ...*

W ostatnim zdaniu każdej historyjki mierzyłem amplitudę N400 wywołaną pojawieniem się słowa krytycznego—rzeczownika (pogrubiony w powyższym przykładzie). Rzeczownik występował albo w warunku zgodnym semantycznie z kontekstem ("rekin"; jak wykazał test luk, słowo to było też najbardziej przewidywalne), lub w warunku niezgodnym ("lekarz", słowo o zerowej przewidywalności i zarazem niezgodne z kontekstem zdaniowym). Niezależnie od manipulacji zgodnością, w połowie prób eksperymentalnych, tuż przed ostatnim zdaniem historyjki, sztucznie wzbudzałem aktywację słowa docelowego (warunek z indukcją predykcji). Indukcja predykcji polegała na wyświetleniu informacji:

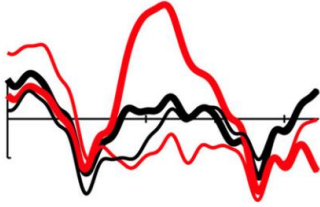
*W następnym zdaniu zobaczysz następujące słowo: **rekin/lekarz***

(osoby badane zawsze widziały tylko jeden rzeczownik i zawsze ten, który faktycznie występował jako słowo krytyczne). W drugiej połowie prób eksperymentalnych pojawiała się neutralna informacja, nie zawierająca wskazówki na temat słowa docelowego (warunek bez wzbudzonej predykcji).

Wedle hipotezy zakładającej, że N400 odzwierciedla stopień w jakim słowo zostało przewidziane (preaktywowane), uprzednia indukacja aktywacji "rekina" albo "lekarza" powinno sprawić, że gdy słowo to później pojawi się jako część zdania, wywoła ono N400 o minimalnej amplitudzie, niezależnie od stopnia jego zgodności ze zdaniem. Z kolei, gdy predykcja słowa krytycznego nie zostanie uprzednio zaindukowana, jego pojawienie się jako część zdania wywoła N400 o amplitudzie proporcjonalnej do stopnia jego zgodności (i przewidywalności): niewielkie N400 dla "rekina" i bardzo duże N400 dla "lekarza".

Z kolei wedle hipotezy zakładającej, że N400 odzwierciedla wyłącznie trudność integracji słowa z kontekstem zdaniowym, uprzednie wzbudzenie predykcji słowa nie będzie miało jakiegokolwiek

wpływu na późniejsze N400 wywołane pojawieniem się słowa krytycznego jako części zdania. Amplituda N400 będzie zależała wyłącznie od stopnia jego zgodności z kontekstem. Innymi słowy, "lekarz" zawsze wywoła N400 o dużej amplitudzie, niezależnie od indukcji predykcji lub jej braku.



Rycina 1. Potencjały EEG wywołane przez słowa docelowe (elektroda Pz). Cienkie linie: warunki ze wzbudzoną predykcją. Pogrubione linie: warunki bez wzbudzonej predykcji. Czarny kolor: warunek zgodny; czerwony kolor: warunek niezgodny.

Wyniki były w całości zgodne z pierwszą hipotezą. Jak pokazuje Rycina 1, wyłącznie niezgodne słowa krytyczne nie poprzedzone indukcją predykcji wywołały N400 o dużej amplitudzie. W pozostałych warunkach N400 było całkowicie wyeliminowane. Podsumowując, dzięki manipulacji eksperymentalnej indukującej lub nieindukującej predykcji słowa mogłem wykazać, że amplituda N400 w całości zależy od stopnia preaktywacji danego słowa w pamięci semantycznej i że preaktywacja ta może pochodzić albo z uprzedniego kontekstu zdaniowego, albo zostać sztucznie wzbudzona eksperymentalnie. Innymi słowy, N400 jest efektem prawdziwie predykcyjnym.

Odkrycie to prowadzi do następnego pytania, dotyczącego treści predykcji: **Czy predykcje semantyczne działają na poziomie słów czy cech semantycznych** (ang. *semantic features*)? Związane z tym jest także pytanie o zasięg predykcji/preaktywacji. Jeśli przedmiotem predykcji są poszczególne słowa, można przypuszczać, że będą dotyczyć tylko jednego lub kilku najbardziej przewidywalnych słów (jeśli predykcja jest wyrażana poprzez prawdopodobieństwo wystąpienia słowa, jest matematycznie niemożliwa sytuacja, w której wiele słów będzie silnie przewidywalnych, gdyż prawdopodobieństwa muszą się sumować do 1.0). Jeśli przedmiotem predykcji są cechy semantyczne, wtedy możliwa jest równoległa facylitacja tysięcy słów—wszystkich słów, których reprezentacja konceptualna zawiera w sobie przewidywaną cechę semantyczną. Co istotne, facylitacji powinny ulegać nawet słowa niezgodne, o ile tylko ich znaczenie zawiera w sobie przewidywalne cechy semantyczne. Na przykład kontekst zdaniowy: "Na swoje następne przyjęcie zaprosił ..." nie wzbudza predykcji żadnego konkretnego słowa, ale wskazuje, że kontynuacja zdania będzie odnosić się do czegoś żyjącego, najprawdopodobniej do jakiejś osoby. Jeśli więc predykcja odbywa się na poziomie cech semantycznych, oczekivalibyśmy najsilniejszej facylitacji dla słów odnoszących się do osób, ale częściowej facylitacji także dla słów niezgodnych, które oznaczają obiekty żywotne (np. „pies”). Z pytaniem tym zmierzyłem się w kolejnym badaniu opisanym w artykule Szewczyk i Federmeier (2022).

#### 4.1.2. Badanie 2: Czy N400 jest wrażliwe na stopień niezgodności semantycznej nawet pośród nieprzewidywalnych słów? (Szewczyk i Federmeier, 2022)

Zreanalizowałem w nim dane z 5 eksperymentów wykorzystujących technikę potencjałów wywołanych EEG (ERP). Bodźcami eksperymentalnymi we wszystkich badaniach były zdania proste zapisane po angielsku (który dla osób badanych był pierwszym językiem). Pomiar amplitudy N400 dokonywany był każdorazowo na ostatnim słowie każdego zdania. Słowo to (słowo krytyczne) występowało albo w warunku zgodnym albo niezgodnym semantycznie. Jak już opisałem powyżej,

dotychczas wiadomo było, że słowa wywołują N400 proporcjonalne do stopnia ich przewidywalności. Nie wiadomo było jednak, czy amplituda N400 zależy także od stopnia **niezgodności** w przypadku słów nieprzewidywalnych. Stan taki wynikał z ograniczeń testu luk, który pozwala zmierzyć stopień przewidywalności zgodnych (przewidywalnych) słów, ale nie pozwala mierzyć stopnia niezgodności słów nieprzewidywalnych (takie słowa, z definicji, nie bywają proponowane jako wypełnienia luk). Aby obejść to ograniczenie, wykorzystałem—na tamten moment przełomowe—narzędzie z dziedziny Sztucznej Inteligencji, tj. model języka oparty o głęboką sieć neuronową (GPT-2). Głębokie sztuczne sieci neuronowe są trenowane na gigantycznych korpusach w przewidywaniu następnego słowa, w oparciu o już zaprezentowaną część zdania. Daje im to zdolności generatywne, tj. pozwala na tworzenie całych fragmentów tekstów, których stopień płynności, koherencji i realizmu sprawia, że są często nieodróżnialne od tekstu stworzonego przez człowieka (najnowszy model tego typu, ChatGPT, przebojem wdarł się do świadomości publicznej ze względu na swoje nietrywialne zdolności odpowiadania na pytania). Zaletą modeli takich jak GPT jest to, że w przeciwieństwie do testu luk, mogą być wykorzystane do estymacji dystrybucji prawdopodobieństwa (a więc przewidywalności) dla **dowolnych** słów. Dla przykładu, w Tabeli 1.1 podaję przykładu kilku zdań wraz z wyliczoną przez GPT estymatą (logarytmu) prawdopodobieństwa ostatniego słowa w zdaniu:

<b>Kontekst zdaniowy</b>	<b>Rzeczownik docelowy</b>	<b>log(p)</b>
The schools closed down because the teachers decided to go on	praying	-16.8
He really wanted to go, but decided to stay home to wait for the	timer	-12.3
Although Jan was usually careful she had forgotten to turn off the	freezers	-11.5
He was soothed by the gentle	medication	-10.7
Shuffle the cards before you	forget	-9.4
At the park next to the beach she saw a man with a	balloon	-7.7
It's hard to admit when one is	scared	-6.6
The exit was marked by a large	rock	-5.1
The teacher saw that some students were cheating and wrote down their	grades	-4.2
The surface of the water was nice and	cold	-3.4
I just had a new sound system installed in my	office	-2.6
There were a lot of old boxes stored in the	basement	-1.1

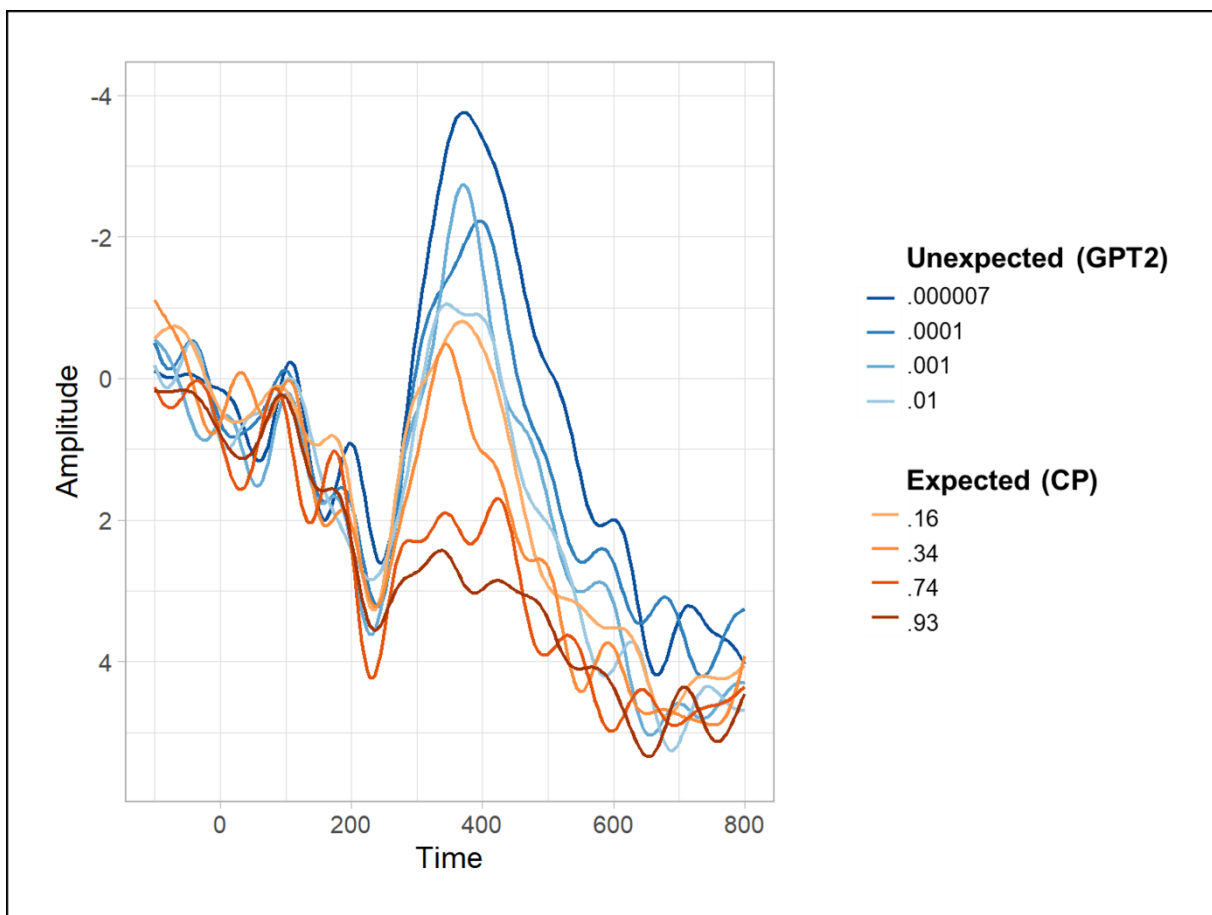
Tabela 1.1. Prawdopodobieństwa w kontekstach zdaniowych przykładowych słów docelowych, oszacowane za pomocą GPT-2. Najmniej prawdopodobne słowa docelowe znajdują się na górze tabeli.

W powyższych przykładach, słowa docelowe przybierają cały wachlarz wartości prawdopodobieństwa, od wartości zbliżonych do 100% ( $\log(p) = -1.1$  odpowiada  $p = 0.38$ ), do znikomych wartości ( $\log(p) = -16.8$  odpowiada  $p = 0.000000051$ ). Większość prawdopodobieństw jest na tyle niska, że wygodniej jest je opisywać za pomocą logarytmu prawdopodobieństwa. Pokazuje to przewagę modeli językowych nad testem luk: Uczestnicy testu luk prawdopodobnie nigdy by nie wymienili żadnego słowa o prawdopodobieństwie niższym niż 0.001 (wartości na skali logarytmicznej niższe niż -4), a nawet gdyby to zrobili, potrzeba by milionów badanych, aby opisać w ten sposób prawdopodobieństwa tysięcy różnych dokończeń zdania. Niemożliwe jest tym samym zmierzenie ich przewidywalności w teście luk.

Moje szczegółowe pytanie badawcze dotyczyło kształtu funkcji łączącej amplitudę N400 z prawdopodobieństwem słowa: czy jest ona liniowa czy logarytmiczna? Liniowy kształt funkcji byłby zgodny z hipotezą zakładającą predykcje na poziomie słów. W przypadku liniowej funkcji, tylko najbardziej prawdopodobne słowa (te o prawdopodobieństwie większym niż 0.01) mają szansę w mierzalnym stopniu wpłynąć na amplitudę N400. Z kolei różnice pomiędzy N400 wywołanym przez

słowa o prawdopodobieństwach 0.001 i 0.0001 byłyby niewykrywalne. Z kolei dla logarytmicznego związku między prawdopodobieństwem słowa a amplitudą N400, różnica w amplitudzie N400 pomiędzy słowami o prawdopodobieństwie 0.1 i 0.01 byłaby taka sama, jak różnica dla słów o prawdopodobieństwie 0.01 i 0.001. Taki kształt funkcji świadczyłby o tym, że aktywacja tysięcy słów może być w różnym stopniu facylitowana przez kontekst. Według tej hipotezy, niezgodność (i nieprzewidywalność) słowa jest stopniowa (proszę zwrócić uwagę, że wartości w kolumnie z logarytmem prawdopodobieństwa w Tabeli 1.1 są równomiernie rozłożone).

Analizy wykazały, że przewidywalność słowa i amplitudę N400 w dominującym stopniu łączy funkcja logarymiczna (Rycina 2). Dlatego *nawet wśród słów niezgodnych* występuje duża wariacja amplitudy N400 (jest ona nawet większa niż wariacja słów zgodnych, tj. tych o prawdopodobieństwie większym niż 0.01). Wspiera to interpretację, że przedmiotem mechanizmu predykcyjnego, którego przejawem jest N400, nie są poszczególne słowa, ale raczej cechy semantyczne. Jako, że te same cechy semantyczne mogą być współdzielone przez tysiące słów, predykcja kolejnych cech semantycznych powinna prowadzić do stopniowej redukcji nieprzewidywalności wielu słów jednocześnie.



Rycina 2. Zależność amplitudy efektu N400 od stopnia przewidywalności słowa w zdaniu. Osobnymi kolorami zaznaczono słowa przewidywalne, których prawdopodobieństwo jest estymowane za pomocą testu luk, i słowa nieprzewidywalne, których prawdopodobieństwo jest estymowane za pomocą modelu językowego GPT2.

Wyniki uzyskane w tym badaniu potwierdziły wnioski z mojego wcześniejszego badania (Szewczyk & Schriefers, 2013). Badanie to także posłuży jako wprowadzenie do **paradygmatu predykcji przed-**

**rzeczownikowych**, który jest wykorzystywany przez wszystkie pozostałe badania z pierwszego cyklu publikacyjnego.

#### **4.1.3. Badanie 3: Czy efekty predykcyjne zachodzą, gdy nie jest przewidywalne żadne konkretne słowo? (Szewczyk i Schriefers, 2013)**

W tym badaniu wykorzystałem istniejący już wcześniej paradygmat eksperymentalny stworzony do badania procesów predykcji—paradygmat, który roboczo nazwę **paradygmatem predykcji przed-rzeczownikowych** (DeLong i in., 2005; Van Berkum i in., 2005; Wicha i in., 2004). Zainspirowany jest on obserwacją, że jeśli predykcja rzeczownika powstaje przed jego wystąpieniem w zdaniu, powinna ona być wykrywalna również przed pojawieniem się tego rzeczownika. W celu wykrycia predykcji rzeczownika, paradygmat wykorzystuje związek gramatyczny (uzgodnienie) między rzeczownikiem a poprzedzającym go przymiotnikiem. Na przykład w wielu językach indoeuropejskich—w tym w języku polskim—sufiks przymiotnika zawiera w sobie informację o rodzaju gramatycznym uzgodnionego z nim rzeczownika (małe drzewo; mała dziewczynka). Biorąc to pod uwagę, jeśli badani silnie przewidują jakiś rzeczownik np. w rodzaju żeńskim, ale napotykają przymiotnik w rodzaju nijakim, ich system percepcji języka powinien wychwycić tę niezgodność, co z kolei powinno znaleźć swoje odbicie w zapisie EEG. Jednakże efekt ten może wystąpić wyłącznie jeśli badani przewidzieli dany rzeczownik, gdyż tylko wtedy mogą oni znać rodzaj gramatyczny jaki powinien przyjąć poprzedzający go przymiotnik. Poprzednie badania pokazały, że mózgi badanych reagują na przymiotnik o rodzaju niezgodnym z rodzajem przewidywanego rzeczownika (DeLong i in., 2005; Van Berkum i in., 2005; Wicha i in., 2004). Badania te miały jednak swoje ograniczenie. Jako zgodnych słów (rzeczowników) krytycznych, używano w nich słów o bardzo wysokim prawdopodobieństwie wystąpienia. Tak więc badania te mogły jedynie wykazać występowanie predykcji w tych rzadkich sytuacjach, gdy jedno słowo jest silnie przewidywalne. W codziennym użyciu języka silnie przewidywalne słowa stanowią mniej niż 20% słów (Luke & Christianson, 2016), a więc wnioski tych badań rozciągały się na relatywnie wąską klasę zdań.

Moje badania rozszerzyło wnioski z poprzednich badań demonstrując, że predykcja zachodzi w odniesieniu do wielu słów równolegle. Aby to pokazać, wykorzystałem unikalną cechę polskiego systemu deklinacji (występującą również w innych językach słowiańskich) — splątanie rodzaju gramatycznego i semantycznej żywotności. Ścisłej mówiąc, język polski nie wyróżnia tylko rodzajów męskiego, żeńskiego i nijakiego, ale np. w obrębie liczby pojedynczej rodzaju męskiego dokonuje dodatkowego rozróżnienia na rodzaje męskoosobowy, męskożywotny (zwany też męskozwierzęcym) i męskonieżywotny (zwany też męskorzeczowym). To rozróżnienie łatwo zademonstrować za pomocą poniższej pary fraz:

*widzę wielki stół* ("stół" jest męskonieżywotny)  
*widzę wielkiego psa* ("pies" jest męskożywotny)

Tak więc w językach takich jak polski, sufiks przymiotnika daje informację nie tylko o rodzaju gramatycznym frazy rzeczownikowej, ale także daje wskazówkę dotyczącą semantycznej żywotności rzeczownika (w bierniku, poza pewnymi wyjątkami, po przymiotniku "wielkiego" musi wystąpić rzeczownik odnoszący się do czegoś żyjącego).

Uczestnicy mojego badania czytali krótkie historyjki podczas gdy rejestrowana była aktywność elektryczna ich mózgu za pomocą EEG. Podobnie jak poprzednio, ostatnie zdanie każdej historyjki zawierało rzeczownik krytyczny. Historyjki zostały skonstruowane w ten sposób, by niezależnie od przewidywalności indywidualnych słów krytycznych, w miejscu słowa krytycznego przewidywalne

były tylko rzeczowniki w rodzaju męskonieżywotnym (i by oznaczały nieżywotne obiekty), albo w rodzajach męskożywotnym lub męskoosobowym (i oznaczały żywotne obiekty).

Manipulowałem dwiema zmiennymi: siłą zawężania (ang. *constraint strength*) i zgodnością semantyczną. Siła zawężania była wysoka, gdy tylko jeden rzeczownik stanowił prawdopodobną kontynuację zdania. Siła zawężania była niska, gdy wiele różnych rzeczowników mogło wystąpić w miejscu słowa docelowego, ale żadne z nich nie było wyjątkowo prawdopodobne. W warunku zgodnym jako słowo docelowe prezentowałem najbardziej prawdopodobne słowo. W warunku niezgodnym prezentowałem nieprawdopodobne i bezsensowne semantycznie słowo, które dodatkowo miało **odwrotną wartość żywotności** (np. gdy wszystkie prawdopodobne rzeczowniki były żywotne, niezgodny rzeczownik był semantycznie nieżywotny i w rodzaju męskonieżywotnym). Niech poniższe zdania posłużą jako przykład historyjki o niskiej sile zawężania:

*W moim domu mama ogłosiła wielką akcję przedsięwziętych porządków. Ona sama zajęła się parterem i kuchnią, zadaniem ojca było zająć się całym pierwszym piętrem. Mi zaś przypadło w udziale porządnie wysprzątać ...*

Historyjka ta może być kontynuowana przez cały szereg rzeczowników: *piwnicę, łazienkę, strych, garaż, pokój, ogród*, itd., i żadne z tych słów nie jest szczególnie prawdopodobne (przez co historyjka ma niski poziom zawężania). W moim badaniu, ta konkretna historyjka była kontynuowana jedną z dwóch fraz rzeczownikowych:

*cały strych* (warunek zgodny)

*całego wilka* (warunek niezgodny)

Potencjały wywołane EEG mierzone na przymiotniku wykazały efekt zgodności semantycznej (bardziej negatywne N400 na przymiotniki poprzedzające niezgodne rzeczowniki), replikując poprzednie badania. Co jednak najważniejsze, efekt ten występował zarówno w przypadku zdań o niskiej jak i wysokiej sile zawężania. Tym samym wykazałem, że efekty predykcyjne zachodzą również w sytuacji, w której żadne poszczególne słowo nie jest przewidywalne, a więc, że system percepcji języka musi przewidywać wiele słów równolegle. Była to pierwsza w literaturze demonstracja równoległej predykcji wielu słów. Jest to istotne odkrycie, bo pokazuje, że zjawisko predykcji odnosi się do wszystkich słów, a nie tylko tych, które są wyjątkowo silnie przewidywalne.

W późniejszych badaniach paradygmat predykcji przedręczownikowych był jeszcze wielokrotnie wykorzystywany przez innych badaczy. Większość z tych badań replikowała efekt na przymiotniku (lub rodzajniku) w językach z odpowiednio bogatą morfoskładnią. Większość z tych badań wykazywała różnicę między sufiksami zgodnymi i niezgodnymi z predykcją rzeczownika. Ale wciąż nie było wiadomo jaki konkretny mechanizm odpowiada za efekt występujący na przymiotniku. Czy wywoływany był on przez samo dostrzeżenie możliwej niezgodności, czy może aktualizacji predykcji? Było to tematem mojego kolejnego badania, opisanego w artykule Szewczyk i Wodniecka (2020).

#### **4.1.4. Badanie 4: Czy przymiotniki, za pomocą swojej końcówki, aktualizują predykcje dot. rzeczowników? (Szewczyk i Wodniecka, 2020)**

W tym badaniu weryfikowałem autorską hipotezę zakładającą, że efekt na przymiotniku to komponent N400 i jest on odbiciem dostępu leksykalnego do nowych reprezentacji semantycznych możliwych rzeczowników. Wedle tej hipotezy, sufiks przymiotnikowy daje wskazówkę dotyczącą rodzaju gramatycznego rzeczownika i ta wskazówka **zostaje natychmiast wykorzystana do aktualizacji predykcji dot. rzeczownika**. Tak więc przymiotnik o sufiksie, którego rodzaj gramatyczny jest niezgodny z rodzajem przewidywanego rzeczownika, wywołuje efekt N400 odzwierciedlający



dostęp do nowych rzeczowników, które: 1) pasują semantycznie do kontekstu i 2) mają rodzaj gramatyczny zgodny z formą przymiotnika. Taki mechanizm pociągałby za sobą dalszą konsekwencję: dostęp semantyczny do rzeczowników, co do których nastąpiła aktualizacja predykcji, powinien być łatwiejszy. Innymi słowy, większe N400 na przymiotniku (generacja/aktualizacja predykcji) powinno przełożyć się na mniejsze N400 na rzeczowniku (zaktualizowana predykcja ułatwia przetwarzanie rzeczownika). Aby przetestować tę hipotezę, manipulowałem formą całej frazy rzeczownikowej, tak by przymiotnik był mniej lub bardziej informatywny w odniesieniu do rodzaju gramatycznego następującego później rzeczownika. Ściślej mówiąc wykorzystałem inną właściwość deklinacji przymiotników w języku polskim: synkretyzmu. Dla przykładu, poniższy fragment zdania:

*Widzę wielkiego ...*

może być kontynuowany wyłącznie przez rzeczownik w rodzaju męskożywotnym lub męskoosobowym (dlatego, że w tej konstrukcji zdania dopełnienie jest w bierniku, a w bierniku sufiks –ego odpowiada wyłącznie tym dwóm rodzajom). Możemy więc powiedzieć, że sufiks -ego jest bardzo informatywny w odniesieniu do rodzaju następującego po nim rzeczownika. Jednak niewielka modyfikacja tego zdania:

*Nie widzę wielkiego ...*

sprawia, że następujący rzeczownik może obrać dowolny rodzaj gramatyczny z wyjątkiem rodzaju żeńskiego (w tej konstrukcji zdaniowej dopełnienie jest w dopełniaczu, a w dopełniaczu sufiks –ego odpowiada prawie wszystkim rodzajom gramatycznym). Tym samym przymiotnik w powyższym przykładzie nie jest zbyt informatywny w odniesieniu do rodzaju gramatycznego rzeczownika. Poprzez takie drobne modyfikacje kontekstu (wprowadzenie przeczenia lub zmianę liczby) mogłem manipulować informatywnością przymiotnika bez zmieniania ogólniejszego znaczenia całej historyjki i bez zmieniania treści słów docelowych (przymiotnika i rzeczownika).

Bazując na teście luk zmierzyłem stopień w jakim prawdopodobieństwo rzeczownika występującego w historyjce jest aktualizowane przez przymiotnik w mniej lub bardziej informatywnej wersji. Jako przykład niech posłuży poniższa mini-historyjka:

*W naszej szkole nauczyciele coraz częściej narzekali na złe zachowanie uczniów czwartych klas. Przy najbliższej okazji dyrektor wygłosił ...*

Testy luk wykazały, że jest cała lista rzeczowników stanowiących dobrą kontynuację ostatniego zdania: *przemówienie, kazanie, orędzie, pouczenie, apel, nagana, mowa, przemowa, gadka, reprimenda*, itp. Co kluczowe, tylko jeden spośród tych rzeczowników jest rodzaju niemęskożywotnego i ma on prawdopodobieństwo 0.2 („apel”). Jeśli rzeczownik "apel" zostanie poprzedzony przymiotnikiem niosącym sufiks jednoznacznie wskazujący na rodzaj niemęskożywotny (np. "długi"), prawdopodobieństwo rzeczownika "apel" wzrośnie do 1.00 (aktualizacja prawdopodobieństwa = +0.80), bo wykluczone zostaną wszystkie alternatywne kontynuacje zdania. Taki przymiotnik będzie bardzo informatywny w odniesieniu do rzeczownika. Jeśli jednak minimalnie zmienimy ostatnie zdanie historyjki poprzez wprowadzenie negacji:

*... Jednak przy najbliższej okazji dyrektor nie wygłosił ...*

przymiotnik stanie się dużo mniej informatywny, bo jego sufiks wykluczy jedynie rzeczowniki w rodzaju żeńskim, dopuszczając przy tym wiele innych rzeczowników wymienionych w teście luk („przemówienie”, „kazanie”, „orędzie”, „pouczenie”). Tym samym, prawdopodobieństwo rzeczownika "apel" wzrośnie tylko do 0.26 (aktualizacja prawdopodobieństwa = +0.06).

W badaniu tym mierzyłem N400 wywołane przez przymiotnik (słowo aktualizujące predykcje nt. rzeczownika) i przez rzeczownik (cel predykcji) by zbadać, czy występuje pomiędzy nimi przetarg. Polegałby on na tym, że silna aktualizacja predykcji na bardziej informatywnym przymiotniku wywoła duże N400, ale dzięki temu przetwarzanie następującego później rzeczownika będzie ułatwione (wywoła on mniejsze N400, niż po mniej informatywnym przymiotniku). Innymi słowy, mierzyłem, czy jest związek przyczynowo-skutkowy pomiędzy N400 na przymiotniku i rzeczowniku.

Zgodnie z moimi przewidywaniami, amplituda N400 mierzona na rzeczowniku precyzyjnie odzwierciedlała stopień aktualizacji wprowadzony przez przymiotnik: N400 dokładnie odpowiadał sumie oryginalnego (tj. bez przymiotnika) prawdopodobieństwa rzeczownika i aktualizacji prawdopodobieństwa wprowadzonej przez przymiotnik. Wynik ten wskazuje, że przymiotnik został wykorzystany do aktualizacji predykcji nt. rzeczownika. Natomiast wyniki na przymiotniku były zaskakujące. Potencjały wywołane EEG na przymiotniku nie zależały od jego informatywności. Prowadziło to do pozornie paradoksalnej sytuacji, w której przymiotnik zmieniał stopień preaktywacji rzeczownika, ale sama zmiana preaktywacji była niemożliwa do uchwycenia za pomocą EEG/ERP.

Z drugiej strony, w osobnej puli prób eksperymentalnych (historyjek) zaobserwowałem, że gdy sufiks przymiotnika był niezgodny z **jakimkolwiek** rzeczownikiem pasującym do kontekstu, wtedy na przymiotniku występował pokaźny efekt N400 (replikowało to wyniki opublikowane w poprzednio opisanym artykule, Szewczyk i Schriefers, 2013). Tym samym na przymiotniku występowała dysocjacja: Jeśli sufiks przymiotnika był zgodny z rodzajem jakiegokolwiek przewidywalnego rzeczownika, wtedy na przymiotniku brak było efektu N400, ale gdy rodzaj był niezgodny z każdym przewidywalnym rzeczownikiem, wtedy występował efekt N400. Przy ówczesnym stanie wiedzy nie byłem w stanie wyjaśnić tej dysocjacji. Zrozumiałem ją dopiero dzięki kolejnemu badaniu opublikowanemu w artykule Szewczyk, Mech i Federmeier (2021).

#### **4.1.5. Badanie 5: Czy przymiotniki, za pomocą swojego znaczenia, aktualizują predykcje dot. rzeczowników? (Szewczyk, Mech i Federmeier, 2021)**

To badanie również dotyczyło relacji między przetwarzaniem przymiotnika i rzeczownika i było przeprowadzone w języku angielskim (na Amerykanach). Skupiało się również na badaniu aktualizacji predykcji, ale w przeciwieństwie do poprzednich badań, była to aktualizacja wprowadzona za pomocą znaczenia przymiotnika, a nie uzgodnienia względem rodzaju gramatycznego.

Ciekawiło nas czy przymiotnik może zarówno zwiększyć jak i zmniejszyć preaktywację rzeczownika, w zależności od swojego dopasowania znaczeniowego do niego. Aby to zbadać, prezentowaliśmy osobom badanym krótkie zdania, przy równoczesnej rejestracji fal mózgowych za pomocą EEG. Ostatnim słowem każdego zdania był rzeczownik krytyczny. Do każdej próby eksperymentalnej przypisane były dwa potencjalne rzeczowniki krytyczne (obydwa były sensownymi i prawdopodobnymi kontynuacjami zdania), z których w danym zdaniu występował tylko jeden. Rzeczowniki były poprzedzone jednym z dwóch możliwych przymiotników. Przymiotniki były dobrane tak, by pasować do jednego, ale nie do drugiego rzeczownika. Poniżej znajduje się przykład jednego ze zdań wykorzystywanych w badaniu:

*The other driver was so angry he threatened him with a **loaded/civil gun/lawsuit**.*

(tłumaczenie: *Drugi kierowca był tak wściekły, że zagroził mu **naładowanym/cywilnym pistoletem/pozwem***).

Przed wystąpieniem przymiotnika, obydwa rzeczowniki ("gun" - pistolet i "lawsuit" - pozew) były prawdopodobne. Jednak ich prawdopodobieństwa (zmierzone testem luk) znacząco się zmieniały w

zależności od tego, którym przymiotnikiem zostały poprzedzone. Na przykład przymiotnik "loaded" (naładowany) znacząco zwiększał prawdopodobieństwo "pistoletu" kosztem prawdopodobieństwa "pozwu". Nasze pytanie badawcze dotyczyło tego, czy system percepcji języka jest w stanie wykorzystać znaczenie przymiotnika by na bieżąco zaktualizować predykcję rzeczownika, zarówno do zwiększenia jak i zmniejszenia aktywacji rzeczownika.

Również w tym badaniu wyniki okazały się zaskakujące. Przymiotniki skutecznie zmieniły amplitudę N400 na rzeczowniku i to w obydwu kierunkach, tj. zarówno zwiększając N400, gdy rzeczownik następował po niezgodnym przymiotniku (np. "naładowany pozew"), jak i zmniejszając N400 dla spójnych par (np. "cywilny pozew"). Jednak podobnie jak w poprzednim badaniu, na samym przymiotniku nie było jakichkolwiek modulacji potencjałów wywołanych, niezależnie od tego, czy przymiotnik wprowadzał pozytywną czy negatywną aktualizację prawdopodobieństwa rzeczownika i niezależnie od jej siły. Jednak co zaskakujące, amplituda N400 na przymiotniku ulegało modulacji w zależności prawdopodobieństwa samego przymiotnika w kontekście zdaniowym (prawdopodobieństwa przymiotnika praktycznie nie da się zmierzyć testem luk, użyta została zatem wspomniana wcześniej głęboka sieć neuronowa, GPT-2).

Jedynym rozsądnym wytłumaczeniem wyników obydwu badań jest to, że po napotkaniu przymiotnika następuje do niego dostęp leksykalny (na co dowodem jest modulacja N400, której wielkość zależy od prawdopodobieństwa samego przymiotnika). Natomiast wbrew naszym założeniom, znaczenie przymiotnika nie jest na bieżąco (tj. jeszcze w czasie przetwarzania przymiotnika) używane do aktualizacji predykcji następującego dalej rzeczownika (N400 na przymiotnik nie zależał od stopnia ani kierunku aktualizacji prawdopodobieństwa rzeczownika). Innymi słowy, przymiotnik jest przetwarzany jako słowo zupełnie niezależne od rzeczownika. Integracja przymiotnika z rzeczownikiem następuje dopiero w momencie, gdy ukazuje się sam rzeczownik. Dopiero tutaj przymiotniki sprawiają, że dostęp do rzeczownika staje się trudniejszy lub łatwiejszy, w zależności od zgodności między parą słów. Innymi słowy, napotkane przymiotniki zostają przechowane w semantycznej pamięci roboczej i reaktywują się w momencie napotkania rzeczownika, by współkierować procesem dostępu leksykalnego do niego.

Interpretacja ta rzuca dużo światła na wszystkie poprzednie badania wykorzystujące paradygmat predykcji przedrzeczownikowych (włączając w to 3 badania opisane powyżej). Efekt występujący na przymiotnikach i rodzajnikach przed rzeczownikiem odzwierciedla predykcje, ale nie cech semantycznych rzeczownika, a cech samego przymiotnika. Innymi słowy, obecność tego efektu świadczy o sile z jaką cechy semantyczne samego przymiotnika zostały preaktywowane przez poprzedzający kontekst. To z tego powodu N400 na przymiotniku było duże, gdy forma przymiotnika nie pasowała rodzajem do żadnego prawdopodobnego rzeczownika, np. gdy przewidywalne były same rzeczowniki nieżywotne, a przymiotnik miał sufiks w formie męskożywotnej. W takiej sytuacji cecha żywotności nie była preaktywowana przez kontekst, ale ponieważ przymiotnik ją sygnalizował, musiała ona zostać zaktywowana, prowadząc do powstania N400. Z kolei, gdy przymiotnik pasował rodzajem do przynajmniej części z przewidywalnych rzeczowników, jego cechy siłą rzeczy były już preaktywowane przez kontekst.

Na koniec trzeba zauważyć, że ogólniejszym wnioskiem moich badań jest, że sufiksy deklinacyjne nie są abstrakcyjnymi elementami gramatycznymi, ale że na równi ze rdzeniami słów biorą udział w przetwarzaniu semantycznym, co jasno wykazało badanie Szewczyk i Schriefers (2013).

#### 4.1.6. Pierwszy cykl publikacyjny: wnioski

Badania przedstawione w ramach pierwszego cyklu publikacyjnego dostarczyły nowych i istotnych informacji nt. mechanizmów przetwarzania informacji semantycznych w mózgu.

Po pierwsze, wskazały one, że podstawową jednostką w percepcji semantycznej nie są słowa, a cechy semantyczne. To dzięki temu predykcje kontekstowe są w stanie jednocześnie ułatwić dostęp do tysięcy słów. W ten sposób słowa—nawet te niezgodne—mogą być preaktywowane dokładnie w takim stopniu w jakim składowe cechy semantyczne są wspierane przez kontekst.

Po drugie, opisane powyżej badania pokazują wielorakość mechanizmów leżących u podłoża predykcji. Ogólnie rzecz biorąc, funkcja łącząca prawdopodobieństwo każdego słowa z kosztem jego przetwarzania wydaje się najbardziej zbliżona do logarytmicznej, co może być rozwiązaniem optymalnym z punktu widzenia przetwarzania informacji (Levy, 2008). Jednakże, dokładny kształt funkcji łączącej prawdopodobieństwo słowa z amplitudą N400 jest wynikiem nałożenia wielu specyficznych mechanizmów. Tak więc o ile w przeważającej części amplituda N400 jest funkcją logarytmu prawdopodobieństwa słowa, w oknie 300-400ms amplituda ta do pewnego stopnia zależy także od liniowej funkcji prawdopodobieństwa, co może odzwierciedlać możliwość, że słowa organizują cechy semantyczne w grupy i przy aktywacji odpowiedniej liczby należącej doń cech, preaktywacji ulega cała grupa odpowiadająca danemu słowu. W przypadku N400 na rzeczownik, jest ono modulowane nie tylko preaktywacją cech semantycznych rzeczownika, ale także, niezależnie, przez cechy semantyczne poprzedzających go przymiotników, które to współkierują procesem dostępu leksykalnego do rzeczownika. W ramach planowanych badań zamierzam szukać neurobiologicznego podłoża tego procesu.

Po trzecie, mimo, że komponent N400 był on przedmiotem tysięcy badań poczynając od jego odkrycia ponad 40 lat temu (Kutas & Hillyard, 1980), a efekty predykcyjne na słowach poprzedzających rzeczownik badane są od lat 15 (DeLong i in., 2005; Van Berkum i in., 2005; Wicha i in., 2004), moje badania dostarczyły mechanicznego wyjaśnienia procesów, których odbiciem są te efekty.

#### 4.1.7. Pierwszy cykl publikacyjny: cechy wyróżniające przedstawione badania

Zaprezentowane powyżej badania są godne uwagi z kilku powodów. Po pierwsze, część z nich w dużym stopniu korzysta z faktu bycia opartymi o język polski, co pozwoliło na wykorzystanie w manipulacjach eksperymentalnych właściwości gramatycznych niedostępnych w językach typowo używanych w badaniach psycholingwistycznych (angielski, holenderski, niemiecki, hiszpański, włoski), takich jak splątanie żywotności i rodzaju gramatycznego, lub synkretyzm sufiksów. Po drugie, badania te są wysoce interdyscyplinarne, łącząc narzędzia i aparaty pojęciowe psychologii, językoznawstwa, informatyki i teorii informacji. Dla przykładu, dwa badania wykorzystały nowatorskie narzędzia informatyczne (GPT, głębokie sieci neuronowe typu Transformer) by dać dostęp do parametrów słów niemożliwych do zdobycia behawioralnie i tym samym uutorować drogę do rozwiązania dotychczas nierozstrzygalnych problemów badawczych. Większość przeprowadzonych przeze mnie badań wykorzystała narzędzia z teorii informacji (Informacja, ang. *surprisal*, entropia, dywergencja Kullbacka-Leiblera), by zgrabnie opisać rozkład prawdopodobieństwa słów oraz jego zmiany. Po trzecie, badania te są też rygorystyczne metodologicznie: Są przeprowadzone na grupach wystarczająco dużych by zagwarantować odpowiednią moc statystyczną, niektóre z nich zawierają bezpośrednie replikacje, walidację krzyżową (ang. *cross-validation*; za pomocą pomiaru *expected log-predictive density*) i wykorzystują najnowocześniejsze metody statystyczne (frekwencyjne i Bayesowskie modele hierarchiczne).

#### 4.1.8. Pierwszy cykl publikacyjny: wkład w artykuły

Jestem pierwszym autorem wszystkich przedstawionych badań. Poniżej przedstawiam wkład w poszczególne badania rozpisany wg metodologii CRediT (Contributor Roles Taxonomy; <https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement>):

**Szewczyk i Schriefers (2018):** Konceptualizacja, Metodologia, Oprogramowanie, Analiza statystyczna (ang. *formal analysis*), Badania (ang. *investigation*), Zasoby (ang. *resources*), Opieka nad danymi (ang. *data curation*), Artykuł – pierwsza wersja manuskryptu (ang. *writing – first draft*), Artykuł – późniejsze redakcje i poprawki (ang. *writing – review & editing*), Wizualizacja, Administracja projektem, Pozyskiwanie funduszy.

**Szewczyk i Federmeier (2022):** Konceptualizacja, Metodologia, Oprogramowanie, Walidacja, Analiza statystyczna, Zasoby, Opieka nad danymi, Artykuł – pierwsza wersja manuskryptu, Artykuł – późniejsze redakcje i poprawki, Wizualizacja, Administracja projektem

**Szewczyk i Schriefers (2013):** Konceptualizacja, Metodologia, Oprogramowanie, Analiza statystyczna, Badania, Zasoby, Opieka nad danymi, Artykuł – pierwsza wersja manuskryptu, Artykuł – późniejsze redakcje i poprawki, Wizualizacja, Administracja projektem, Pozyskiwanie funduszy

**Szewczyk i Wodniecka (2020):** Konceptualizacja, Metodologia, Oprogramowanie, Walidacja, Analiza statystyczna, Badania, Zasoby, Opieka nad danymi, Artykuł – pierwsza wersja manuskryptu, Artykuł – późniejsze redakcje i poprawki, Wizualizacja, Administracja projektem, Pozyskiwanie funduszy

**Szewczyk, Mech i Federmeier (2021):** Konceptualizacja, Metodologia, Oprogramowanie, Analiza statystyczna, Badania, Zasoby, Opieka nad danymi, Artykuł – pierwsza wersja manuskryptu, Artykuł – późniejsze redakcje i poprawki, Wizualizacja, Administracja projektem.

Ponieważ artykuł Szewczyk i Federmeier (2022) zawiera już w swojej treści deklarację wkładu poszczególnych autorów (CRediT statement), nie dubluje tej deklaracji w osobnym oświadczeniu, a jedynie załączam skan strony ją zawierającej.

#### 4.2. Drugi cykl publikacyjny: Nauka kognatów i fałszywych kognatów w drugim języku: Rola podobieństwa znaczenia i formy w nabywaniu słów

Drugi cykl publikacyjny został przeprowadzony z moją koleżanką językoznawczynią, prof. ucz. dr hab. Agnieszką Otwinowską-Kasztelanic. Skupia się on na procesie nabywania słów w języku obcym (ang. *second language*, L2). Zadawaliśmy w nim pytanie o to czy, jak i kiedy podobieństwo między językami wspomaga uczenie się nowego słownictwa. Językoznawcy od dawna zakładali, że uczenie się drugiego języka jest łatwiejsze, gdy jest on podobny do rodzimego języka. Jak zauważył Lado (1957:2): "elementy wspólne dla dwóch języków będą łatwiejsze do nauczenia się". Uczenie się słów w dowolnym języku opiera się między innymi o dwa procesy zachodzące w słowniku mentalnym (ang. *mental lexicon*) uczącego się: 1) izolację i zapamiętanie znaczenia, czyli pojęcia oznaczanego przez słowo; 2) nauczenie się formy wyrazu. Zarówno pojęcia, jak i formy wyrazów mogą być jednakowe, podobne lub różne pomiędzy językami. Tak więc podobieństwo między językami może przybierać różne formy, w zależności od poziomu, na którym ono zachodzi. Ekwiwalenty tłumaczeniowe większości słów nie mają podobnej formy (ale w oczywisty sposób dzielą znaczenie, np. ang. "achievement" = pol. "osiągnięcie"). Są też dwa typy słów, które mają podobną formę pomiędzy językami. Są to kognaty (ang. *cognates*), czyli ekwiwalenty językowe oraz fałszywe kognaty (ang. *false cognates*), zwane też „fałszywymi przyjaciółmi” (ang. *false friends*). W przypadku kognatów, zarówno forma jak i znaczenie słów są bardzo podobne (np. ang. "atmosphere" = pol. "atmosfera"). W

przypadku fałszywych kognatów, forma jest podobna, ale nie znaczenie (np. ang. "rumour" oznaczające plotkę, jest podobne do polskiego słowa "rumor").

Jak już wspomniałem, językoznawcy badający stosowane aspekty nabywania drugiego języka (dziedzinę tę określa się jako SLA, ang. *Second Language Acquisition*) od dawna interesowali się kognatami i fałszywymi kognatami. Wielu autorów proponowało, że kognaty wspomagają naukę drugiego języka (L2), zwłaszcza, gdy uczący są świadomi ich podobieństwa do pierwszego języka (L1). Opierając się na tym założeniu, wielu badaczy SLA wykazywało, że osoby uczące się L2 często nie zauważają kognatów w tekście lub nie zwracają uwagi na formalne i znaczeniowe podobieństwo słów (Dressler i in., 2011; Nagy i in., 1993; Schmitt, 1997; Singleton, 2006), tym samym nie wykorzystując ich potencjału w nauce L2. Sugerowali, że eksplicytnie uświadamianie uczniom takiego podobieństwa może przyspieszyć naukę języka obcego. Jednak do niedawna nie przeprowadzono żadnych metodologicznie rygorystycznych badań testujących, czy kognaty (i fałszywe kognaty) faktycznie różnią się łatwością nauczenia i czy świadomość jest niezbędnym modulatorem tego procesu. Innym ograniczeniem badań SLA jest to, że mglisty termin "świadomość podobieństwa" nie został odpowiednio zoperacjonalizowany, lub jego operacjonalizacja (*Noticing Hypothesis*; Schmidt, 1990, 1992) nie była brana pod uwagę.

Psycholingwiści obrali podejście empirycznie bardziej zasadnicze. Ich badania wykazały tak zwany efekt przewagi kognatów (ang. *cognate facilitation effect*) opierający się na szybszym przetwarzaniu kognatów niż zwykłych słów, np. szybszym i poprawniejszym tłumaczeniu kognatów (de Groot, 1993; Jacobs i in., 2016), szybszych czasach reakcji na kognaty w zadaniu decyzji leksykalnej (Dijkstra i in., 2010; Lemhöfer i in., 2008; Mulder i in., 2014), ich łatwiejszym zapamiętywaniu i przywoływaniu z pamięci (Ellis & Beaton, 1993; Lotto & de Groot, 1998) oraz ich większej odporności na zapomnienie (De Groot & Keijzer, 2000).

Z kolei badania nad fałszywymi kognatami wykazały, że prezentacja fałszywego kognata prowadzi do tymczasowej aktywacji jego różnych znaczeń w obydwu językach. Na przykład, prezentacja słowa "pies" w kontekście L2 (np. gdy całe badanie jest przeprowadzane po angielsku) wzbudzi w umyśle osoby dobrze znającej polski i angielski nie tylko reprezentację semantyczną ciasta (ang. „pies” = pol. „ciasta”; Durlik i in., 2016), ale także cechy semantyczne psa, ze względu na to że angielskie „pies” jest identyczna ze słowem „pies” w j. polskim. To automatyczne wzbudzenie znaczenia słowa w L1 może doprowadzić do interferencji z jego znaczeniem w L2 i w konsekwencji utrudniać naukę poprawnego znaczenia w L2. Jednak również tutaj badań psycholingwistycznych na temat fałszywych kognatów jest niewiele i dają one sprzeczne wyniki.

Co więcej, badania psycholingwistyczne również mają swoje ograniczenia. Na przykład mierzenie czasów decyzji leksykalnych, gdy słowa są prezentowane pojedynczo i wyizolowane z kontekstu, niekoniecznie jest dobrym wskaźnikiem łatwości nauki słów w L2. A kiedy już testowane było uczenie słów w laboratoriach, było to robione w nieekologiczny sposób (w zadaniach znacząco różnych od tych, w jakich badani mają kontakt ze słowem w środowisku naturalnym lub nawet klasie szkolnej) i przy użyciu niewielkich prób bodźców.

W naszych badaniach z Agnieszką Otwinowską-Kasztelanic staraliśmy się przezwyciężyć opisane wyżej ograniczenia obydwu dziedzin. Poniżej opisuję trzy badania, które łączą ekologiczne podejście do doboru bodźców i zadań z eksperymentalnym rygiorem. O ile wiem, jest to najbardziej wszechstronna linia badań na ten temat.

#### 4.2.1. Badanie 6: Które słowa są łatwiejsze do nauczenia? (Otwinowska i Szewczyk, 2019)

W tym badaniu zastanawialiśmy się, które spośród trzech typów słów—kognat, fałszywy kognat i zwykłe słowo—ma większą szansę na to, że zostanie nabyte, zakładając, że osoba ucząca się ma równie częsty kontakt z każdym z nich. By udzielić odpowiedzi na to pytanie przeprowadziliśmy duże badanie przekrojowe. Było ono retrospekcyjne, tzn. nie testowaliśmy w nim samego procesu nabywania słów, ale mierzyliśmy, których słów badani już się nauczyli, dopasowując trzy typy słów pod względem poziomu konkretności i frekwencji w L2 (będących wyznacznikami częstości kontaktu z każdym typem słów). W ten sposób nasze pomiary integrowały wszystkie sposoby w jakie ludzie (przynajmniej w ramach testowanej populacji) typowo uczą się słów. Inaczej mówiąc, w jakikolwiek sposób słowa nie były nabywane przez naszych badanych, my mierzyliśmy, czy typ słowa wpływał na łatwość ich uczenia.

Wybraliśmy 105 słów docelowych (po 35 kognatów, fałszywych kognatów i zwykłych słów), które zostały starannie dopasowane pod względem długości, frekwencji leksykalnej w L2, kategorii gramatycznej (testowaliśmy rzeczowniki i przymiotniki) i konkretności (abstrakcyjne i konkretne), by zrównać stopień w jakim badani mieli szansę spotkać się z tymi słowami w przeszłości. Dla kognatów i fałszywych kognatów zmierzaliśmy też ich stopień podobieństwa ortograficznego używając do tego dystansu Levenshteina (ang. *Levenshtein Distance*, LD) znormalizowanego długością słów. Co również istotne, bodźce w badaniu zostały dobrane w sposób ekologiczny, czyli tak, aby osoby badane na poziomie językowym od B1 (CEFR) wzwyż miały szansę spotkać się z danym słowem w popularnych podręcznikach do angielskiego występujących w Polsce.

W tym (i w kolejnym badaniu) mierzyliśmy znajomość słów przy użyciu testów tłumaczeniowych—w przypadku tego badania, z angielskiego (L2) na polski (L1). Osoby badane ( $n = 150$ ) były proszone o niezgadywanie i o niewpisywanie odpowiedzi, jeśli nie znają słowa w L2. Podejrzewaliśmy, że niektóre osoby badane i tak mogą zgadywać, zwłaszcza w przypadku słów L2, których forma jest podobna do słów L1, co skutkowałoby zawyżeniem wyniku w teście (np. forma angielskiego słowa "horizon" przypomina polskie słowo "horyzont", prowokując badanych nie znających tego słowa do podania odgadniętego—i poprawnego—tłumaczenia). Taka strategia nie działałaby w przypadku fałszywych kognatów, gdyż ich zgadywanie prowadziłoby do niepoprawnych tłumaczeń. Ponieważ jednak fałszywe kognaty nie znane przez badanych i tak nie byłyby poprawnie przetłumaczone, strategia odgadywania wpływałaby jedynie na zawyżanie wyników znajomości kognatów, ale nie fałszywych kognatów.

Aby pozbyć się tego systematycznego zakrzywienia wyników na korzyść kognatów wprowadziliśmy poprawkę na zgadywanie. Po pierwsze, przy każdym tłumaczonym słowie poprosiliśmy badanych o określenie stopnia w jakim są pewni, że znają znaczenie słowa w L2 (na skali Likerta, od „zgaduję” do „mam całkowitą pewność”). Następnie, osobno dla każdego badanego, określiliśmy próg pewności poniżej którego było wysoce prawdopodobne, że badany zgaduje dane słowo. Próg ten został określony na podstawie pewności podawanych przy niepoprawnie „odgadniętych” tłumaczeniach fałszywych kognatów, dokonanych na podstawie ich podobieństwa do słów w L1 (tylko przy takich słowach mieliśmy pewność, że badany zgaduje) i różnicy w stosunku do pewności poprawnych tłumaczeń.

Dane zostały przeanalizowane hierarchicznymi modelami regresji logistycznej, w których oszacowywaliśmy prawdopodobieństwo, że dane słowo jest znane badanemu. Wstępne analizy wykazały, że poprawka na zgadywanie jest niezbędna. Niektórzy badani próbowali zgadywać tłumaczenia słów w L2 na podstawie ich podobieństwa formy do słów w L1. Jednakże, gdy to robili, znajdowało to odzwierciedlenie w mniejszej pewności tłumaczenia, co umożliwiło nam nałożyć

poprawkę. U niektórych badanych pewność tłumaczenia nie odróżniała odgadanych od nieodgadanych tłumaczeń i dane tych osób nie były uwzględniane w analizie.

Wyniki po nałożeniu poprawki pokazały, że w porównaniu ze zwykłymi słowami, badani mieli większe szanse znać kognaty, a mniejsze szanse znać fałszywe kognaty. Ponadto, badani mniej biegli w L2 lepiej znali kognaty, ale głównie te, które były bardzo podobne do swoich odpowiedników w L1 (np. ang. "antagonism" = pol. "antagonizm"), zaś w dużo mniejszym stopniu te mniej podobne (np. ang. "loyalty" = pol. "lojalność"). Badani bardziej biegli w L2 znali kognaty niezależnie od ich stopnia podobieństwa formy do L1. Ten wynik wskazuje, że korzystanie z podobieństwa formy może do pewnego stopnia zależeć od znajomości afiksów derywacyjnych w L2 i reguł zamiany liter na fonemy (np., ang. "nerv-ousness" = pol. "nerwow-ość", ang. "v" ≈ pol. "w", ang. „-ousness” = pol. „-ość”).

Podsumowując, mimo, że nie mierzyliśmy samego procesu nabywania słów, wykazaliśmy, że mając ustaloną szansę na napotkanie danego słowa, w porównaniu do zwykłych słów, osoby uczące się L2 będą miały większą szansę zapamiętać kognat, a mniejszą fałszywy kognat. Wniosek ten prowadzi do kolejnego pytania: Jakie sposoby uczenia się (np., intencjonalne uczenie się w klasie, uczenie incydentalne) wpływają na to, że kognaty są nabywane łatwiej, a fałszywe kognaty trudniej. Innym narzucającym się pytaniem było, czy nauczyciele uczący L2 mogą zrobić cokolwiek by uczniowie w większym stopniu wykorzystywali podobieństwo formalne słów między uczonym i znanym językiem. Skupiliśmy się na tych pytaniach w naszym kolejnym badaniu (Otwinowska, Foryś-Nogała, Kobosko i Szewczyk (2020).

#### **4.2.2. Badanie 7: uczenie się słów w klasie (Otwinowska, Foryś-Nogała, Kobosko i Szewczyk, 2020)**

W przeciwieństwie do poprzedniego badania, w tym badaniu chcieliśmy prześledzić sam proces uczenia się, by wychwycić źródło wpływu podobieństwa formy między L1 i L2 na uczenie się słów. W tym celu przeprowadziliśmy dwa quasi-eksperymenty w schemacie podłużnym. Miały one miejsce w szkołach średnich (gimnazjum), w ramach regularnych zajęć z języka angielskiego. Przez okres 6 tygodni uczniowie poznawali wybrany przez nas zestaw 90 słów kluczowych w L2 (30 kognatów, 30 fałszywych kognatów i 30 zwykłych słów, częściowo pokrywających się ze słowami użytymi w poprzednim badaniu). Słowa były wprowadzone na lekcjach, w przygotowanych przez nas ćwiczeniach odpowiednio dobranych do poziomu i wieku uczniów. Nasze ćwiczenia i zadania były wbudowane do regularnego programu nauczania w klasach uczestniczących w badaniu. Materiały dla uczniów zostały stworzone przez profesjonalnego autora podręczników do nauki angielskiego, a zarazem jedną ze współauterek badania (prof. Agnieszkę Otwinowską-Kasztelanic). Stopień, w jakim uczniowie nauczyli się słów docelowych był mierzony za pomocą testu tłumaczenia słów z polskiego na angielski (L1 na L2) przed (pre-test) i po (post-test) okresie uczenia. W sumie badanie zostało przeprowadzone na sześciu klasach prowadzonych przez 3 nauczycieli w dwóch szkołach (wszystkie klasy były na podobnym poziomie i korzystały z tego samego podręcznika). Niestety dane z jednej ze szkół zostały w całości wykluczone, gdyż nauczyciel nie zastosował się do instrukcji. Mimo to zebraliśmy dane od prawie 100 uczniów.

Oprócz manipulowania stopniem i rodzajem podobieństwa słów docelowych do znanych słów w L1, manipulowaliśmy także świadomością podobieństw międzyjęzykowych. Jak już wcześniej wspominałem, wielu językoznawców zajmujących się nabywaniem drugiego języka podejrzewa, że uczniowie korzystają z podobieństwa między językami tylko jeśli są go świadomi. Ponieważ pogląd ten nie został nigdy empirycznie zweryfikowany, postanowiliśmy to zrobić w ramach naszego badania. Podzieliliśmy badane klasy na dwie grupy: kontrolną i eksperymentalną. Obydwie grupy w ramach lekcji uczestniczyły w kilku dodatkowych warsztatach dotyczących strategii efektywnej nauki



języków (przeprowadzone przez współautorki badania). Grupa eksperymentalna w ramach swoich warsztatów otrzymała dodatkowo ćwiczenia zwiększające świadomość podobieństw między językami (trening w zauważaniu kognatów, mini-wykład o strategiach wyszukiwania znajomo wyglądających wyrazów w tekście w L2). Każdy z nauczycieli prowadził 2 grupy, eksperymentalną i kontrolną. W warsztatach manipulujących świadomością uczniów nie uczestniczyli nauczyciele prowadzący klasy, nie wiedzieli więc oni, które grupy są grupami eksperymentalnymi. Ciekawiło nas, czy zwiększenie świadomości podobieństw międzyjęzykowych jest pomocne albo konieczne w zwiększeniu korzyści w nabywaniu kognatów.

Wyniki pokazały, że nawet jeszcze przed rozpoczęciem naszej interwencji edukacyjnej, osoby badane znały więcej kognatów niż dopasowanych zwykłych słów (replikując nasze poprzednie badanie). Z drugiej strony, znały one podobną ilość fałszywych kognatów, co zwykłych słów (co nie replikowało wyników poprzedniego badania). Ponadto, jak wykazało porównanie pre- i post-testów, nasze 6-tygodniowe ćwiczenia polepszyły znajomość słów docelowych, ale przyrost w znajomości słów był jednak dla wszystkich ich typów. Okazało się także, że trening świadomości podobieństw między językami nie ma żadnego wpływu na przyrost słownictwa, w tym uczenie się kognatów, nie wspierając założeń wielu językoznawców zajmujących się nabywaniem drugiego języka.

Interpretując te wyniki pomyśleliśmy, że kognaty nie były uczone od lepiej zwykłych słów, gdyż w warunkach uczenia w klasie wszystkim typom słów uczniowie poświęcali maksymalną uwagę, co doprowadziło do sufitowego efektu uczenia. Zastanawialiśmy się, czy sytuacja nie byłaby inna w przypadku uczenia incydentalnego, tj. w sytuacjach, w których uwaga nie jest skupiona na uczeniu, a raczej na rozumieniu tekstu. W uczeniu incydentalnym nieznanne słowa są uczone mimowolnie, na podstawie kontekstu w jakim one występują. Założyliśmy, że w takich sytuacjach przewaga kognatów jest najwyraźniejsza, ze względu na łatwy dostęp do ich znaczenia (znajoma forma kognata w L2 automatycznie aktywuje jego poprawne znaczenie). Testowaliśmy tę hipotezę w kolejnym (jeszcze nieopublikowanym badaniu). Od razu jednak zaznaczę, że również to badanie nie pokazało różnic w szybkości uczenia kognatów, fałszywych kognatów i zwykłych słów.

Wobec tej sprzeczności między efektami formalnego podobieństwa słów L2 do L1 mierzonymi poprzecznie i podłużnie, postanowiliśmy popatrzeć na proces nabywania słów w większym przybliżeniu. W tym celu przeprowadziliśmy badanie laboratoryjne, opisane w artykule Marecka, Szewczyk i in. (2021).

#### **4.2.3. Badanie 8: Uczenie słów w warunkach laboratoryjnych (Marecka, Szewczyk i in., 2021)**

W przeciwieństwie do naszych poprzednich badań, tutaj poświęciliśmy trafność ekologiczną w zamian zyskując możliwość śledzenia procesu nabywania reprezentacji uczonych słów. Konceptualizując problem badawczy, założyliśmy, że minimalnie 3 rodzaje reprezentacji mentalnej uczestniczą w nabywaniu słów: forma słowa (ortograficzna lub fonologiczna), pojęcie oznaczane przez słowo, a także związek między formą a pojęciem. Założyliśmy, że osoby uczące się L2 już znają pojęcia oznaczane przez nasze słowa docelowe, tak więc trudność w nabywaniu słów ogranicza się do nabycia **formy i związku między nią a znanym pojęciem**.

Następnie skonstruowaliśmy zadanie pozwalające nam osobno mierzyć stopień w jakim badani wykorzystują znajomość 1) form słów i 2) związków między formami a pojęciami w uczeniu się słów sztucznego języka. Zadanie to było oparte o paradygmat skojarzonych par (ang. *paired-associates learning paradigm*), w którym badani zapamiętywali związek między ilustracją dobrze znanego przedmiotu (np. rysunek przedstawiający drabinę), a pseudosłowem (nowo nabywaną etykietą w

„nowym języku”). Przypomina to uczenie się z pomocą obrazków (ang. *flashcards*) w klasie, gdy nauczyciel przedstawia dane pojęcie za pomocą obrazka jednocześnie podając uczniom formę słów. W naszym badaniu pseudosłowa były wybrane tak, by symulować albo zwykłe wyrazy (np. rysunek drabiny był skojarzony z formą "rekaduf"), kognaty (np. rysunek słonia skojarzony z etykietą "słoń"), albo fałszywe kognaty (np. rysunek sukienki skojarzony z etykietą "bonan", co prowadziło do podobieństwa formalnego ze słowem banan, jednak kłóciło się ze znaczeniem na obrazku).

Zadaniem osób badanych było opanowanie 24 skojarzonych par (12 symulujących zwykłe wyrazy, 6 kognaty i 6 fałszywe kognaty). Wpierw, jednorazowo zaprezentowano badanym wszystkie pary ilustracji i wyrazów. Pozostała część procedury składała się z prezentacji bloków testujących (i uczących) naprzemiennie produkcję i rozumienie. W blokach produkcji, badanym prezentowano obrazek, a ich zadaniem było wpisać przypisaną mu etykietę (nazwę). W blokach testujących rozumienie, badanym prezentowano etykietę otoczoną czterema obrazkami, z prośbą o wskazanie obrazka skojarzonego z etykietą.

Co istotne, po każdorazowym udzieleniu odpowiedzi, badani dostawali informację zwrotną w postaci poprawnej odpowiedzi. W ten sposób każda próba eksperymentalna zarówno testowała znajomość pary obrazek-etykieta, jak i służyła jako wskazówka do dalszej nauki. Pozwalało nam to śledzić nabywanie słów wszystkich typów, z próby na próbę eksperymentalną, zarówno poprzez testowanie produkcji jak i rozumienia. W sumie każdą z 24 skojarzonych par testowaliśmy i trenowaliśmy 8 razy w bloku produkcji i 8 razy w bloku na rozumienie.

Odkryliśmy, że w tym zadaniu słowa symulujące kognaty były uczone znacznie szybciej niż pozostałe typy słów (tj. ich pełne opanowanie następowało we wcześniejszych blokach), zarówno w testach produkcji, jak i rozumienia. Ponadto, fałszywe kognaty były nabywane szybciej niż zwykłe słowa w testach produkcji, ale nie w testach na rozumienie. Wyniki te oznaczały, że w nauce kognatów badani wspierali się zarówno uprzednią znajomością podobnych form (znajomość słowa "słoń" pomogło im w nabyciu formy "słoń"), jak i uprzednią znajomością powiązania form ze znaczeniem (wiedza, że forma "słoń" odpowiada pojęciu krowy pomogła im w powiązaniu formy "słoń" z tymże pojęciem). Nieco wbrew intuicjom, okazało się, że także nauka fałszywych kognatów była łatwiejsza, ale w tym przypadku, korzyść płynęła wyłącznie ze znajomości podobnej formy (znajomość słowa "banan" pomogła badanym zapamiętać formę "bonan", mimo że ta była ona skojarzona z ilustracją sukienki).

Podsumowując, poprzez wykorzystanie nowatorskiej procedury badawczej mogliśmy osobno przetestować i oszacować szybkość uczenia dwóch aspektów znajomości słowa. Było to pierwsze badanie pokazujące, że przy uczeniu intencjonalnym badani mogą korzystać z podobieństw międzyjęzykowych, nawet w przypadku fałszywych kognatów.

#### **4.2.4. Drugi cykl publikacyjny: wnioski**

Podsumowując, przedstawione powyżej badania składają się na intrygujący obraz. Jeśli chodzi o uczenie się kognatów, niektóre nasze badania pokazały, że jeśli badany równie często jest eksponowany na zwykłe słowo i na kognata, będzie miał większą szansę i szybciej nauczy się kognata niż zwykłego słowa. Z drugiej strony, podłużne badanie w klasie nie wykazało szybszego nabywania kognatów przez młodzież szkolną (jeszcze nieopublikowane badanie skupiające się na uczeniu incydentalnym także nie stwierdziło takiej przewagi). Tylko w laboratoryjnym badaniu nad nabywaniem skojarzonych par kognaty były uczone szybciej niż zwykłe słowa.

Jak wytłumaczyć rozbieżność, w ramach której kognaty są znane lepiej, ale z drugiej strony w ekologicznych badaniach nie da się uchwycić momentu nabywania tej przewagi? Jak zawsze, ostateczna odpowiedź kryje się w przyszłych badaniach, ale na ten moment wydają się

prawdopodobne dwie niewykluczające się odpowiedzi. Według pierwszej, przewaga kognatów zarysowuje się wyłącznie w czasie nabywania leksykonu produkcyjnego (tj. leksykonu używanego przy ekspresji mowy). Słowa w dowolnym języku są przeważnie od siebie na tyle różne, że do rozróżnienia ich od siebie wystarczy wytworzenie zgrubej reprezentacji formy słowa. Ale w produkcji języka (mówienie, pisanie) sytuacja jest inna, gdyż wymaga ona znajomości dokładnej formy słowa. Być może tutaj właśnie kognaty otrzymują wsparcie ze strony swojego odpowiednika tłumaczeniowego w L1. Być może właśnie dlatego jedynym badaniem, w którym prześledziliśmy bardziej efektywne uczenie się kognatów, był eksperyment, w którym badani uczyli się kognatów przez produkcję.

Drugim wyjaśnieniem sprzeczności między konkluzjami naszych badań jest to, że badania pokazujące lepszą uprzednią znajomość kognatów korzystały z testów tłumaczenia. Być może takie testy prowadzą badanych do iluzji znajomości kognata, iluzji, która wymyka się nawet poprawce na zgadywanie (tj. widząc słowo formalnie podobne do znanego słowa w L1, ma dużo niższy próg stwierdzenia że zna to słowo i bycia tego pewnym). W badaniu testującym uczenie w klasie taka iluzja wpływałaby na wyniki zarówno pre-testów jak i post-testów, zatem nie wpływałaby na pomiar stopnia przyrostu słownictwa (który nie wykazało szybszej nauki kognatów). Wytłumaczenie to tłumaczyłoby czemu wszystkie testy tłumaczeniowe wykazały przewagę znajomości kognatów.

Wnioski dotyczące nabywania fałszywych kognatów wydają się z pozoru jeszcze bardziej sprzeczne. W pierwszym, przekrojowym badaniu, fałszywe kognaty były znane gorzej niż dopasowane słowa kontrolne. Na starcie podłużnego badania w klasie, fałszywe kognaty nie były znane gorzej od zwykłych słów (i to w obydwu eksperymentach przeprowadzonych w ramach tego badania). Nie były także wolniej ani szybciej uczone. Z kolei w badaniach nad uczeniem par skojarzonych, w porównaniu do zwykłych słów, forma fałszywych kognatów była uczona szybciej. Jak zatem grupa wyrazów może z jednej strony być gorzej znana, a z drugiej uczona szybciej? Wyjaśnienia tej rozbieżności mogą być zasadniczo takie same jak w przypadku kognatów. Szybsze uczenie fałszywych kognatów w badaniu z uczeniem par skojarzonych, ale brak takiej przewagi w klasie, może wynikać z intensywnego treningu w produkcji słów, któremu poddani zostali badani w eksperymencie z uczeniem par skojarzonych. Z kolei gorsze wyniki tłumaczenia fałszywych kognatów w badaniach przekrojowych mogą być efektem chwilowej iluzji znajomości słowa. Badani mogli zostać chwilowo i nieświadomie zmyleni międzyjęzykowym podobieństwem fałszywego kognata do znanego słowa w L1, co doprowadziło do wrażenia, że fałszywy kognat jest znany osobom badanym.

Jak widać z powyższego podsumowania, ta linia badań nie jest ukończona i wymaga kontynuacji (w szczególności, zweryfikowania czy w ekologicznych warunkach, uczenie przez produkcję faktycznie prowadzi do lepszej znajomości kognatów i fałszywych kognatów i czy później przewaga taka zaznacza się też w rozumieniu). Póki co, wstępne wnioski są następujące: 1) zwiększanie świadomości podobieństw międzyjęzykowych nie prowadzi do jakichkolwiek korzyści w szybkości uczenia słów w L2; 2) nie ma znaczącej przewagi w szybkości nabywania kognatów, albo spowolnienia w nabywaniu fałszywych kognatów w czasie uczenia przez rozumienie (np. uczeniu się polskich odpowiedników słów w L2); 3) kognaty i fałszywe kognaty mogą wspomagać naukę produkcji słów w L2, zwłaszcza u tych badanych, którzy już znają afiksy derywacyjne w L2 i reguły zamiany liter na fonemy.

#### **4.2.5. Drugi cykl publikacyjny: cechy wyróżniające przedstawione badania**

Na koniec chciałbym uwypuklić kilka aspektów powyższych badań, które na to zasługują.

Po pierwsze nasze badania pokazują konieczność empirycznej weryfikacji poglądów szeroko akceptowanych w danej dziedzinie badań. Pokazały one, że niektóre z takich poglądów (np. rola świadomości podonienstwa) mogą być po prostu nieprawdziwe.

Po drugie, wszystkie nasze badania miały wysoką moc statystyczną. Testowały większe niż typowo spotykane w literaturze grupy badanych (badanie przekrojowe  $n=150$ ; uczenie w klasie  $n=97$ , w tym 4 klasy i 2 nauczycieli; badanie z uczeniem par skojarzonych  $n=75$ ), z możliwie szerokim rozrzutem stopnia znajomości L2 i statusem społeczno-ekonomicznym (naturalnie pewnym wyjątkiem było badanie w klasie). Co równie istotne (i często pomijane w badaniach nad uczeniem słów), nasze zestawy bodźców eksperymentalnych także były duże (badanie przekrojowe  $n=105$ ; badanie w klasie  $n=90$ ; uczenie par skojarzonych  $n=24$ , gdzie typowo badania tego typu używają 6-10 bodźców).

Po trzecie, wszystkie badania wykorzystujące testy tłumaczenia mierzyły i brały poprawkę na zgadywanie, co od samego początku zidentyfikowaliśmy jako potencjalnie silne źródło zakłóceń w wynikach (ang. *confounder*). Wyróżnia to nasze badania spośród wszystkich innych badań testujących rolę podobieństw międzyjęzykowych, które zwyczajnie ignorowały ryzyko zgadywania.

Po czwarte, wszystkie badania wykorzystywały mieszane modele regresji, nowoczesne narzędzie statystyczne maksymalizujące moc statystyczną przy jednoczesnej kontroli wariancji błędu między bodźcami i między badanymi. Badanie nad uczeniem par skojarzonych wykorzystywało specjalnie dobraną funkcję prawdopodobieństwa (ang. *likelihood*; bazującej na dystrybucji gamma) w celu precyzyjnej identyfikacji bloku, w którym badani nabywali słowo.

Po piąte, przedstawiony program jest doskonałym przykładem prób mierzenia się z pytaniem badawczym z wielu różnych perspektyw, wykorzystując różne narzędzia, badając różne tryby nabywania słów i różne populacje osób badanych. Pozwoliło nam to zidentyfikować podobieństwa i różnice w wynikach poszczególnych paradygmatów badawczych, a dzięki temu w dalszej kolejności zorientować się, które aspekty wyników wynikają ze specyficznych właściwości użytych narzędzi badawczych, a które są uniwersalne i niezależne od metody.

#### 4.2.6. Drugi cykl publikacyjny: mój wkład w artykuły

Z formalnego punktu widzenia, w badaniu poprzecznym (Otwinowska & Szewczyk, 2019) współdzieliłem pozycję pierwszego autora. W artykule o uczeniu w klasie (Otwinowska i in. 2020), byłem autorem korespondencyjnym. W trzecim badaniu nad uczeniem par skojarzonych (Marecka i in. 2021), byłem drugim autorem. Podobnie jak w przypadku pierwszego cyklu publikacyjnego, wkład w poszczególne badania prezentuję posługując się taksonomią CRediT:

**Otwinowska & Szewczyk, 2019:** konceptualizacja (ang. *conceptualization*), metodologia (pewność tłumaczeń, poprawka na zgadywanie), oprogramowanie, analiza statystyczna (ang. *formal analysis*), opieka nad danymi (ang. *data curation*), artykuł – pierwsza wersja manuskryptu, artykuł – późniejsze redakcje i poprawki, wizualizacja.

Pozostałe artykuły powstały w ramach grantu, którego byłem współautorem (OPUS, 2016/21/B/HS6/01129; kierownik: Prof. Otwinowska-Kasztelanic).

**Otwinowska i in. 2020:** konceptualizacja, metodologia, oprogramowanie, analiza statystyczna, opieka nad danymi, artykuł – pierwsza wersja manuskryptu, artykuł – późniejsze redakcje i poprawki, opieka merytoryczna (ang. *supervision*), zdobywanie funduszy.

**Marecka i in. 2021:** konceptualizacja, metodologia (w tym zaproponowanie procedury naprzemiennego testowania rozumienia i produkcji uczonych par), oprogramowanie, analizy

statystyczne (analizy dotyczące identyfikacji bloku, w którym nabywane są słowa), artykuł – pierwsza wersja manuskryptu, artykuł – późniejsze redakcje i poprawki, wizualizacja, zdobywanie funduszy.

Ponieważ artykuł Marecka i in. (2021) zawiera już w swojej treści deklarację wkładu poszczególnych autorów (CrediT statement), nie dubluje tej deklaracji w osobnych oświadczeniach współautorów, a jedynie załączam skan strony ją zawierającej.

## **5. Informacja o wykazywaniu się istotną aktywnością naukową realizowaną w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury, w szczególności zagranicznej.**

### **5.1. Artykuły opublikowane w czasopismach naukowych i narzędzia inne niż zgłoszone jako osiągnięcie naukowe.**

Zgodnie z tym co wykazałem w artykule II wykazu osiągnięć naukowych, mój wkład do literatury nie ogranicza się do artykułów zgłoszonych jako osiągnięcie naukowe i zawiera w sobie 32 inne publikacje i narzędzia. Krótko przedstawię dwie grupy publikacji spośród nich.

Pierwsza grupa skupia się na przetwarzaniu pseudosłów, tj. słowo-podobnych ciągów liter, które spełniają reguły fonotaktyczne danego języka, ale nie odpowiadają żadnej jednostce leksykalnej. Pseudosłowa są często wykorzystywane do testowania zdolności fonologicznych osób badanych (najczęściej dzieci). Dają one możliwość badania przetwarzania fonologicznego bez opierania się o eksplicytną wiedzę leksykalną. Jednym z typowych zadań wykorzystujących pseudosłowa jest zadanie powtarzania pseudosłów, w którym badani są proszeni o natychmiastowe powtórzenie pseudosłów prezentowanych im słuchowo. Test ten został stworzony jako narzędzie do pomiaru pojemności pamięci fonologicznej (Gathercole & Baddeley, 1989), a ogólniej są w miarę czuły i swoisty wskaźnikami zaburzeń rozwoju fonologicznego (takiego jak specyficzne zaburzenie językowe oraz niektóre formy dysleksji). W **artykule Szewczyk, Marecka, Chiat i Wodniecka (2018)** zaproponowałem nową metodę pomiaru językopodobności pseudosłów: średnią frekwencję wszystkich ngramów. Na przykład, pseudosłowo „ket” można rozbić na szereg ngramów: 4 bigramy („#k”, „ke”, „et”, „t#”), 3 trigramy („#ke”, „ket”, „et#”), itd., gdzie „#” jest markerem nagłosu i wygłosu pseudosłowa. Następnie, na podstawie korpusu języka możemy oszacować częstotliwość wszystkich ngramów i uśrednić ich logarytmy. Uzyskana wartość daje nam wskaźnik nasycenia pseudosłowa zbitkami liter (lub dźwięków) charakterystycznych dla danego języka. Wykazałem, że oprócz długości pseudosłowa, ten wskaźnik to najlepszy predyktor szansy, że dziecko będzie w stanie powtórzyć dane pseudosłowo (pseudosłowa składające się z rzadszych ngramów są dużo trudniejsze do powtórzenia). Tym samym moja miara zdetronizowała inny wskaźnik, dotychczas uznawany w literaturze jako najbardziej trafny, oparty o gęstość sąsiedztwa leksykalnego. Na tej podstawie wywnioskowałem, że głównym poznawczym wyznacznikiem poprawności w teście nie jest—jak przez długi czas sądzono—pojemność pamięci fonologicznej, ale raczej zasób słownika subleksykalnego w danym języku. Zaproponowałem, że zasób ten determinuje, które złożone zbitki dźwięków dana osoba jest w stanie efektywnie rozpoznać, przechować w pamięci i wyartykułować. Wskaźnik ten został także wykorzystany do parametryzacji angielskich pseudosłów w badaniu przeprowadzonym ze współpracownikami z University of Manchester i University of London (**Polišenská i in., 2021**), a także polsko- i angielsko-podobnych pseudosłów w badaniu nad uczeniem słów (**Marecka i in., 2018**). Był także wykorzystany w dużym badaniu z wykorzystaniem funkcjonalnego rezonansu magnetycznego nad źródłami dysleksji (**Łuniewska i in., 2018**). Statystyki przeprowadzone na ngramach pozwalają także na generowanie nowych pseudosłów o określonych parametrach (manipulowanie stopniem podobieństwa do danego języka, na różnych poziomach ogólności). Sprawilo to, że moja metoda leżała u podstaw najnowszego polskiego wystandaryzowanego i znormalizowanego testu

powtarzania pseudostów (**TPP; Szewczyk i in., 2015**), towarzyszącemu większej baterii testów służących do diagnozy językowej dzieci w wieku 4-9 lat (Test Rozwoju Językowego; Smoczyńska i in., 2015).

Kolejna grupa badań wywodzi się z mojej współpracy z Laboratorium Psychologii Języka i Dwujęzyczności prowadzonego przez prof. Zofię Wodniecką. Badania te skupiają się na dwujęzyczności i jej związku z funkcjami wykonawczymi. W sumie, badanie te pokazują ograniczenia śmiałych hipotez zakładających uogólniony i dwukierunkowy związek między treningiem dwujęzyczności i przyrostem potencjału funkcji wykonawczych. Mimo, że jakaś forma kontroli poznawczej musi być zaangażowana w proces separacji, selekcji i równoważenia aktywacji dwóch języków w umyśle osoby dwujęzycznej (**Casado i in., 2022; Durlík i in., 2016; Wodniecka i in., 2020**), wydaje się mało prawdopodobne, by ta forma kontroli poznawczej była współdzielona z innymi niejęzykowymi zadaniami, które nominalnie również zależą od kontroli poznawczej. Problem ten wydaje się mieć fundamentalną naturę, gdyż pomiary tej samej funkcji wykonawczej przeprowadzone przy użyciu różnych narzędzi korelują w znikomym stopniu (**Kałamata i in., 2020**), zadania uznawane za mierzące hamowanie mogą w istocie mierzyć selekcję uwagową (**Kałamata, Szewczyk, i in., 2018**), a same pomiary mogą być obarczone szumem, którego istotnym komponentem jest strategia przetwarzania kontrolowanego (**Kałamata, Drożdżowicz, i in., 2018**). Nawet sam pomiar stopnia dwujęzyczności nie jest prosty, ze względu na brak jasności, które potencjalne indeksy dwujęzyczności (np. wiek nabycia drugiego języka, jego bierna i czynna znajomość, wzorce użycia w życiu codziennym) mają przełożenie na system poznawczy (**Kałamata i in., 2022**). Z drugiej strony, udało nam się lepiej zrozumieć sieci kontroli poznawczej, które są zaangażowane w przetwarzanie dwóch języków, a także wyodrębnić behawioralną miarę zaangażowania kontroli poznawczej, która na tle literatury wyróżnia się ekologicznością i nie mieszaniami wielu różnych typów kontroli (**Casado i in., 2022; Wodniecka i in., 2020**). Na koniec wspomnę, że ta linia badań dostarczała i wciąż dostarcza precyzyjnych modeli przetwarzania dwujęzycznego, które są dobrze zdefiniowane i mocno osadzone w pomiarach behawioralnych, czego wciąż brakuje w literaturze nt. neuronalnych i poznawczych podstaw dwujęzyczności.

## 5.2. Realizacja projektów naukowych

W czasie studiów doktoranckich uzyskałem dwa ogólnokrajowe granty (z MNiSW i z NCN), w tym grant OPUS dla doświadczonych naukowców (zajmując drugą lokatę na liście rankingowej). Projekty te skupiały się na badaniu predykcji w percepcji języka przy użyciu eksperymentów EEG i bodźców wykorzystujących właściwości języka polskiego. Oprócz umożliwienia mi realizacji tych badań, projekty te także dały mi początkowe doświadczenie w kierowaniu i zarządzaniu projektami naukowymi. Kilka lat później otrzymałem finansowanie w ramach prestiżowego grantu Marie Curie Individual Fellowship (wersja globalna), które pozwoliło mi na 3-letnie studia podoktorskie na Uniwersytecie Illinois (USA) i w Instytucie Dondersa (Holandia). W czasie mojej kariery byłem współautorem, konsultantem i postdokiem w sześciu innych projektach naukowych (finansowanych przez NCN, NPRH i FNP). W szczególności, przez dwa lata byłem zatrudniony jako postdok w granicie Sonata Bis kierowanym przez dr hab. Zofię Wodniecką. Uczestniczyłem także w dwóch projektach o bardziej aplikacyjnym charakterze. Pierwszym z nich był projekt COST ISO804 "Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment", którego celem było rozwinięcie narzędzi do diagnozy specyficznego zaburzenia językowego (SLI). Oprócz realizacji zadań badawczych, w projekcie tym służyłem w roli zastępcy członka grupy sterującej. Uczestnictwo w tym projekcie dało mi pierwsze doświadczenie we współpracy międzynarodowej w wielowątkowym projekcie naukowym. Zapoczątkowało też moją współpracę z badaczami SLI z Londynu i Manchesteru (prof. Chiat, dr Polišenská), a także z psycholingwistami i językoznawcami z Warszawy (prof. Haman,

prof. Otwinowska-Kasztelanic), co zaowocowało wieloma wspólnymi badaniami. Zaowocowało to także zaproszeniem do innego dużego projektu finansowanego przez Komitet Badań Edukacyjnych (kierownik: dr Magdalena Smoczyńska), skupiającym się na stworzeniu polskich narzędzi do diagnozy SLI i do ćwiczenia zdolności językowych dzieci potencjalnie dotkniętych tym zaburzeniem.

### **5.3. Staże i pobyty naukowe zrealizowane w zagranicznych instytucjach naukowych**

W ramach realizacji doktoratu wielokrotnie udawałem się na krótkie wyjazdy do Nijmegen w Holandii, gdzie znajdują się aż dwie szanowane instytucje badawcze w dużej mierze skupiające się na badaniach psycholingwistycznych: Donders Centre for Brain, Cognition and Behaviour i Max Planck Institute for Psycholinguistics. Prawdopodobnie sprawia to, że mimo bycia względnie niedużym miastem, Nijmegen skupia w jednym miejscu największą liczbę psycholingwistów i neurolingwistów na świecie. Mój doktorat miał dość nietypową strukturę. Wszystkie badania były przeprowadzane w Polsce (co pozwoliło mi, z dużą korzyścią dla badań, na wykorzystanie języka polskiego), ale opiekę naukową sprawował nad nim naukowiec osadzony w Nijmegen (prof. Herbert Schriefers). Ten epizod kariery zakończył się w 2016, kiedy to obroniłem rozprawę doktorską na Uniwersytecie Radboud w Nijmegen. W międzyczasie, odbyłem 3 wizyty w Londynie w ramach współpracy z prof. Shulą Chiat. Jak już wspominałem, później uzyskałem 3-letnie stypendium Marie Curie, w ramach którego przez 2 lata przebywałem w USA (współpraca z prof. Karą Federmeier, światowym ekspertem w wykorzystaniu EEG/ERP w badaniach nad przetwarzaniem języka), a później rok w Instytucie Dondersa w Nijmegen. Tam współpracowałem z prof. Peterem Hagoortem (ekspertem w badaniach nad neuronalnym podłożem przetwarzania języka) i z prof. Florisem de Lange (ekspertem w badaniach nad predykcjami i przetwarzaniem wzrokowym). Po ukończeniu stypendium Marie Curie zaproponowano mi przedłużenie współpracy z prof. Hagoortem i prof. de Lange w ramach zatrudnienia w Instytucie Maxa Plancka. W sumie (nie licząc doktoratu) mój pobyt w Nijmegen trwał 2,5 roku.

### **5.4. Członkostwo w komitetach redakcyjnych, krajowych i międzynarodowych organizacjach i towarzystwach naukowych**

Jestem członkiem Society of Neurobiology of Language, Society for Psychophysiological Research (SPR), Cognitive Neuroscience Society, Society for Human Sentence Processing (HSP), a także European Society for Cognitive Psychology. W zeszłym roku byłem członkiem rady naukowej konferencji SPR. Od roku też recenzuje zgłoszenia na konferencję HSP. Jestem też członkiem komitetu redakcyjnego czasopisma *Glossa: Psycholinguistics*. Ponadto, okazjonalnie piszę recenzje dla wielu innych międzynarodowych czasopism (do tej pory pisałem recenzje dla 22 czasopism).

### **5.5. Informacja o osiągnięciach dydaktycznych**

Prowadziłem szereg kursów dla studentów psychologii, głównie dotyczących anatomii mózgu, funkcjonalnej anatomii mózgu i psycholingwistyki. Były to następujące kursy: Biopsychologia 1; Biopsychologia 2; Psychofizjologia; Leworęczność, płęć i dysfunkcje mózgowie a asymetria półkulowa; Psychologia języka; Psychologia procesów poznawczych; Wprowadzenie do analizy potencjałów wywołanych przy użyciu EEG/ERP. Niedawno prowadziłem też kurs dla uczestników programu anglojęzycznych studiów doktorskich COGNES zatytułowany „Data Science in R”.

Oprócz regularnych kursów na uczelni, prowadziłem warsztat z mechanizmów predykcji w przetwarzaniu języka na językoznawczej szkole letniej w Dacicach (Republika Czeska). Niedawno prowadziłem też jeden moduł na Neuroimaging Toolkit, tygodniowym kursie z neuroobrazowania odbywającym się w Instytucie Dondersa.

## 5.6. Informacja o osiągnięciach organizacyjnych

Jako student studiów magisterskich, założyłem sekcję koła naukowego skupioną na neuronaukach („Sekcja mózgu”). Oprócz regularnych spotkań w trakcie których odbywały się prezentacje naukowe, zapraszaliśmy gości i zorganizowaliśmy dwa wyjazdy naukowe, wraz z członkami zaprzyjaźnionego koła naukowego działającego przy Uniwersytecie w Greifswaldzie.

Jako student studiów doktorskich, wyszedłem z inicjatywą stworzenia, zaprojektowałem, a następnie stworzyłem system informatyczny pośredniczący w rekrutacji studentów na badania psychologiczne odbywające się w Instytucie Psychologii UJ. Do systemu corocznie zapisywało się kilkaset studentów z całego uniwersytetu i był on podstawowym źródłem osób badanych przez wiele lat.

Stworzyłem także platformę informatyczną do przechowywania, selektywnego udostępniania i kodowania wyników ogromnego programu badawczego odbywającego się w ramach projektu COST ISO804. Na bieżąco korzystało z niego około 50 badaczy zaangażowanych w projekt: osoby zbierające dane, osoby transkrybujące oraz badacze przeprowadzający analizy. System był aktywnie wykorzystywany przez 5 lat do zbierania surowych danych, wyników i notatek z badań. Były na nim przechowywane dane z ponad 797 badanych testowanych 14 narzędziami.

.....

(podpis wnioskodawcy)



## Literatura cytowana

- Casado, A., **Szewczyk, J.**, Wolna, A., & Wodniecka, Z. (2022). The relative balance between languages predicts the degree of engagement of global language control. *Cognition*, 226, 105169. <https://doi.org/10.1016/j.cognition.2022.105169>
- de Groot, A. M. B. (1993). Word-type Effects in Bilingual Processing Tasks. In B. Weltens & R. Schreuder (Eds.), *The Bilingual Lexicon* (pp. 1–315). John Benjamins Publishing Company.
- De Groot, A. M. B., & Keijzer, R. (2000). What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting. *Language Learning*, 50(1), 1–56. <https://doi.org/10.1111/0023-8333.00110>
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121. <https://doi.org/10.1038/nn1504>
- Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, 62(3), 284–301. <https://doi.org/10.1016/j.jml.2009.12.003>
- Dressler, C., Carlo, M. S., Snow, C. E., August, D., & White, C. E. (2011). Spanish-speaking students' use of cognate knowledge to infer the meaning of English words\*. *Bilingualism: Language and Cognition*, 14(2), 243–255. <https://doi.org/10.1017/S1366728910000519>
- Durlik, J., **Szewczyk, J.**, Muszyński, M., & Wodniecka, Z. (2016). Interference and Inhibition in Bilingual Language Comprehension: Evidence from Polish-English Interlingual Homographs. *PLOS ONE*, 11(3), e0151430. <https://doi.org/10.1371/journal.pone.0151430>
- Ellis, N., & Beaton, A. (1993). Factors Affecting the Learning of Foreign Language Vocabulary: Imagery Keyword Mediators and Phonological Short-Term Memory. *The Quarterly Journal of Experimental Psychology Section A*, 46(3), 533–558. <https://doi.org/10.1080/14640749308401062>
- Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28(2), 200–213. [https://doi.org/10.1016/0749-596X\(89\)90044-2](https://doi.org/10.1016/0749-596X(89)90044-2)
- Jacobs, A., Fricke, M., & Kroll, J. F. (2016). Cross-Language Activation Begins During Speech Planning and Extends Into Second Language Speech. *Language Learning*, 66(2), 324–353. <https://doi.org/10.1111/lang.12148>
- Kałamala, P., Chuderski, A., **Szewczyk, J.**, Senderecka, M., & Wodniecka, Z. (2022). Bilingualism caught in a net: A new approach to understanding the complexity of bilingual experience. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0001263>
- Kałamala, P., Drożdżowicz, A., **Szewczyk, J.**, Marzecová, A., & Wodniecka, Z. (2018). Task strategy may contribute to performance differences between monolinguals and bilinguals in cognitive control tasks: ERP evidence. *Journal of Neurolinguistics*, 46, 78–92. <https://doi.org/10.1016/j.jneuroling.2017.12.013>
- Kałamala, P., **Szewczyk, J.**, Chuderski, A., Senderecka, M., & Wodniecka, Z. (2020). Patterns of bilingual language use and response inhibition: A test of the adaptive control hypothesis. *Cognition*, 204, 104373. <https://doi.org/10.1016/j.cognition.2020.104373>

- Kałamała, P., **Szewczyk, J.**, Senderecka, M., & Wodniecka, Z. (2018). Flanker task with equiprobable congruent and incongruent conditions does not elicit the conflict N2. *Psychophysiology*, *55*(2), e12980. <https://doi.org/10.1111/psyp.12980>
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. <https://doi.org/10.1126/science.7350657>
- Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, *11*(5), 539–550. <https://doi.org/10.3758/BF03196991>
- Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press.
- Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 12–31. <https://doi.org/10.1037/0278-7393.34.1.12>
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. <https://doi.org/10.1016/j.cognition.2007.05.006>
- Lotto, L., & de Groot, A. M. B. (1998). Effects of Learning Method and Word Type on Acquiring Vocabulary in an Unfamiliar Language. *Language Learning*, *48*(1), 31–69. <https://doi.org/10.1111/1467-9922.00032>
- Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88*, 22–60. <https://doi.org/10.1016/j.cogpsych.2016.06.002>
- Łuniewska, M., Chyl, K., Dębska, A., Kacprzak, A., Plewko, J., Szczerbiński, M., **Szewczyk, J.**, Grabowska, A., & Jednoróg, K. (2018). Neither action nor phonological video games make dyslexic children read better. *Scientific Reports*, *8*(1), 549. <https://doi.org/10.1038/s41598-017-18878-7>
- Marecka, M., **Szewczyk, J.**, Jelec, A., Janiszewska, D., Rataj, K., & Dziubalska-Kořaczyk, K. (2018). Different phonological mechanisms facilitate vocabulary learning at early and late stages of language acquisition: Evidence from Polish 9-year-olds learning English. *Applied Psycholinguistics*, *39*(1), 1–35. <https://doi.org/10.1017/S0142716417000455>
- Marecka, M., **Szewczyk, J.**, Otwinowska, A., Durlík, J., Foryś-Nogala, M., Kutylowska, K., & Wodniecka, Z. (2021). False friends or real friends? False cognates show advantage in word form learning. *Cognition*, *206*, 104477. <https://doi.org/10.1016/j.cognition.2020.104477>
- Mulder, K., Dijkstra, T., Schreuder, R., & Baayen, H. R. (2014). Effects of primary and secondary morphological family size in monolingual and bilingual word processing. *Journal of Memory and Language*, *72*, 59–84. <https://doi.org/10.1016/j.jml.2013.12.004>
- Nagy, W. E., García, G. E., Durgunođlu, A. Y., & Hancin-Bhatt, B. (1993). Spanish-English Bilingual Students' Use of Cognates in English Reading. *Journal of Reading Behavior*, *25*(3), 241–259. <https://doi.org/10.1080/10862969009547816>
- Otwinowska, A., Foryś-Nogala, M., Kobosko, W., & **Szewczyk, J.** (2020). Learning Orthographic Cognates and Non-Cognates in the Classroom: Does Awareness of Cross-Linguistic Similarity Matter? *Language Learning*, *70*(3), 685–731. <https://doi.org/10.1111/lang.12390>

- Polišenská, K., Chiat, S., **Szewczyk, J.**, & Twomey, K. E. (2021). Effects of semantic plausibility, syntactic complexity and n-gram frequency on children's sentence repetition. *Journal of Child Language*, 48(2), 261–284. <https://doi.org/10.1017/S0305000920000306>
- Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning<sup>1</sup>. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Schmidt, R. W. (1992). Awareness and Second Language Acquisition. *Annual Review of Applied Linguistics*, 13, 206–226. <https://doi.org/10.1017/S0267190500002476>
- Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 199–227). Cambridge University Press.
- Singleton, D. (2006). Lexical transfer: Interlexical or intralexical. In J. Arabski (Ed.), *Cross-linguistic influences in the second language lexicon* (Vol. 17, pp. 130–143). UK: Multilingual Matters.
- Smoczyńska, M., Haman, E., Maryniak, A., Czaplewska, E., Krajewski, G., Banasik, N., kochańska, M., & Łuniewska, M. (2015). *Test Rozwoju Językowego*.
- Szewczyk, J. M.**, & Federmeier, K. D. (2022). Context-Based Facilitation of Semantic Access Follows Both Logarithmic and Linear Functions of Stimulus Probability. *Journal of Memory and Language*, 123, 104311. <https://doi.org/10.1016/j.jml.2021.104311>
- Szewczyk, J. M.**, Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, 179, 23–36. <https://doi.org/10.1016/j.cognition.2018.06.002>
- Szewczyk, J. M.**, Mech, E. N., & Federmeier, K. D. (2021). The power of “good”: Can adjectives rapidly decrease as well as increase the availability of the upcoming noun? *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, 48(6), 856–875. <https://doi.org/10.1037/xlm0001091>
- Szewczyk, J. M.**, & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language*, 68(4), 297–314. <https://doi.org/10.1016/j.jml.2012.12.002>
- Szewczyk, J. M.**, & Schriefers, H. (2018). The N400 as an index of lexical preactivation and its implications for prediction in language comprehension. *Language, Cognition and Neuroscience*, 33(6), 665–686. <http://dx.doi.org/10.1080/23273798.2017.1401101>
- Szewczyk, J. M.**, Smoczyńska, M., Haman, E., Łuniewska, M., kochańska, M., & Załupska, J. (2015). *Test Powtarzania Pseudosłów TPP*. Instytut Badań Edukacyjnych.
- Szewczyk, J. M.**, & Wodniecka, Z. (2020). The mechanisms of prediction updating that impact the processing of upcoming word: An event-related potential study on sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(9), 1714–1734. <https://doi.org/10.1037/xlm0000835>
- Taylor, W. L. (1953). “Cloze Procedure”: A New Tool for Measuring Readability. *Journalism Quarterly*, 30(4), 415–433. <https://doi.org/10.1177/107769905303000401>
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467. <https://doi.org/10.1037/0278-7393.31.3.443>

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, *16*(7), 1272–1288.

Wodniecka, Z., **Szewczyk, J.**, Kałamała, P., Mandera, P., & Durlik, J. (2020). When a second language hits a native language. What ERPs (do and do not) tell us about language retrieval difficulty in bilingual language production. *Neuropsychologia*, *141*, 107390.  
<https://doi.org/10.1016/j.neuropsychologia.2020.107390>