

Marzena Fornal  <https://orcid.org/0000-0002-0263-9081>
Akademia Humanistyczno-Ekonomiczna w Łodzi
e-mail: mfornal@ahelodz.pl

Teoria przetwarzania predykcyjnego a problem ogólnych pojęć abstrakcyjnych w perspektywie językoznawstwa kognitywnego

Streszczenie

Współczesny paradygmat ucieleśnienia w językoznawstwie kognitywnym stanowi cenne ramy pojęciowe dla wyjaśnienia ugruntowania pojęć konkretnych, lecz napotyka zasadnicze trudności w wyjaśnianiu mechanizmu tworzenia pojęć abstrakcyjnych (tak zwany problem odcieleśnienia pojęć). Coraz częściej wskazuje się, że rozwiązanie tej trudności leży w połączeniu paradygmatu ucieleśnienia z teorią przetwarzania predykcyjnego. Chociaż teoria ta aspiruje do bycia ogólną teorią mózgu w naukach kognitywnych, ma pewne ograniczenia, chociażby dotyczące wyjaśnienia istotnych cech ogólnych pojęć abstrakcyjnych. Artykuł analizuje teorię przetwarzania predykcyjnego pod kątem jej zdolności do wyjaśnienia kompozycyjności, produktywności, systematyczności i ogólności myślenia pojęciowego. Pomimo wskazanych w artykule ograniczeń teoria przetwarzania predykcyjnego w połączeniu z paradygmatem ucieleśnionego języka stanowi obiecującą propozycję w ramach kognitywizmu drugiej generacji.

Słowa kluczowe: przetwarzanie predykcyjne, ucieleśnione poznanie, ucieleśniony język, problem pojęć abstrakcyjnych, Clark, Hohwy

Wprowadzenie

Popularny współcześnie paradygmat ucieleśnienia w językoznawstwie kognitywnym, choć dobrze wyjaśnia kwestię ugruntowania pojęć konkretnych, nie radzi sobie ze wskazaniem mechanizmu wyłaniania się pojęć abstrakcyjnych (tak zwany problem odcieleśnienia pojęć, na który wskazuje między innymi Guy Dove, 2011; 2014; 2016; 2018; por. też Fornal,

2022). Obecnie coraz częściej paradygmat ucieleśnionego poznania łączony jest z teorią przetwarzania predykcyjnego. Teoria ta, jak wskazują niektórzy, aspiruje do miana ogólnej teorii mózgu w naukach kognitywnych (Hohwy, 2013; Wiese, Metzinger, 2017). Wydaje się, że z uwagi na jej komputacyjny charakter oraz ogólność, nie tylko będzie dobrze pasowała do pewnego ideału takiej teorii, ale też może przybliżyć nas do wyjaśnienia, w jaki sposób mózg generuje pojęcia ogólne o wysokim poziomie abstrakcji. Jak jednak postaram się wykazać w niniejszym artykule, teoria ta ma również swoje ograniczenia i na ten moment w jej ramach nie jesteśmy w stanie wyjaśnić niektórych, konstytutywnych cech ogólnych pojęć abstrakcyjnych. Niemniej jednak teoria ucieleśnionego języka uzupełniona o teorię przetwarzania predykcyjnego stanowi obiecującą propozycję w ramach kognitywizmu drugiej generacji w językoznaństwie.

1. Teoria przetwarzania predykcyjnego – wprowadzenie

Teoria przetwarzania predykcyjnego oparta jest na hipotezie mózgu bayesowskiego, zgodnie z którą:

1. Racjonalność definiowana jest jako zdolność rozumowania w sytuacji niepewności.
2. Wnioskowanie, jakiego w naturalnych warunkach dokonuje ludzki system poznawczy, prawie nigdy nie jest dedukcyjne. Ludzkie rozumowanie najlepiej postrzegać jako rozwiązywanie problemów probabilistycznych, a nie logicznych.
3. (Rzekomo) logiczne problemy z rozumowaniem, ujawniające pozornie irracjonalne zachowania (czyli tak zwane błędy poznawcze), są lepiej rozumiane z probabilistycznego punktu widzenia. Błędy poznawcze nie są irracjonalne. Ludzie przyjmują niepoprawne wnioski, ponieważ ich przesłanki są prawdopodobne.
4. Jednym z ważnych zadań, jakie wykonuje nasz mózg, jest wydobywanie sygnału z szumu.
5. Możliwe jest, że inferencja bayesowska jest implementowana w układzie nerwowym, a ludzki system poznawczy zachowuje się bayesowsko zarówno w swoich nieświadomych wnioskowaniach dotyczących spostrzegania i pamięci, jak i w swoich świadomych rozważaniach, takich jak ocena teorii.
6. Nie jest wykluczone, że „bayesowskie” są nie tylko mózgi ludzi, ale i innych ssaków – odejście od antropocentryzmu.

Z powyższych punktów wyraźnie wynika, że takie podejście całkowicie zrywa z klasycznym modelem racjonalności. Niektórzy wskazują nawet, że możemy mówić tu o „zwrócenie probabilistycznym” w kognitywistyce i sztucznej inteligencji dla zrozumienia ludzkiej racjonalności (Kwisthout, van Rooij, 2019).

Wspomnianą wyżej zmianę rozumienia ludzkiej racjonalności zawdzięczamy między innymi pracom Thomasa Bayesa, na których wyrosła aksjomatyka prawdopodobieństwa XX wieku. Jak zauważył Bayes, między przesłankami a wnioskiem, czy też między wynikiem eksperymentu lub obserwacji a hipotezą, zachodzi w wielu przypadkach relacja inferencji, która może być w zadowalającym stopniu obiektywnie mierzona przy pomocy funkcji prawdopodobieństwa. Co za tym idzie, wyjaśnienia (szczególnie w naukach empirycznych)

powinny mieć charakter probabilistyczny, a nie przyczynowy (ze względu na to, że odkrywane prawidłowości są tylko probabilistyczne). Kluczową rolę odgrywa tu konfirmacja hipotez i teorii w oparciu o świadectwa empiryczne. Opis mechanizmu modyfikowania prawdopodobieństwa hipotezy polega na systematycznym stosowaniu twierdzenia Bayesa (Kawalec, 2003). Twierdzenie Bayesa jest wyeksponowanym elementem omawianej w tym punkcie teorii, gdyż właśnie ono określa: jaki wpływ na prawdopodobieństwo hipotezy **H** mają dane **E**. Prawdopodobieństwo hipotezy **H** w świetle danych **E** odpowiada prawdopodobieństwu danych **E** przy założeniu hipotezy **H**, pomnożonemu przez dotychczasowe prawdopodobieństwo hipotezy **H** i podzielonemu przez prawdopodobieństwo danych **E**. Zgodnie z bayesianizmem racjonalny agent określa na wyjściu prawdopodobieństwo hipotezy w sposób dowolny (prawdopodobieństwo a priori), byle zgodny z aksjomatami rachunku prawdopodobieństwa wskazanymi przez Kołmogorowa:

Aksjomat 1: Każde zdarzenie losowe **A** charakteryzuje pewna liczba **P(A)** zwana prawdopodobieństwem zdarzenia **A**, spełniająca nierówność $0 \leq P(A) \leq 1$. $0 \leq P(E)$.

Aksjomat 2: Prawdopodobieństwo zdarzenia pewnego równa się 1, zaś zdarzenia niemożliwego wynosi 0. $P(\Omega) = 1$.

Aksjomat 3: Prawdopodobieństwo sumy zdarzeń parami wyłączających się równe jest sumie prawdopodobieństw tych zdarzeń. Aksjomaty są podstawą wyznaczania miary niepewności (jaką jest prawdopodobieństwo) w podejściu subiektywnym: $P(E_1 E_2 E_3 \dots) = \sum P(E_x)$.

Następnie pod wpływem każdego kolejnego świadectwa empirycznego agent modyfikuje rozkład prawdopodobieństwa danej hipotezy zgodnie z twierdzeniem Bayesa.

Aksjomaty Kołmogorowa są wystarczające do udowodnienia twierdzenia Bayesa, jednak aby takie wnioskowania mogły mieć **charakter sekwencyjny**, konieczne jest **uzupełnienie aksjomatów rachunku prawdopodobieństwa** o tzw. **regułę warunkowania**. Zasada warunkowania określa sposób uzgodnienia prawdopodobieństwa hipotezy po dokonaniu jednego eksperymentu **E1**, a przed dokonaniem kolejnego **E2** (tzw. wnioskowanie sekwencyjne): $P(H E1) = Q(H)$. Wpływ wyników nowego eksperymentu **E2** na hipotezę obliczamy oczywiście z twierdzenia Bayesa (**TB**) dla „warunkowanej” funkcji, a więc dla $Q(H)$. Stopień przekonania zaktualizowany po otrzymaniu nowych informacji obserwacyjnych lub eksperymentalnych reguła warunkowania traktuje jako prawdopodobieństwo pierwotne w twierdzeniu Bayesa przed otrzymaniem kolejnego wyniku obserwacji lub eksperymentu¹.

Współcześnie w ramach podejścia bayesowskiego została sformułowana teoria przetwarzania (kodowania) predykcyjnego. W kognitywistyce teoria ta często stanowi punkt wyjścia dla tworzenia modeli wyjaśniających różne zjawiska poznawcze. Hipoteza predykcyjna opiera się na tezie, że percepcja, procesy uwagowe oraz kontrola zachowania mają wspólny, nadrzędny cel, jakim jest unikanie zaskoczenia. „Zaskoczenie” w ramach

¹ Ze względu na problem prawdopodobieństwa pierwotnego standardowy (subiektywny) wariant bayesianizmu często wzbogaca się o tezy (na przykład o zasadę ekspercką), które mają ustrzec to stanowisko przed arbitralnym przypisywaniem wartości stopniom przekonania w punkcie wyjścia, czyli przed uzyskaniem jakichkolwiek wyników obserwacji czy eksperymentów. Przykładem jest obiektywny bayesianizm Kazimierza Ajdukiewicza (Kawalec, 2012).

przetwarzania predykcijnego (dalej: PP) definiowane jest jako błąd predykcijny, czyli rozbieżność pomiędzy sygnałem zmysłowym docierającym do systemu poznawczego ze środowiska a wewnętrznymi wygenerowanymi przez ten system przewidywaniami (Friston, 2009; 2012; Clark, 2013; 2016; Hohwy, 2016; Wiese, Metzinger, 2017). Zminimalizowanie błędu predykcijnego opiera się na zasadzie: „Spraw, by różnica między przewidzianymi a rzeczywistymi stanami aparatu zmysłowego była jak najmniejsza”. Zgodnie z PP mózg generuje odpowiednie predykcje, szacując poziom swojej niepewności (Clark, 2016: 57), a przebieg informacji w mózgu jest hierarchiczny i dwustronny. Procesy odgórne odpowiadają za przewidywania tego, jak świat będzie oddziaływał na aparat zmysłowy. Z kolei procesy oddolne obejmują informacje napływające ze świata i przekazują sygnały dotyczące rozmiaru błędu.

Hipoteza mózgu bayesowskiego głosi, że centralny układ nerwowy konstruuje i testuje wewnętrzne modele świata zewnętrznego na drodze realizacji specyficznych procesów poznawczych, które są niczym innym jak aproksymacją wnioskowania bayesowskiego (Hohwy, 2013; 2016; Harkness, Keshava, 2017). Czyniąc użytek z hierarchicznie zorganizowanych inferencji, mózg tworzy odpowiednie **modele generatywne**, które odgórnie konstruują hipotezy interpretujące oddolnie napływające informacje, które pochodzą z sygnału zmysłowego. Zatem predykcje są z jednej strony warunkowane posiadanym przez dany system poznawczy wewnętrznym modelem świata (na który składa się uprzednio zdobyta wiedza), z drugiej zaś przez zmienne informacje napływające ze świata. Obie te warstwy wzajemnie się ograniczają, pozostając w dynamicznych zależnościach (Hohwy, 2013: 69–70). Każdy poziom modelu generatywnego minimalizuje błędy predykcyjne na niższym poziomie: od wyższych procesów poznawczych po procesy neuronalne. Takie podejście wiąże się jednak z postulowaniem istnienia hierarchicznie zorganizowanych **(neuro)reprezentacji**, które są tworzone na poszczególnych poziomach modelu generatywnego świata. Statystyczne przewidywania, jakie generuje mózg, funkcjonują właśnie jako reprezentacje tego, co wydarza lub ma się wydarzyć w świecie, z którego pochodzą stany percepcyjne. W tym miejscu warto zaznaczyć, że **paradygmat ucieleśnienia generalnie odrzuca istnienie reprezentacji mentalnych**.

2. Konserwatywne przetwarzanie predyktywne i zasada swobodnej energii

W ramach zyskującej na popularności teorii postrzegającej mózg jako narzędzie prognozy możemy wyróżnić dwa główne nurty. Pierwszy z nich jest w literaturze określany jako **konserwatywne przetwarzanie predykcyjne**. Podejście to ogranicza poznanie do procesów i struktur mózgowych, nie traktuje ciała ani środowiska jako składowych procesów poznawczych. Koncepcja ta nie przystaje do obowiązującego paradygmatu ucieleśnionego poznania. Konserwatywne PP jest bowiem stanowiskiem **internalistycznym** i **reprezentacjonistycznym**, z kolei paradygmat ucieleśnionego poznania wyraźnie przyjmuje postawę antyreprezentacjonistyczną (Gładziejewski, 2016).

Możemy jednak w ramach przetwarzania predykcijnego wyróżnić inny nurt, opierający się na **zasadzie swobodnej energii** (ang. *free energy principle* – FEP), określany

jako **enaktywizm swobodnej energii** lub **radykałne przetwarzanie predykcyjne** (Piekarski, 2020). Z powodu wielu odmian i braku wyraźnych deklaracji pojęciowych nie jest łatwo jasno zdefiniować i określić, czym jest enaktywizm. Na pewno wpisuje się on w nurt ucieleśnionego poznania, uznając, że aktywność poznawcza jest procesem zachodzącym w wyniku oddziaływania autonomicznego i samoorganizującego się systemu poznawczego ze środowiskiem. Aktywną rolę w tym procesie odgrywa również fizyczne ciało systemu poznawczego (Ward, Silverman, Villalobos, 2017: 4). Ponadto w ramach enaktywizmu często percepcję interpretuje się jako formę działania oraz utrzymuje się, że struktury poznawcze wyłaniają się z powtarzających się **wzorców sensomotorycznych**, które umożliwiają działania percepcyjne (Varela, Thompson, Rosch, 1991: 173).

Enaktywizm swobodnej energii ściśle wiąże się z przyjęciem stanowiska **antyrepresentacjonistycznego**, które opiera się na przekonaniu, iż reprezentacje mentalne nie istnieją, a jeżeli nawet istnieją, to wyjaśnienie wykorzystujące to pojęcie nie ma wartości eksplanacyjnej. Zgodnie z tym podejściem minimalizacja błędów predykcyjnych opiera się na zasadzie swobodnej energii (FEP), która dotyczy utrzymywania żywego organizmu w stanie dalekim od równowagi termodynamicznej (Friston, Stephan, 2007; Friston i in., 2010; Friston, 2013a; Sims, 2016). Zdaniem zwolenników tej teorii przy pomocy FEP można wyjaśnić zdolność złożonych systemów, takich jak na przykład ludzki mózg, do samoorganizacji, która jest niczym innym jak tendencją do unikania zaburzeń, czyli pewną wewnętrzną potrzebą utrzymania homeostatycznej równowagi ze środowiskiem, również za pomocą mechanizmów allostacyjnych (Clark, 2016: 305).

Enaktywizm swobodnej energii bez wątpienia jest alternatywnym sposobem interpretacji przetwarzania predykcyjnego, który zrywa z inferencjalizmem i konstruktywizmem, który obecny jest w konserwatywnym PP. Względem tego podejścia pojawiają się jednak inne zarzuty. Kluczowe pojęcia, takie jak model generatywny czy wnioskowanie bayesowskie, tracą zupełnie na ostrości i nie do końca wiadomo, jak należałoby je rozumieć. Jeżeli modelem generatywnym jest cały organizm (Friston, 2013a: 274), a jego zadaniem nie jest reprezentacja świata, lecz kierowanie interakcjami podmiotu w jego otoczeniu w celu utrzymania równowagi pomiędzy mózgiem, ciałem i środowiskiem (Bruineberg, Rietveld, 2014: 7), to nie jest w pełni jasne, w jaki sposób konstytuowana jest relacja między mózgiem a modelem generatywnym oraz jak model ten może kierować działaniami podmiotu, skoro nie reprezentuje on w żaden sposób struktury środowiska. Kolejnym problemem jest nieinferencyjne odczytanie teoremu Bayesa, który tutaj rozumiany jest jako rodzaj działania (por. Ramstead, Kirchhoff, Friston, 2019), co może jednak prowadzić do konkluzji, że podmiot nie posiada swojego modelu świata, lecz to on jest tym modelem. Enaktywizm swobodnej energii rozciąga poznanie poza organizm na świat tak, że zacierają się granice między zjawiskami poznawczymi i niepoznawczymi. Co za tym idzie, stanowisko to nie oferuje żadnych unikalnych wyjaśnień dla bardziej złożonych zjawisk poznawczych, które są specyficzne dla ludzkiego poznania (wbrew temu, co twierdzą jego zwolennicy). **W kontekście poruszanego problemu generowania abstrakcyjnych pojęć ogólnych, jak się wydaje, teoria FEP również nie wnosi niczego do paradygmatu ucieleśnionego języka.**

3. Hierarchiczne probabilistyczne modele generatywne

Nadzieją na stworzenie komplementarnego stanowiska, które będzie łączyło zalety konserwatywnego PP z FEP, jest stanowisko hybrydowe wprowadzające architekturę hierarchicznych probabilistycznych modeli generatywnych (dalej HPGM z ang. *hierarchical probabilistic generative model*), która jest podstawową strukturą reprezentacyjną w umyśle (Clark, 2016; Kiefer, Hohwy, 2017). Podejście to zdaniem niektórych autorów (por. m.in. Friston, Kiebel, 2009; Bruineberg, Rietveld, 2014; Hohwy, 2015a; 2015b; 2016; 2017; Bruineberg, Kieverstein, Rietveld, 2016; Kirchhoff i in., 2018; Ramstead, Kirchhoff, Friston, 2019) pozwala na lepsze osadzenie PP w neuroauce, a z drugiej strony **lepiej wyjaśnia myślenie pojęciowe niż samo stanowisko poznania ucieleśnionego (i języka)**. Możliwość połączenia dwóch perspektyw, tj. homeostatycznej (FEP) i architektonicznej (PP), jest widoczna w zastosowaniu tak zwanych koców Markowa². Zgodnie z omawianą interpretacją wewnętrzne stany organizmu można zdefiniować jako koce Markowa, które oddzielają system od jego środowiska. Są one zaprojektowane tak, aby zminimalizować energię swobodną w celu utrzymania integralności strukturalnej i funkcjonalnej, co prowadzi do wewnątrzsystemowej homeostazy i samoregulacji. Tak więc warunkiem minimalizacji energii swobodnej jest zdolność organizmu do odróżnienia się od środowiska (Friston, 2013a). W odniesieniu do PP należy stwierdzić, że kolejne poziomy modelu są zaimplementowane w mózgu jako hierarchicznie zorganizowane koce Markowa, stale zaangażowane w proces minimalizacji błędów przewidywania poprzez minimalizację energii swobodnej (Friston, 2009).

Koce Markowa ograniczają ilość informacji przekazywanych do wnętrza systemu. W ten sposób gwarantują utrzymanie poziomu entropii na jak najniższym poziomie (por. Ashby, 1963: 247). Informacje docierające do systemu służą do uzasadnienia przewidywań generowanych przez model. Koce Markowa znajdują się na każdym poziomie modelu generatywnego, a minimalizacja błędów precyzji (energii swobodnej), a tym samym różnica między przewidywaniami a wnioskowaniem na podstawie wejść sensorycznych, może być wyjaśniona wewnętrznymi stanami organizmu (Kirchhoff i in., 2018: 3–4; por. też Ramstead, Badcock, Friston, 2018). Jak się wydaje, można tu mówić o **silnym internalizmie**, ale ważne okazuje się również pojęcie **aktywnego wnioskowania**, tj. ingerencji (poprzez działanie) w przyczynową strukturę świata, która pozwala na uprawdopodobnienie przyjętych hipotez predykcyjnych.

Konsekwentne odczytanie konserwatywnego podejścia (nawet przy silnej interpretacji zobowiązania do reprezentacji) pozwala na włączenie go w nurt ucieleśnionego poznania. Koc Markowa ustanawia warunkową niezależność wewnętrznych stanów organizmów od zewnętrznych stanów środowiska. Te pierwsze są otwarte na te drugie tylko w sensie warunkowym. Następnie w kocu Markowa można wyróżnić **stany sensoryczne i aktywne**.

² Koncepcja łańcuchów Markowa wywodzi się z pracy Andrieja Markowa, który badał abstrakcyjne systemy, które pamiętają swoje przeszłe trajektorie tylko wtedy, gdy przechowują bieżące sygnały. Łańcuch Markowa pierwszego rzędu ($n = 1$) to system, w którym prawdopodobieństwo każdego zdarzenia zależy tylko od wyniku poprzedniego zdarzenia. Łańcuchy wiążą się z istnieniem procesów stochastycznych, czyli procesów odnoszących się do danej rodziny zmiennych losowych, które można zdefiniować na pewnej przestrzeni probabilistycznej. Procesy stochastyczne spełniające własność Markowa nazywane są procesami Markowa.

Rozróżnia się je w następujący sposób: stany wewnętrzne (aktywne) nie mogą wpływać na stany sensoryczne, podczas gdy stany zewnętrzne (sensoryczne) nie mogą wpływać na stany aktywne (por. Friston, 2012). Taka warunkowa niezależność pozwala na statystyczne oddzielenie wewnętrznych stanów koca (na przykład komórki lub organizmu) od zewnętrznych stanów jego środowiska. Stany wewnętrzne zatem poprzez minimalizację zaskoczenia organizmu kodują rozkład prawdopodobieństwa dotyczący stanów zewnętrznych, tj. przyczyn wrażeń zmysłowych odbieranych przez dany koc. Energia swobodna jest zatem funkcją opisującą rozkład prawdopodobieństwa zakodowany przez wewnętrzne stany koca Markowa. Zaskoczenie jest funkcją stanów samego koca Markowa. Innymi słowy, wolna energia jest funkcją probabilistycznych przekonań zakodowanych przez stany wewnętrzne na temat stanów zewnętrznych (czyli oczekiwań dotyczących prawdopodobnych przyczyn bodźców sensorycznych) (Ramstead, Badcock, Friston, 2018: 3–4). Reprezentacje nie są statycznymi obrazami rzeczywistości, ale **wewnętrznymi reprezentacjami strukturalnymi**, które kierują działaniem i pozwalają nam rozpoznać błędy reprezentacyjne. Hohwy (2018) dodaje, że PP pozwala nam lepiej wyjaśnić, w jaki sposób ucieleśnione podmioty wchodzą w dynamiczne interakcje ze środowiskiem. Granica między umysłem a światem jest z jednej strony **samowierzytelniająca** (ang. *self-evidencing*) (Hohwy, 2016), co oznacza, że przyczyny stymulacji sensorycznej są znane pośrednio poprzez wnioskowanie o informacjach docierających z wejść sensorycznych. Z drugiej strony ma ona charakter przyczynowy – istnieje dynamiczne sprzężenie zwrotne między umysłem a światem, możliwe dzięki percepcji i działaniom wykonywanym w tym świecie.

4. Wyzwania stojące przed HPGM

Zanim przejdziemy do omówienia ograniczeń, jakimi charakteryzują się HPGM, warto porównać podejście pierwszego kognitywizmu, opierającego się na gramatyce generatywnej i pojęciu symbolu amodalnego, z PP skupiającym się na pojęciu prawdopodobieństwa (Tenenbaum i in., 2011; Goodman, Tenenbaum, Gerstenberg, 2015). Pierwsze z podejść akcentuje kompozycyjny charakter myślenia pojęciowego i języka, opartego na manipulowaniu symbolami amodalnymi na podstawie określonych reguł (por. Chomsky, 1982). Wprowadzona w tym podejściu architektura kognitywna dobrze nadaje się do modelowania myślenia abstrakcyjnego czy logicznego rozumowania ogólnego, ograniczającego się do sfery pewności, a co za tym idzie, zgodnego z tradycyjnym modelem racjonalności. Mimo pewnych sukcesów eksplanacyjnych tego podejścia obecnie wskazuje się, że takie podejście jest sprzeczne z faktami neurobiologicznymi (Pearl, 1988; Churchland i in., 2012). Z drugiej strony nowsze architektury kognitywne, takie jak PP, są znacznie bardziej biegłe w radzeniu sobie z niepewnością i dużo bliższe wynikom badań w zakresie neurobiologii. Nie radzą sobie jednak z wyjaśnieniem ustrukturyzowanych i kompozycyjnych zdolności reprezentacyjnych, które wydają się niezbędne do uchwycenia i wyjaśnienia poznania wyższego rzędu, a także naszej zdolności do operowania na ogólnych pojęciach abstrakcyjnych. Jak twierdzi Clark (2015; 2016: 24, 171–174), HPGM są w stanie uchwycić to, co atrakcyjne w obu tych, zdawałoby się, przeciwstawnych podejściach,

łącząc realizm neurobiologiczny i wnioskowanie statystyczne z hierarchicznie ustrukturyzowanymi zdolnościami reprezentacyjnymi, kluczowymi w generowaniu ogólnych pojęć abstrakcyjnych (Clark, 2016: 24, 171–174).

Wbrew temu, co twierdzi Clark, uważam, że przetwarzanie predykcyjne postulujące HPGM nie jest w stanie uchwycić ani wyjaśnić dwóch podstawowych i powiązanych ze sobą cech myślenia pojęciowego: po pierwsze, jego **ogólności** – czyli faktu, że możemy myśleć i rozumować o zjawiskach na dowolnym poziomie abstrakcji; po drugie, jego **kompozycyjności** – czyli sposobu, w jaki pojęcia są produktywnie łączone ze sobą w nowy sposób, aby lepiej wyrażać nasze myśli. Cechy przed chwilą zidentyfikowane wydają się mieć kluczowe znaczenie dla myślenia pojęciowego w ogóle (wliczając w to operowanie na pojęciach abstrakcyjnych).

Ogólność naszego myślenia przejawia się w tym, że możemy myśleć i rozumować o zjawiskach na dowolnym poziomie abstrakcji w sposób, który elastycznie łączy reprezentacje na różnych poziomach. Istniejące poglądy dotyczące reprezentacji mentalnych w ramach przetwarzania predykcyjnego nie są w stanie jak dotąd wyjaśnić tego faktu. To, że możemy angażować się w wysoce abstrakcyjne myślenie o zjawiskach na dowolnym poziomie skali czasoprzestrzennej, podważa pogląd, że każdy prosty, hierarchicznie iterowany proces minimalizacji błędów predykcji wyczerpuje prawdę o reprezentacjach poznawczych. Co ważniejsze, możemy myśleć i rozumować nie tylko o konkretnych zjawiskach w dowolnej skali czasoprzestrzennej, ale możemy myśleć o zjawiskach spoza tej skali, na przykład o wnioskowaniu bayesowskim, liczbie naturalnej czy postmodernizmie. Takie zjawiska nie odgrywają żadnej roli przyczynowej ani nie są związane z sensomotorycznymi pobudzeniami, a zatem nie są predykcyjne dla żadnych modalności. **O ile relacje międzypoziomowe w hierarchicznym wnioskowaniu można wyjaśnić jako rekurencyjny proces minimalizacji błędu predykcji, nie jest oczywiste, w jaki sposób reprezentacje abstrakcyjnych pojęć mogą w ogóle występować.**

Podobnie przetwarzanie predykcyjne nie jest w stanie wyjaśnić naszej zdolności do elastycznego rozumowania o zjawiskach, które łączy reprezentacje z różnych poziomów hierarchicznych. Kognitywna architektura przetwarzania predykcyjnego, opierająca się na kocach Markowa i warunkowej niezależności nieprzylegających do siebie poziomów hierarchii postuluje, że wpływ reprezentacji na nieprzylegających do siebie poziomach jest zawsze zapośredniczony przez reprezentacje na poziomach pomiędzy nimi. Używając koców Markowa (Pearl, 1988: 120–121), można opisać taką serię węzłów, że dla danego węzła **A** jego zachowanie można w pełni przewidzieć, znając stany innych węzłów, które tworzą jego właściwy koc. Stany sąsiednich węzłów określają statystycznie (nie przyczynowo) stan docelowego węzła **A**. Oznacza to, że koc Markowa dla węzła **A** zawiera całą wiedzę potrzebną do przewidzenia zachowania tego węzła i wynikających z niego węzłów³. Innymi słowy, znajomość stanów węzłów zewnętrznych nie jest konieczna do

³ W sieciach bayesowskich kocy Markowa są opisywane za pomocą pojęć rodziców i dzieci. Wyjaśnienie zachowania danego dziecka wymaga odniesienia się do zachowania jego rodzica i innych dzieci. Nie jest więc konieczna znajomość stanów węzłów poprzedzających rodziców danego dziecka, tj. dziadków, pradziadków itp. W praktyce oznacza to, że jeśli chcemy przewidzieć stan danego węzła, wystarczy znać stany węzłów tworzących jego koc Markowa.

poznania stanów węzłów wewnętrznych. Systemy biologiczne i sztuczne można w zadowalający sposób modelować jako sieci składające się z węzłów (ang. *node*), które reprezentują jednostki składowe systemu, oraz krawędzi (ang. *edge*), które oznaczają interakcje między węzłami. W odniesieniu do mózgu węzły mogą oznaczać neurony, populacje neuronów lub całe obszary mózgu w zależności od skali przestrzennej, którą badacze są zainteresowani.

Problem polega na tym, że myślenie pojęciowe i rozumowanie nie mają tego ograniczenia: możemy łączyć reprezentacje zjawisk na różnych poziomach abstrakcji w ustrukturyzowane przekonania i wykorzystywać je następnie w rozumowaniu w sposób, który wydaje się całkowicie niezgodny z rozumieniem reprezentacji mentalnej w kategoriach hierarchicznych probabilistycznych modeli generatywnych. Możemy na przykład pomyśleć o tym, jak prawa fizyki musiałyby się zmienić, aby pies mógł poruszać się szybciej niż prędkość światła. Są to dwie, nieprzylegające do siebie domeny pojęciowe, które jednak możemy łączyć w elastyczny sposób. Nasza zdolność do elastycznego myślenia i rozumowania o zjawiskach o dowolnej skali czasowo-przestrzennej i o dowolnym poziomie abstrakcji wiąże się z naszą zdolnością do łączenia pojęć w ustrukturyzowane schematy pojęciowe niezależnie od ich treści. W przypadku teorii przetwarzania predykcyjnego pojawia się problem polegający na tym, że hierarchiczne modele generatywne wydają się nie spełniać tej zależności: **Relacja różnych reprezentacji mentalnych względem siebie jest całkowicie zapośredniczona przez ich miejsce w określonej strukturze sieci reprezentacji, ostro ograniczając tym samym dopuszczalne kombinacje reprezentacji.**

Wiąże się z tym kolejny problem, mianowicie przetwarzanie predykcyjne nie jest w stanie uchwycić bogatej **kompozycyjności** myśli. Znaczenie złożonego wyrażenia językowego (takiego jak zdanie) jest zdeterminowane nie tylko przez jego strukturę, ale i znaczenie każdej z jego poszczególnych części składowych. To oznacza, że nasze myśli są **bogato złożone**: proste reprezentacje o wysokim poziomie konkretności (reprezentacje atomowe) włączają się w reprezentacje coraz bardziej złożone i abstrakcyjne (reprezentacje molekularne). Język naturalny możemy określić jako system reprezentacji o charakterze nieskończonym. Kompozycyjność systemu reprezentacyjnego leży u podstaw myślenia pojęciowego i stanowi niezbędne wyjaśnienie dwóch podstawowych cech języka: jego **produktywności** i **systematyczności** (Fodor, 1975; Fodor, Pylyshyn, 1988). Produktywność określa nieskończony charakter wyrażen językowych. Systematyczność wskazuje na regularne relacje pomiędzy pojęciami. Dzięki temu zdolność do operowania na jednym pojęciu pociąga za sobą zdolność do operowania na innym pojęciu. Niemniej jednak samo stwierdzenie, że język jest systematyczny, produktywny i kompozycyjny nie mówi nam wiele o genezie powstawania ogólnych pojęć abstrakcyjnych. Różne systemy reprezentacji mają różne własności kompozycyjne, a tym samym generują różne formy produktywności i systematyczności. Myśl i język są zatem bogato kompozycyjne, a HPGM nie jest w stanie wyjaśnić tej specyficznej cechy myśli i języka. Wystarczy wykazać, że niektóre cechy wyższego poznania, w tym operowanie na pojęciach abstrakcyjnych jest co najmniej tak ekspresywne, jak logika pierwszego rzędu, aby uznać, że HPGM są nieodpowiednie do uchwycenia produktywności myślenia pojęciowego, gdyż nie są one

bogato kompozycyjne. Ich moc eksplanacyjna jest równoważna logice tradycyjnej (Russell, Norvig, 2010: 58), a więc odnosi się tylko do faktów dotyczących światów możliwych, pomijając obiekty i ich wzajemne powiązania.

Zakończenie

Wymienione wyżej cechy: kompozycyjność, produktywność, systematyczność i ogólność, są nie tylko cechami języków naturalnych, ale także pozostają kluczowe dla myślenia pojęciowego czy rozumowania w ogóle (Fodor, 1975; Jackendoff, 2002). Na ten moment jednak przetwarzanie predykcyjne nie jest w stanie uchwycić ogólności i bogatej kompozycyjności naszego myślenia, w tym operowania na pojęciach ogólnych i abstrakcyjnych. Ograniczając reprezentacje do określonych miejsc w strukturze sieci (koce Markowa), nie sposób jest uwzględnić w wyjaśnieniach radykalnej ogólności domeny pojęć abstrakcyjnych czy sposobu, w jaki pojęcia mogą być elastyczne, systematyczne i produktywne. Fakt, że możemy się uczyć i rozumieć języki naturalne, jest trudny do zrozumienia bez założenia systemu reprezentacji przynajmniej tak ekspresyjnego, jak sam język (Pylyshyn, Fodor, 2015). Rozwój takiego systemu reprezentacji mógł istotnie wpłynąć na wyłonienie się wyższych zdolności poznawczych (Bermúdez, 2005: rozdz. 10), a także na procesy minimalizacji błędu predykcji. Teoria przetwarzania predykcyjnego powinna zatem uwzględnić charakterystyczny wkład języka naturalnego w rozwój procesów poznawczych, (Clark, 2016) oraz wyjaśnić, jak ludzka myśl dziedziczy systematyczność, którą wykazuje struktura gramatyczna języka (Clark, 2000: 77), aby uzyskać pełne i satysfakcjonujące stanowisko. Oczywiście, pełna teoria powinna również uwzględniać aktywną rolę ciała w procesach poznawczych, a tym samym zostać włączona w nurt ucieleśnionego poznania (i języka). Ciało bowiem, jak dotąd wykazano, reguluje aktywność poznawczą, gwarantując ściśle powiązanie pomiędzy poznaniem a działaniem. Ponadto działa niczym dystrybutor w tym sensie, że przenosi część obciążenia poznawczego. Służy także do ograniczania informacji przetwarzanych przez podmiot (Venter, 2021). Dlatego uważam, że mimo ograniczeń, na jakie wskazałam, teoria przetwarzania predykcyjnego stanowi obiecującą alternatywę, która jednak wymaga dalszego rozwijania.

Bibliografia

- Ashby W.R. (1963), *Wstęp do cybernetyki*, PWN, Warszawa.
- Bermúdez J. (2005), *Philosophy of psychology: A contemporary introduction*, London.
- Bruineberg J., Kiverstein J., Rietveld E. (2016), *The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective*, „Synthese”, vol. 195, s. 1–28, <https://doi.org/10.1007/s11229-016-1239-1>
- Bruineberg J., Rietveld E. (2014), *Self-organization, free energy minimization, and optimal grip on a field of affordances*, „Frontiers in Human Neuroscience”, vol. 8(599), s. 1–14, <https://doi.org/10.3389/fnhum.2014.00599>

- Chomsky N. (1982), *Zagadnienia teorii składni*, Wrocław.
- Churchland M., Cunningham J., Kaufman M., Foster J.D., Nuyujukian P., Ryu S.I., Shenoy K.V. (2012), *Neural population dynamics during reaching*, „Nature”, vol. 487, s. 51–56, <https://doi.org/10.1038/nature11129>
- Clark A. (2000), *Mindware*, New York.
- Clark A. (2013), *Whatever next? Predictive brains, situated agents, and the future of cognitive science*, „Behavioral and Brain Sciences”, vol. 36(3), s. 181–204, <https://doi.org/10.1017/S0140525X12000477>
- Clark A. (2015), *Predicting peace: The end of the representation wars – A reply to Michael Madary*, [w:] T. Metzinger, J.M. Windt (red.), *Open MIND: 7(R)*, Frankfurt am Main, <https://doi.org/10.15502/9783958570979>
- Clark A. (2016), *Surfing Uncertainty. Prediction, Action, and the Embodied Mind*, Oxford.
- Dove G. (2011), *On the need for embodied and dis-embodied cognition*, „Frontiers in Psychology”, vol. 1.
- Dove G. (2014), *Thinking in words: Language as an embodied medium of thought*, „Topics in Cognitive Science”, vol. 6.
- Dove G. (2016), *Three symbol ungrounding problems: Abstract concepts and the future of embodied cognition*, „Psychonomic Bulletin & Review”, vol. 23.
- Dove G. (2018), *Language as a disruptive technology: abstract concepts, embodiment and flexible mind*, „Philosophical Transactions of the Royal Society B: Biological Sciences”, vol. 373.
- Fodor J. (1975), *The language of thought*, Cambridge.
- Fodor J., Pylyshyn Z. (1988), *Connectionism and cognitive architecture: A critical analysis*, „Cognition”, vol. 28(1–2), s. 3–71, [https://doi.org/10.1016/0010-0277\(88\)90031-5](https://doi.org/10.1016/0010-0277(88)90031-5)
- Fornal M. (2022), *Problem ogólnych pojęć abstrakcyjnych w kontekście językoznawstwa kognitywnego*, „Językoznawstwo”, nr 1(16), s. 9–27.
- Friston K.J. (2009), *The free-energy principle: A rough guide to the brain?*, „Trends in Cognitive Sciences”, vol. 13(7), s. 293–301.
- Friston K.J. (2012), *A free energy principle for biological systems*, „Entropy”, vol. 14(11), s. 2100–2121, <https://doi.org/10.3390/e14112100>
- Friston K.J. (2013a), *Active inference and free energy*, „Behavioral and Brain Sciences”, vol. 36(3), s. 212–213, <https://doi.org/10.1017/S0140525X12002142>
- Friston K.J. (2013b), *Consciousness and hierarchical inference*, „Neuropsychanalysis”, vol. 15(1), s. 38–42.
- Friston K.J., Daunizeau J., Kilner J., Kiebel S.J. (2010), *Action and behavior: A free-energy formulation*, „Biological Cybernetics”, vol. 102, s. 227–260, <https://doi.org/10.1007/s00422-010-0364-z>
- Friston K.J., Kiebel S.J. (2009), *Predictive coding under the free-energy principle*, „Philosophical Transactions of the Royal Society B”, vol. 364, s. 1211–1221.
- Friston K.J., Stephan K.E. (2007), *Free energy and the brain*, „Synthese”, vol. 159, s. 417–458.
- Gładziejewski P. (2016), *Predictive coding and representationalism*, „Synthese”, vol. 193, s. 559–582, <https://doi.org/10.1007/s11229-015-0762-9>

Goodman N., Tenenbaum J., Gerstenberg T. (2015), *Concepts in a probabilistic language of thought*, [w:] E. Margolis, S. Laurence (red.), *The conceptual mind: New directions in the study of concepts*, Cambridge, s. 623–654.

Harkness D.L., Keshava A. (2017), *Moving from the what to the how and where – Bayesian models and predictive processing*, [w:] T. Metzinger, W. Wiese (red.), *Philosophy and Predictive Processing*, Frankfurt am Main, s. 1–10, <https://doi.org/10.15502/9783958573178>

Hohwy J. (2013), *The Predictive Mind*, Oxford.

Hohwy J. (2015a), *The neural organ explains the mind*, [w:] T. Metzinger, J.M. Windt (red.), *Open MIND*, 19(T), Frankfurt am Main, s. 1–22, <https://doi.org/10.15502/9783958570016>

Hohwy J. (2015b), *The diversity of Bayesian explanation – a reply to Dominic L. Harkness*, [w:] T. Metzinger, J.M. Windt (red.), *Open MIND*, 19(R), Frankfurt am Main, s. 1–6, <https://doi.org/10.15502/9783958570870>

Hohwy J. (2016), *The self-evidencing brain*, „Noûs”, vol. 50(2), s. 259–285, <https://doi.org/10.1111/nous.12062>

Hohwy J. (2017a), *How to entrain your evil demon*, [w:] T. Metzinger, W. Wiese (red.), *Philosophy and Predictive Processing*, Frankfurt am Main, s. 1–15, <https://doi.org/10.15502/9783958573048>

Hohwy J. (2017b), *Priors in perception: Top-down modulation, Bayesian perceptual learning rate, and prediction error minimization*, „Consciousness and Cognition”, vol. 47, s. 75–85, <https://doi.org/10.1016/j.concog.2016.09.004>

Hohwy J. (2018), *The predictive processing hypothesis*, [w:] A. Newen, L.S. de Bruin, S. Gallagher (red.), *The Oxford handbook of 4e cognition*, Oxford, s. 129–145.

Jackendoff R. (2002), *Foundations of language*, New York.

Kawalec P. (2003), *Zagadnienia metodologiczne w bayesowskiej teorii konfirmacji*, „Roczniki Filozoficzne”, t. LI, s. 113–142.

Kawalec P. (2012), *Bayesianizm w polskiej tradycji probabilizmu – studium stanowiska Kazimierza Ajdukiewicza*, „Ruch Filozoficzny”, nr 1(69), s. 111–122.

Kiefer A., Hohwy J. (2017), *Content and misrepresentation in hierarchical generative models*, „Synthese”, vol. 195, s. 2387–2415, <https://doi.org/10.1007/s11229-017-1435-7>

Kirchhoff M., Parr T., Palacios E., Friston K., Kiverstein J. (2018), *The Markov blankets of life: Autonomy, active inference and the free energy principle*, „Journal of the Royal Society Interface”, vol. 15, s. 1–11, <https://doi.org/10.1098/rsif.2017.0792>

Kwisthout J., van Rooij I. (2019), *Computational resource demands of a predictive Bayesian brain*, „Computational Brain & Behavior”, vol. 3(3), s. 1–15, <https://doi.org/10.1007/s42113-019-00032-3>

Pearl J. (1988), *Probabilistic reasoning in intelligent systems*, San Francisco.

Piekarski M. (2020), *Mechanizmy predykcyjne i ich normatywność*, Warszawa, <https://doi.org/10.47943/lib.9788363487447>

Pylyshyn Z., Fodor J. (2015), *Minds without meanings: An essay on the content of concepts*, Cambridge.

- Ramstead M.J.D., Badcock P.B., Friston K.J. (2018), *Answering Schrödinger's question: A free-energy formulation*, „Physics of Life Reviews”, vol. 24, s. 1–16, <https://doi.org/10.1016/j.plrev.2017.09.001>
- Ramstead M.J.D., Kirchhoff M.D., Friston K.J. (2019), *A tale of two densities: Active inference is enactive inference*, „Adaptive Behavior”, vol. 28(4), s. 1–15, <https://doi.org/10.1177/1059712319862774>
- Russell S., Norvig P. (2010), *Artificial intelligence: A modern approach*, London.
- Schwartenbeck P., FitzGerald T., Dolan R.J., Friston K. (2013), *Exploration, novelty, surprise, and free energy minimization*, „Frontiers in Psychology”, vol. 4(710), s. 1–5, <https://doi.org/10.3389/fpsyg.2013.00710>
- Sims A. (2016), *A problem of scope for the free energy principle as a theory of cognition*, „Philosophical Psychology”, vol. 29, s. 967–980, <https://doi.org/10.1080/09515089.2016.1200024>
- Tenenbaum J., Kemp C., Griffiths T., Goodman N. (2011), *How to grow a mind: Statistics, structure, and abstraction*, „Science”, vol. 331(6022), s. 1279–1285, <https://doi.org/10.1126/science.1192788>
- Varela F., Thompson E., Rosch E. (1991), *The embodied mind: Cognitive science and human experience*, Cambridge–London.
- Venter E. (2021), *Toward an Embodied, Embedded Predictive Processing Account*, „Frontiers in Psychology”, vol. 12, <https://doi.org/10.3389/fpsyg.2021.543076>
- Ward D., Silverman D., Villalobos M. (2017), *Introduction: The varieties of enactivism*, „Topoi”, vol. 36, s. 365–375, <https://doi.org/10.1007/s11245-017-9484-6>
- Wiese W., Metzinger T. (2017), *Vanilla PP for Philosophers: A Primer on Predictive Processing*, [w:] T. Metzinger, W. Wiese (red.), *Philosophy and Predictive Processing*, Frankfurt am Main.

Abstract

The theory of predictive processing and the problem of general abstract concepts from the perspective of cognitive linguistics

The contemporary embodiment paradigm in cognitive linguistics provides a valuable conceptual framework for explaining the grounding of concrete concepts but faces fundamental difficulties in explaining the mechanism of abstract concept formation (the so-called disembodiment problem of concepts). It has been increasingly pointed out that the solution to this difficulty lies in combining the embodiment paradigm with the theory of predictive processing. Although this theory aspires to be a general brain theory in the cognitive sciences, it has some limitations, albeit in explaining the salient features of general abstract concepts. The article analyzes the theory of predictive processing in terms of its ability to explain the composability, productivity, systematicity and generality of conceptual thinking. Despite the limitations pointed out in the article, predictive processing theory, combined with the embodied language paradigm, is a promising proposal within second-generation cognitivism.

Keywords: predictive processing, embodied cognition, embodied language, problem of abstract concepts, Clark, Hohwy