

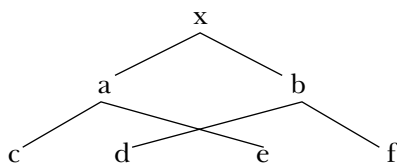
MAGDALENA DERWOJEDOWA

KONSTRUKCJE NIECIĄGŁE W JĘZYKU NATURALNYM

1. WSTĘP

Celem niniejszej pracy jest przedstawienie pewnego typu konstrukcji składniowych o szczególnej linearyzacji i metod ich opisu w dwóch (uważanych niekiedy za przeciwstawne) teoriach lingwistycznych oraz przedstawienie pewnej hipotezy ich powstawania. Choć są tekstowo typowe i obecne w wielu językach, tak w subkodzie mówionym, jak i pisanym, ich opis w zdyscyplinowanych gramatykach wymaga specjalnych rozwiązań. Chodzi tu o takie konstrukcje składniowe, których składniki są rozdzielone przez składnik innej konstrukcji, co schematycznie można przedstawić następująco: jeśli a, b, c, d, e, f są składnikami pewnej konstrukcji, to drzewo ją reprezentujące ma następującą postać:

(1)

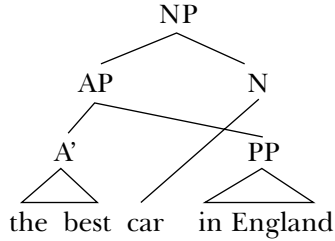


Konstrukcje takie będziemy nazywać *nieciągłymi* lub *nieciągłościami*. Najpierw omówimy typy tych konstrukcji, następnie krótko scharakteryzujemy gramatyki zależnościowe i gramatyki składników bezpośrednich. Kolejna część poświęcona będzie opisowi nieciągłości. W ostatniej części omówimy hipotezę, która próbuje wyjaśnić powstawanie konstrukcji nieciągłych.

2. KONSTRUKCJE NIECIĄGŁE

Konstrukcja jest nieciągła, jeśli w porządku linearnym jej składniki rozdziela składnik dominowany przez inny nadrzędnik (w ciągu *cde* w przykładzie (1) jest to składnik *d*), w zdaniu:

(2)



fraza *best... in England*, której odpowiada polska (również nieciągła) fraza¹:

(3) najlepszy samochód w Anglii

Nieciągłość składników charakteryzuje też angielskie *phrasal verbs*, niemieckie *trennbare Verben* czy polskie konstrukcje z aglutynantem:

(4) Mary will *wake* you *up* in the morning.

Mary AUX obudzić ty PART w ART rano

(5) Ich *rufe* dich später *an*.

ja dzwonię cię później PREF

(6) Byłoby dobrze, *gdyby-ś* wcześniej *przyszł*

szeroko rozumianą parentezę:

(7) John talked, of course, about politics.

John rozmawiał oczywiście o polityce

(8) Jest to – co pragnę podkreślić – zjawisko skrajnie rzadkie

składniki przesunięte ku początkowi, jak w pytaniach:

(9) *Whom will* the president *invite*?

kogo AUX ART prezydent zaprosić

(10) *De quel* sujet *as* tu *parlé*?

Na jaki temat AUX ty mówić+ PASTPART

(11) *Jakie* Marysia wzięła ze sobą *sukienki*?lub ku końcowi (ang. *extraposition*):

¹ Strzałki wyznaczają powiązania nadrzędnik–podrzędnik, por. pkt 3.

- (12) *Men entered who were wearing black suits.*
 mężczyźni weszli, którzy AUX noszący czarne garnitury

a także konstrukcje w językach o szyku dopuszczającym swobodę składników (takich jak na przykład łacina czy polski, a nawet niemiecki (tzw. *scrambling*):

- (13) Z koncertmistrzem perkusista dyskutował.
- (14) Der Frau hat der Mann die
 ART (DATSG) AUX kobieta ART (NOMSG) mężczyzna DIE (NOMPL)
 Blumen gegeben.
 kwiaty dać+PASTPART
- (15) Puer puellae bellae rosam dat.
 chłopiec dziewczynie ładnej różę dał
- (16) Wypłynąłem na *suchego* przestwór *oceanu*...
- (17) *Niech* Pani *się* tym tak bardzo *nie* martwi.

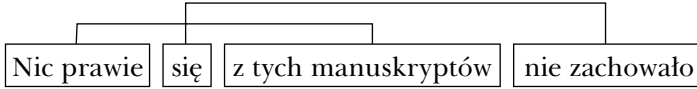
Nieciągłe mogą być morfemy, a więc składniki wyrazów. Intrepretowany tradycyjnie jako forma fleksyjna rozkaznik jest formą nieciąglą, np. *niech prze-czyt- -a/-aj q*. Nieciągłe są też formy Partizip II (*g e- -sag- -t*), a także niektóre derywaty, np. *plask- -o- -den- -ny* czy *n a- -pracować s i ě*; wiele interesujących przykładów nieciągłych morfemów podają klasyczne prace L. Bloomfielda (1933), Z. S. Harrisa (1945) i Ch. Hocketta (1958, § 17.6, §21.4) oraz I. A. Mel'čuka (1963). Współcześnie w lingwistyce zwykle te formy, które mimo złożonej budowy stanowią jedno słowo (rozumiane jako obiekt graficzny wyróżniony spacjami lub znakami równoważnymi), traktuje się jako nieciągłości morfologiczne, natomiast formy takie jak (6) i (17) czy formy tzw. czasów złożonych są interpretowane składniowo. Mogą one też być źródłem nieciągłości (por. (10) i (14)).

Konstrukcje przytoczone w przykładach (2)–(17) to zatem nieciągłości składniowe. Można je podzielić na kilka typów. Najwyraźniej przeciwstawiają się pozostałym zdania (7)–(8), w których składnik zdania jest rozerwany przez składnik nie powiązany składniowo z pozostałymi. McCawley (1982) mówi w tym wypadku o jednostkach parenetycznych. Wiśniewski (1994), a z pewnymi modyfikacjami również Derwojedowa (1996) określają takie jednostki jako dostawione. Derwojedowa (2000, s. 119) nieciągłość tego typu nazywa *zewnątrzną* i wskazuje, że źródłem takiej nieciągłości są wykrzyknienia, zwroty adresatywne, a także rozmaite komentarze, wyrażenia modalne itp.

Przykłady (13)–(17) różnią się od pozostałych tym, że powstały na skutek działania różnych czynników pozagramatycznych: stylistycznych, pragmatycznych czy eufonicznych; zawsze dają się one sprowadzić do nienacecho-

wanej linearyzacji. W niektórych (skrajnie rzadkich²) wypadkach składniki przeplatają się (Szpakowicz, Świdziński 1992):

(18)



Ostatnia grupa to konstrukcje typu (2)–(6) i (9)–(11), a więc takie, których nieciągłość wynika z reguł konstruowania poprawnych jednostek składniowych – zdania te nie mogą mieć innych linearyzacji; Derwojedowa (2000, s. 118) nazywa taką nieciągłość *obligatoryjną*.

W dalszej części artykułu będziemy rozważać tylko nieciągłości typu (3)–(6) i (9)–(14), a więc nieciągłości obligatoryjne i zewnętrzne.

3. GRAMATYKI ZALEŻNOŚCIOWE I GRAMATYKI SKŁADNIKÓW BEZPOŚREDNICH

Przez „gramatykę” rozumiemy sformalizowany opis języka naturalnego, który może być podstawą programu komputerowego (przynajmniej) analizującego zdania, to znaczy rozpoznającego je jako poprawne i przypisującego im pewną strukturę (lub struktury) bądź odrzucającego je jako niepoprawne. Zakładamy też, że oceny uzyskane tą drogą nie różniłyby się od ocen rodzimego użytkownika.

Gramatyki można w zasadzie podzielić na zależnościowe i składników bezpośrednich³. Twórcą podejścia zależnościowego jest L. Tesnière (1959), a najbardziej zaawansowanymi i rozbudowanymi opisami tego typu są teoria Sens ↔ Tekst (ros. *Смысл ↔ Текст*) I. A. Meł’čuka (1974; 1988) i *Word Grammar* R. Hudsona (1990). Analiza opisuje relacje (fr. *connexions*) łączące poszczególne wyrazy danej konstrukcji składniowej⁴.

Opis za pomocą składników bezpośrednich (ang. *immediate constituents*) wiąże się często z pracą R. Wellsa (1947), pojęcie to jednak pojawia się już u L. Bloomfielda (1933) i u innych strukturalistów amerykańskich (Pike 1943; Harris 1951; Gleason 1955; Hockett 1958)⁵. Opis taki jest wykorzystywany również w różnych wersjach generatywizmu (Chomsky 1957, 1965,

² Spośród niemal 5000 analizowanych ze względu na szyk zdań polskich znaleziono tylko siedem nieplątów, por. Derwojedowa (2000, s. 72).

³ Por. Gaifman (1965), Hays (1964, s. 519–522) i Rambow, Joshi (1997) w sprawie równoważności gramatyk składników bezpośrednich i gramatyk zależnościowych.

⁴ „[...] une phrase de type *Alfred parle* n’est pas composée de deux éléments [...] mais de trois éléments, 1° *Alfred*, 2° *parle*, 3° *la connexion* qui les unit et sans laquelle il n’y aurait pas de phrase” (Tesnière, 1959, s. 66).

⁵ Opisem języka polskiego w tej metodologii jest Saloni, Świdziński (2001).

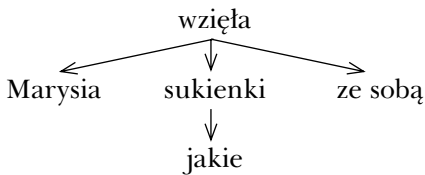
1981). Analiza na składniki bezpośrednie polega na wyodrębnianiu w kolejnych krokach składników danej konstrukcji (fraz).

Najważniejsza z naszego punktu widzenia różnica między tymi podejściami jest taka, że formalizmy generatywne zakładają, iż zdanie jest analizowane w ustalonym porządku, natomiast opisy zależnościowe porządek linearny traktują jako zewnętrzny względem zależności składniowych⁶.

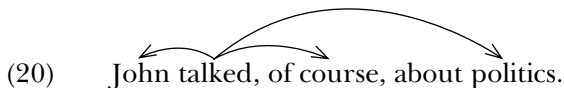
4. KONSTRUKCJE NIECIĄGŁE A GRAMATYKA

W gramatykach zależnościowych wzajemne powiązania składników przedstawia się w postaci strzałek łączących nadrzędniki z podrzędnikami, jak w przykładzie (20), lub jako tak zwane drzewa zależności:

(19)



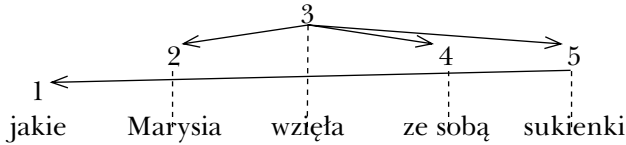
przypominające znany ze szkoły wykres „rozbioru zdania”; warto jednak zwrócić uwagę, że w odróżnieniu od rozbiorów szkolnych nie wyróżnia się tu żadnego argumentu (aktanta) – podmiot jest więc traktowany na równi z (tradycyjnie rozumianymi) dopełnieniami. Podobnie w zdaniu z przykładów (7)/(20) – wtrącony składnik (cyrkumstant) jest podporządkowany predykatowi:



Ograniczenia linearyzacji opisuje się za pomocą kryterium projekcyjności (por. Lecerf (1960); Hays (1964), Mel'čuk (1988, s. 35–38, 42), Mel'čuk, Pertsov (1987, s. 183–186), Obrębski (2002, s. 27–33). Procedura sprawdzania projekcyjności przebiega następująco: wierzchołki drzewa zależności numeruje się w kolejności, w jakiej występują w zdaniu, a następnie spuszcza linie proste prostopadłe do prostej leżącej poniżej drzewa; w konstrukcjach projekcyjnych nie przecinają one gałęzi (por. Jordanska 1963, Hays 1964, s. 518–519). Jak widać, interesujące nas konstrukcje są kandydatami na konstrukcje nieprojekcyjne:

⁶ „Let it be reminded, that in the Meaning-Text framework, linear order of symbols is never allowed to carry information in semantic and syntactic representation; [...] Thus the absence of linear order in our D[ependency]-trees is a corollary of this general principle” (Mel'čuk 1988, s. 16).

(21)



Jordanska (1963) podała listę takich struktur dla języka rosyjskiego (co umożliwia ich odfiltrowanie i specjalne traktowanie podczas analizy). Na przestrzeni lat przedstawiono wiele metod parsowania struktur nieprojekcyjnych w gramatyce zależnościowej, jak przeszukiwanie wstecz i preferencja dla ciągłości (Covington 1990), wskazanie wprost lokalizacji składnika w strukturze (Johnson 1985), pseudoprojekcyjność (Kahane et al. 1998), drzewa topologiczne (Duchier, Debusmann 2001), stopień nieprojekcyjności (Nivre 2004), z wykorzystaniem uczenia (Nivre et al. 2005), koncepcje zaczerpnięte z *Tree Adjoining Grammars* (Rambow, Joshi 1997), domeny szyku (Broker 1996), wykorzystanie cech (*features*), (Øvrelid, Nivre 2007) oraz statystyczne (Eryigit, Oflazer 2006). Warto też wspomnieć, że samo zjawisko znacząco wpływa na wyniki parsowania (Nivre 2004 s. 78), a wspomniane podejścia dają coraz lepsze wyniki.

Można w zasadzie wskazać dwa podejścia do nieciągłości w gramatyce składników bezpośrednich: transformacyjne i nietransformacyjne (Ojeda 2005). Wczesne rozwiązania generatywne nie miały, jak wiadomo, komponentu transformacyjnego. N. Chomsky (1955, s. 405) dopuszcza permutację typu:

(22) (PREFIX, VERB, NP) → (VERB, NP, PREFIX),

wystarczającą, by tworzyć zdania z czasownikami rozdzielnie złożonymi; niewielka zmiana po lewej stronie umożliwia też opisanie *phrasal verbs*. Nietrudno zauważyć, że również w wypadku niektórych innych konstrukcji będzie to strategia skuteczna. V. H. Yngve (1960, s. 78) włącza regułę o postaci:

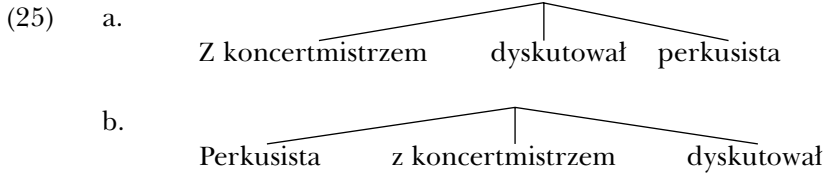
(23) $A \rightarrow B \dots C$,

która umożliwia otoczenie składnika przez inny. Taki mechanizm pozwala, rzecz jasna, opisać interesujące nas struktury, na przykład dla *phrasal verbs* wystarczy wprowadzić regułę o postaci:

(24) PHRASAL_VERB → VERB ... PREP,

by z jednostek słownikowych typu *wake up*, *put off* i frazy rzeczownikowej uzyskać struktury takie jak (4). Podobna reguła pozwala opisać zdania typu (6) czy (9), a z drobnymi modyfikacjami również inne konstrukcje nieciągłe.

Warto w tym miejscu wspomnieć o oryginalnym rozwiązaniu M. Świdzińskiego (1992)⁷. Autor odrzuca struktury przeplatające się typu (15)–(17). Podstawowa struktura zdania elementarnego⁸ obejmuje oprócz zdaniotwórczej frazy finitywnej trzy frazy wymagane i (fakultatywnie) frazę luźną w układach odzwierciedlających linearyzację, przy czym poszczególnym uporządkowaniom przypisuje się różne strukturyzacje; tak więc zdania typu (13)–(14) analizowane są zależnie od porządku składników (przykładowo) jako:



Takie rozwiązanie nie umożliwia jednak opisania zdań typu (9) – gałęzie drzewa nie mogą się bowiem przecinać. Świdziński odrzuca też interpretację, w której składnik pytajny jest podporządkowany głównemu wierzchołkowi – oznaczałaby ona bowiem, że w takich konstrukcjach jest on wymagany przez formę czasownikową, co nie odpowiada intuicji (Świdziński 1992, s. 68). Jak słusznie zauważa sam autor, jest to poważne ograniczenie tej gramatyki. W innych wypadkach (na przykład form typu (6)) reguły są bardzo szczegółowe, jeszcze w innych (głównie zawierających składniki luźne) pojawiają się motywowane technicznie i niekiedy dość nieintuicyjne rozwiązania, jednak znacząca grupa interesujących nas konstrukcji została opisana z dużą subtelnością.

Do rozwiązań nietransformacyjnych należą prace strukturalistów amerykańskich, którzy konstrukcje nieciągłe dopuszczają (por. Wells 1947, s. 102–106; Gleason 1955, s. 142; Hockett 1958, §17.6; por. też Blevins 1990, s. 5–6, 18–25). Spójrzmy na przykładową analizę tego ostatniego:

⁷ Gramatyka Świdzińskiego (1992) jest najobszerniejszym i jak dotąd najbardziej wyczerpującym opisem formalnym polszczyzny. Jest to gramatyka metamorficzna zapisana w postaci reguł PROLOG-u; por. też Szpakowicz (1978), Bień (1996). Została ona zaimplementowana (program *Święgra*), por. Woliński (2005).

⁸ Można je (z pewnym uproszczeniem) utożsamiać z jednostką zorganizowaną wokół osobowej formy czasownika, por. Saloni, Świdziński (2001, s. 43); frazy wymagane to te składniki bezpośrednie, bez których zdanie jest eliptyczne, frazy luźne to frazy niewymagane.

(26)

Is	John	Going	with	you
Aux		V	PP	
		VP		
		VP		
		S		

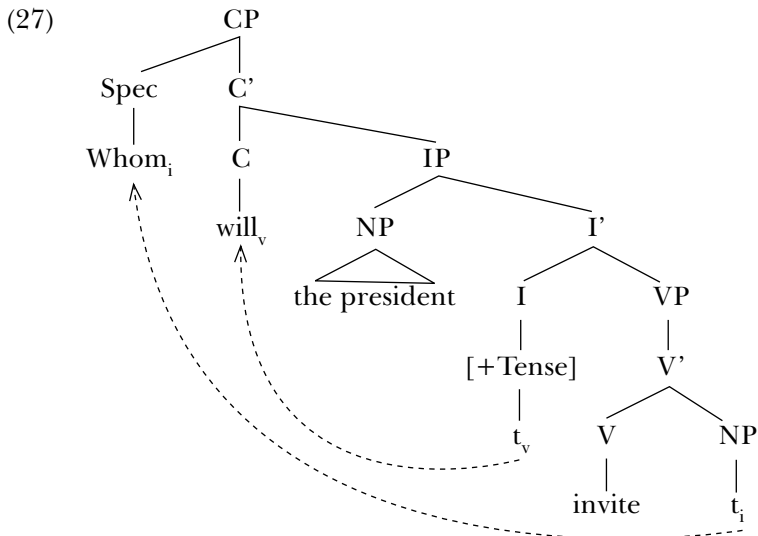
Podwójna linia oznacza, że składnik *John* jest przyłączany dopiero w węźle zdaniowym (w „naturalnym” miejscu podmiotu). Rozwiązanie to skrytykował Chomsky (1957, s. 41); wydaje się jednak, że jego główny zarzut, o niemożności opisanego struktur z czasownikami posiłkowymi, jest nietrafiony – w przedstawionej analizie nieciągłość jest wprost założona. Słusznie uważa Blevins (1990, s. 36), że krytyka ta opiera się na narzuconej przez Chomsky’ego formalizacji pregeneratywnych gramatyk składników bezpośrednich jako odpowiadających gramatykom bezkontekstowym, a nie na zasadach przez nie same sformułowanych. Również zarzuty P. Postal’a (1964, s. 70, 94) są motywowane jego przeświadczeniem, że „tagmemic, like other versions of PSG [Phrase Structure Grammar], has no natural place for discontinuity”. Trudno jednak nie zgodzić się z tym ostatnim, że brak jest ogólnego mechanizmu, który mówiłby, jak konstrukcje tego typu w gramatykach dystrybucyjnych analizować.

Gramatyki generatywne, począwszy od pracy Chomsky’ego (1965, s. 169–173), odrzucają permutacje, wprowadzając na ich miejsce pierwotnie transformacje⁹, a następnie wytarcia, dodania, przesunięcia i inne mechanizmy. Zalety transformacji szczegółowo przedstawia Postal (1964, s. 67–70). Odrzuciwszy możliwość włączenia permutacji do gramatyk struktur frazowych, transformacje permutujące uważa za jedyne, które umożliwiają poprawny opis struktur nieciągłych, i przedstawia następujące argumenty: możliwość zachowania spójnego opisu strukturalnego i ograniczeń selekcyjnych (nie są potrzebne żadne dodatkowe reguły tworzące struktury nieciągłe, są one bowiem przekształcane przez transformacje nakładane na znacznik frazowy), zachowanie jawnych związków między intuicyjnie powiązаныmi strukturami (np. zdaniem twierdzącym i pytaniem tworzonym przez inwersję), dobrze zdefiniowane reguły generowania i dobrze zdefiniowany algorytm automatycznego wyprowadzenia struktury.

Postal zwraca też uwagę na to, że nie każda permutacja prowadzi do nieciągłości, *de facto* formułując własną definicję frazy nieciągłej: permutacji

⁹ Transformacje do lingwistyki wprowadził Z. S. Harris (1957); co prawda transformacje te różnią się od tego, co proponuje gramatyka generatywno–transformacyjna, jednak główna idea – przekształcania powiązanych jednostek – pozostaje taka sama.

dwóch składników, z których przynajmniej jeden nie jest dominowany przez nadrzędnik. Ojeda (2005) pokazuje, że permutacje w zasadzie pozwalają uzyskać ten sam efekt, podobnie zresztą jak późniejsze rozwiązania przyjmujące przesunięcia i wytarcia, w których drzewa reprezentacji składniowej podlegają dwóm regułom: nieprzecinania się gałęzi (co jest oczywiste) i wyłączności, czyli powiązania albo relacją poprzedzania (uszeregowania linearnego), albo dominacji (bezpośredniej nadrzędności; por. Partee et al. 1990, s. 429–450). Składniki mogą się jednak przesuwać na skutek działania czynników gramatycznych. Ogólny mechanizm przesunięcia (ang. *move- α*) pokażemy na przykładzie tzw. *WH-movement*¹⁰:



Składnik pytajny został przeniesiony do pozycji inicjalnej w zdaniu ze swojej bazowej pozycji podrzędnika frazy nominalnej (t_i), podobnie odbywa się też inwersja formy czasownika posiłkowego (t_v). Reguła *move- α* pozwala też zanalizować zdania takie jak (2) i (3): na skutek jej działania składnik *samochód* (będący nadrzędnikiem całej konstrukcji nominalnej) zostaje przesunięty w lewo:

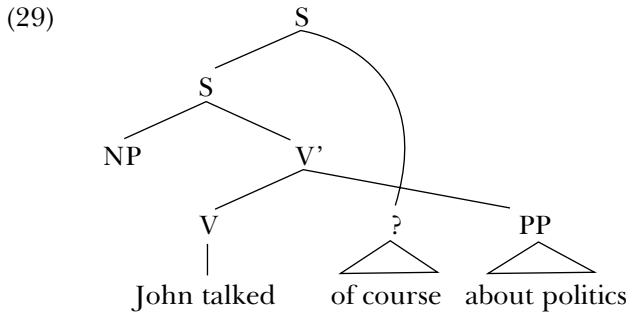
(28) The best car in England t_i will win.



Za pomocą przesunięć są też opisywane ekstrapozycje (por. przykład (12); Bouma, van Noord 1998; Culicover, Rochemont 1997).

¹⁰ Prezentujemy go według założeń teorii Government & Binding; por. Chomsky (1981); Haegeman (1996); Mecner (2005).

W tym samym aparacie McCawley (1982) analizuje zdania typu (7)–(8), w których ciągłość składnika została naruszona przez element syntaktycznie niepowiązany. Proponuje on, by zasady nieprzecinania i wyłączności osłabić tak, ażeby dopuścić składniki nieciągłe. Jednocześnie składnik niepowiązany (dostawiony) przyłącza na najwyższym poziomie struktury, w węźle zdaniowym:



Espinal (1991) wręcz proponuje przenieść analizę fragmentu drzewa, reprezentującego strukturę dostawioną, na płaszczyznę prostopadłą, co – rzecz jasna – nie rozwiązuje problemu. Wydaje się jednak, że sam mechanizm rozerwania struktury został uchwycony dobrze, a niektóre składniki tego typu można uchwycić za pomocą list (np. wiele interiekcji).

Bunt, Thesingh, van der Sloot (1987), wychodząc od konkluzji McCawleya, proponują oparte na przedefiniowaniu sąsiedztwa i wprowadzeniu lokalnych kontekstów reguły, umożliwiające włączenie reguł opisujących nieciągłości do nietransformacyjnych gramatyk frazowych.

5. HIPOTEZA ZAGŁĘBIENIA

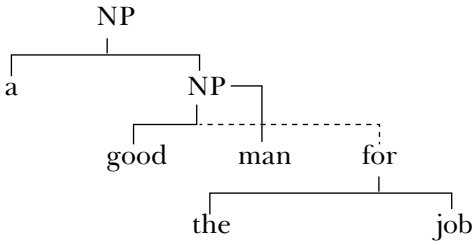
Hipoteza V. H. Yngve’ego (1960), którą chcemy pokrótce przedstawić, próbuje wyjaśnić, dlaczego niektóre konstrukcje nieciągłe są obligatoryjne (por. 2). Przyjmuje ona, że gramatyka struktur frazowych lub gramatyka składników bezpośrednich mogą być podstawą modelu, który produkuje zdania zgodne z szykiem i ortografią danego języka (w tym wypadku języka angielskiego). Autor zakłada też, że zdania są tworzone na bieżąco, stopniowo, ponieważ „niesłusznie byłoby zakładać, że zdania są w pełni formowane w umyśle mówiącego jeszcze przedtem, nim zacznie je on wypowiadać” (s. 71). Ostatnim założeniem jest to, że proponowany model procedury automatycznej (zamierzony na potrzeby tłumaczenia maszynowego) podlega takiemu samemu ograniczeniu, jak przetwarzanie języka przez człowieka – ma skończoną pamięć podręczną.

Na model składają się (nieuporządkowany, skończony) zbiór reguł (w tym reguł pozwalających otaczać składnik przez inne składniki, por. przykład (23)) i mechanizm, który jest automatem skończonym. Możliwe jest wielokrotne stosowanie tej samej reguły, a mechanizm może przechowywać wyniki skończonej liczby operacji pośrednich w pamięci podręcznej. W wyniku działania mechanizmu otrzymuje się zgodne z gramatyką zdanie oraz całą jego ścieżkę derywacyjną.

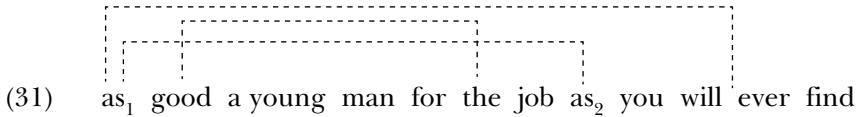
W proponowanym modelu można produkować konstrukcje nieciągłe, mają one jednak wpływ na użycie pamięci pomocniczej: liczba wyników, które trzeba przechować, może bowiem przekroczyć jej pojemność. Z tego powodu, kiedy poznamy ograniczenia automatu, konieczne jest określenie, jak dostosować gramatykę tak, by generowała tylko zdania, które automat działający na strukturze składnikowej jest w stanie przetworzyć. Następnie należy ustalić, jak bardzo język naturalny jest dostosowany do (naturalnego) ograniczenia pamięci krótkotrwałej – wydaje się bowiem, że obserwacje dotyczące angielskiego można pod tym względem uogólnić.

Podstawowe dla oszacowania koniecznej pamięci podręcznej jest obliczenie głębokości budowanej struktury (drzewa). W tym celu gałęzie wychodzące z każdego węzła są numerowane (0... n) od prawej do lewej. Głębokość liczy się, dodając wartości wzdłuż gałęzi prowadzących do danego węzła końcowego; głębokość zdania to maksymalna taka wartość. Zauważmy, że struktury rozgałęziające się w lewo (zwane regresywnymi) mają większą głębokość, a więc wymagają większej pamięci niż struktury rozgałęziające się w prawo (progresywne), które można właściwie rozwijać w nieskończoność, gdyż kolejne nadrzędniki nie zajmują pamięci operacyjnej (i ludzkiej). Jakie jest jednak ograniczenie wymagających pamięci struktur regresywnych? Yngve sięga po sławny wynik G. A. Millera (1956) i przyjmuje, że jest to siedem symboli (± 2 , naturalnie). Takie samo ograniczenie przyjmuje dla języka naturalnego oraz dla gramatyk dostosowanych, a więc odzwierciedlających ograniczone możliwości ludzkiej pamięci bezpośredniej. Słusznie przy tym zauważa, że wiele konstrukcji o dużej głębokości ma też warianty płytsze (a przez to łatwiejsze do przetworzenia), natomiast sama głębokość może wpływać na zmianę językową (ku konstrukcjom mniej kosztownym dla pamięci). Nieciągłość jest jednym ze sposobów ograniczania głębokości struktury składniowej – w drzewie (30) rozbudowana fraza przymiotnikowa *good for the job*, stojąca po lewej stronie rzeczownika *man*, zostaje spłycona.

(30)



Ten sam mechanizm działa w strukturach typu:



„tak dobry, jakiego tylko można do tej pracy znaleźć”

interpretowanych jako mające aż trzy nieciągłości (zaznaczone przerywanymi liniami). W obu wypadkach w wyniku przesunięcia rozbudowanych określeń na prawo konstrukcja ulega spłyceciu, a tym samym zwalnia pamięć podręczną. Yngve podkreśla przy tym, że w spontanicznej mowie, do której przede wszystkim odnosi swój model, śledzenie granic nieciągłości ułatwiają wykładniki prozodyczne (por. też Sweeney, Love 1995).

Lewis (1996, s. 10) podkreśla, że propozycję tę wyróżnia dobre osadzenie w modelu przetwarzania i ustalenie ograniczenia pamięci podręcznej na podstawie niezależnych badań psycholingwistycznych. Z drugiej strony zauważa, że konstrukcje regresywne są odrzucane ze względu na przyjęty model przetwarzania, a nie na obserwacje – przetwarzanie wstępujące radzi sobie bowiem dobrze z rozgałęziającymi się w lewo strukturami.

Yngve (1998) ostatecznie uznał swoją hipotezę za nieweryfikowalną, uważając, że nie sposób stwierdzić, czy przyjęty przez niego model można stosować do działalności językowej człowieka. Potwierdzenie zyskały natomiast niektóre przewidywania hipotezy (Yngve 1998, s. 634), a dość obszerne badanie korpusowe G. Sampsona (1997) zdaje się potwierdzać rzadkość regresji.

Późniejsze badania, np. Crain, Fodor (1995), zwracają uwagę na wpływ gramatyk na strategię parsowania. L. Frazier (1995) pokazuje, że trudno jest wskazać, czy mamy do czynienia z ograniczeniami ludzkiego mechanizmu przetwarzającego język, czy ze strategią tego mechanizmu w odpowiedzi na pewne cechy samego języka. Jednocześnie dowodzi, że przyjęte w hipotezie zagłębienia drzewa binarne nie znajdują potwierdzenia w badaniach nad nabywaniem języka ani w badaniach prowadzonych wśród dorosłych (s. 153)

– często wybierana jest bardziej płaska struktura z większą liczbą rozgałęzień. Główny argument przeciw hipotezie głębokości dotyczy tego, że w rzeczywistości nadawca nie wie, ile składników będzie liczyła konstrukcja, można bowiem dołączyć dowolną liczbę składników luźnych (nie wymaganych przez centrum konstrukcji). Głębokość struktury może zatem policzyć tylko lokalnie, co nie wyjaśnia preferencji dla spłaszczania struktur (przyłączania kolejnych elementów możliwie nisko). Rozwiązanie proponowane przez badaczkę uwzględnia semantykę i uwarunkowania psychologiczne, tzn. pomińanie informacji (leksykalnie) nieistotnej lub przetworzonej.

Inni badacze (por. np. Gibson (1998); Kempen (1998); Schlesewsky et al. (2003)) przyjmują, że nieciągłe konstrukcje powierzchniowe są wyprowadzane od kanonicznie ustrukturyzowanych ciągłych struktur głębinowych. Podczas przetwarzania niepełna struktura jest tymczasowo przechowywana, póki nie zostanie uzupełniona (por. np. Sweeney, Love 1995). Jeszcze inne podejście przewiduje topologię zdania w danym języku: poszczególne elementy strukturalne znajdują swoje miejsce w strukturze linearnej, wypełniając przewidziane dla nich (przez reguły gramatyki) „przegródki” (Kempen, Harbusch 2003).

6. PODSUMOWANIE

Gdy bada się rzeczywiste teksty, trzeba wziąć pod uwagę struktury składniowe rozdzielone przez składniki do nich nie należące. Współczesne gramatyki, a także parsery, są w stanie wiele z nich zanalizować. Po części nieciągłości są motywowane pragmatycznie (przesunięcia w strukturze tematyczno-rematycznej), jednak są takie, które wynikają z reguł formowania zdań. Choć stwarzają one pewne trudności w przetwarzaniu maszynowym, a i każdy, kto uczył się niemieckiego czy zmagał się z angielskim *preposition stranding*, zaliczy je do trudności, zysk (wynikający ze spłaszczenia struktury, z dodatkowej informacji, a niekiedy z jednego i drugiego) najwyraźniej wart jest dodatkowego wysiłku. Wydaje się też, że struktury nieciągłe są istotnym argumentem za tym, że konieczna w języku linearyzacja jest w znacznym stopniu niezależna od składni.

BIBLIOGRAFIA

- Bień J.S. (1996), *Processing Polish with Metamorphosis Grammar*, ftp://ftp.mimuw.edu.pl/pub/users/jsbien/tybinga96/pol-mg. ps. Dostęp 30.04.2008.
- Blevins J. P. (1990), *Syntactic Complexity. Evidence for discontinuity and multidomination*, [rozprawa doktorska], Amherst: University of Massachusetts.
- Bloomfield L. (1933), *Language*, New York.

- Bouma G., van Noord G. (1998), *Word order constraints on verb clusters in German and Dutch*, [w:] *Complex Predicates in Nonderivational Syntax*, [pod red.] E. Hinrichs, T. Nakazawa, A. Kathol, *Syntax and Semantics*, t. 30, New York: Academic Press.
- Bröcker N. (1996), *Separating surface order and syntactic relations in a dependency grammar*, [w:] COLING-ACL 98 – Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the ACL, 10–14 August, Montreal (Canada): Morgan Kaufman Publishers.
- Bunt H., Theisingh J., van der Sloot K. (1987), *Discontinuous constituent in trees, rules, and parsing*, [w:] Proceedings of the Third Conference on European Chapter of the Association for Computational Linguistics, Copenhagen, s. 203–210.
- Chomsky N., (1995), *The logical structure of linguistic theory*. New York: Plenum Press 1955, jako wersja brudnopisowa .pdf URL <http://alpha-leonis.lids.mit.edu/chomsky/>. Dostęp 30.04.2008.
- (1957), *Syntactic Structures*, Mouton & Co.
- (1965), *Aspects of Theory of Syntax*, MIT Press (*Zagadnienia teorii składni*, tłum. I. Jakubczak, Wrocław 1982).
- (1981), *Lectures on Government and Binding*, Dordrecht.
- Covington M. A., (1990), *Parsing discontinuous constituents with dependency grammar*, „Computational Linguistics”, t. 16, s. 234–236.
- Crain S., Fodor J.D., (1995), *How can grammar help parsers*, [w:] *Natural Language Processing*, [pod red.] R. Dowty, L. Karttunen i A. Zwicky, Cambridge University Press, s. 94–128.
- Culicover P.W., Rochement M.S. (1997), *Deriving dependent right adjuncts in English*, [w:] *Rightward Movement*, Proceedings of the Tilburg Conference on Rightward Movement, Oct. 6–10, 1994, [pod red.] H. van Riemsdijk, D. LeBlanc, D. Beermann, Amsterdam: John Benjamins.
- Derwojedowa M. (1996), *Zdaniowe i niezdaniowe wypowiedniki polskie*, „Poradnik Językowy”, s. 22–36.
- (2000), *Porządek składników zdania elementarnego w języku polskim*, Elipsa.
- Duchier D., Debusmann R. (2001), *Topological dependency trees: a constraint-based account of linear precedence*, [w:] Proceedings of the 39th Annual Meeting of the ACL and 10th Conference of the European Chapter of the ACL, July 9–11, 2001, Toulouse, France, [pod red.] R. Zamparelli, Morgan Kaufmann Publishers, CD, s. 180–187.
- Eryigit G., Oflazer K. (2006), *Statistical dependency parsing for Turkish*, [w:] Proceedings of EAACL 2006 – The 11th Conference of the European Chapter of the Association for Computational Linguistics, April 2006, Trento, Italy.
- Espinal M. (1991), *The representation of disjunct constituent*, „Language”, t. 67, s. 726–762.
- Frazier L. (1995), *Syntactic complexity*, [w:] *Natural Language Processing*, [pod red.] R. Dowty, L. Karttunen, A. Zwicky, Cambridge University Press, s. 145–172.
- Gaifman H. (1965), *Dependency systems and phrase-structure systems*. „Information and Control”, t. 8, s. 304–307.
- Gibson E. (1998), *Linguistic complexity: locality of syntactic dependencies*, „Cognition”, s. 1–76.

- Gleason H.A. (1955), *An Introduction to Descriptive Linguistics*, New York: Holt.
- Haegeman L. (1996), *Introducing to Government a. Binding Theory*, USA: Oxford UK-Cambridge USA.
- Harris Z. S., (1945), *Discontinuous Morphemes*. „Language”, t. 21 (3) s. 121–127.
- (1951), *Methods in Structural Linguistics*, Chicago.
- (1957), *Co-occurrence and transformation in linguistic structure*. „Language”, t. 33 (3) s. 283–340. Dostęp 05.05.2008.
- Hays D. G. (1964), *Dependency theory: a formalism and some observations*. „Language”, t. 40(4) s. 511–525.
- Hockett C. (1958), *A course in modern linguistic*, New York: Macmillan (*Kurs językoznawstwa współczesnego*, tłum. Z. Topolińska i M. Jurkowski, Warszawa 1968).
- Hudson R. (1990), *English Word Grammar*. London: Basil Blackwell.
- Johnson M. (1985), *Parsing with discontinuous constituents*, [w:] Proceedings of the 23rd annual meeting on Association for Computational Linguistics, Chicago, 8–12 July, Illinois, USA, s. 127–132.
- Jordanska L., Иорданская Л. (1963), *О некоторых свойствах правильной синтаксической структуры*, „Вопросы языкознания” т. 4, s. 103–112.
- Kahane S., Nasr A., Rambow O. (1998), *Pseudoprojectivity: A polynomially parsable non-projective dependency grammar*, [w:] Proceedings of ACL-COLING, August 10–14, 1998, Montreal, Canada.
- Kempen G. (1998), *Sentence parsing*, [w:] *Language comprehension: A biological perspective*, [pod red.] A.D. Friederici, 1 ed., Berlin/Heidelberg: Springer, s. 213–228.
- Kempen G., Harbusch K. (2003), *Word order scrambling as a consequence of incremental sentence production*, [w:] *Mediating Between Concepts and Grammar*, [pod red.] H.T. K. Hard Holden, Berlin: Mouton De Gruyter, s. 141–164.
- Lecerf Y. (1960), *Programme des conflits, modele des conflits*, „La Traduction Automatique”, t. 4–5.
- Lewis R. L. (1996), *Inference in short-time memory: the magical number two (or three) in sentence processing*, „Journal of Psycholinguistic Research”, t. 25 (1), s. 93–115.
- McCawley J. (1982), *Parentheticals and Discontinuous Constituent Structure*, „Linguistic Inquiry”, t. 13 (1), s. 91–106.
- Mecner P. (2005), *Elementy gramatyki umysłu*, Kraków.
- Mel'čuk L. / Мельчук, И. А., *О «внутренней флексии» в индоевропейских и семитских языках*, „Вопросы языкознания”, т. 3.
- (1974), *Опыт теории лингвистических моделей «смысл ↔ текст»*, Москва.
- (1988), *Dependency Syntax: Theory and Practice*, State University of New York Press.
- Mel'čuk I. A., Pertsov N. (1987), *Surface syntax of English. A formal model within the meaning-text framework*, „Linguistic and Literary Studies in Eastern Europe”, t. 13, John Benjamins.
- Miller G. A. (1956), *The magical number seven, plus or minus two. Some limits for our capacity for processing information*, „Psychological Review”, t. 62 (2), s. 81–97. URL <http://psych.utoronto.ca/users/peterson/psy430s2001/Miller%20GA%20Magical%20Seven%20Psych%20Review%201955.pdf>. Dostęp 30.04.2008.
- Nivre J. (2004), *Constraints on non-projective dependency parsing*, [w:] Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, April, 3rd-7th, 2006 Toronto, Italy, s. 73–80.

- Nivre J., Hall J., Nilsson J. (2005), *Memory-based dependency parsing*, [w:] Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL), May 6–7, 2004, Boston, Massachusetts, [pod red.] H.T. Ng i E. Riloff, s. 49–56.
- Obrębski T. (2002), *Automatyczna analiza składniowa języka polskiego z wykorzystaniem gramatyki zależnościowej*, [rozprawa doktorska], Instytut Podstaw Informatyki PAN.
- Ojeda A. (2005), *Discontinuous constituents*, [w:] *Encyclopedia of Language and Linguistics*, Volumes 1–14, [pod red.] K. Brown, t. 3, 2nd ed., s. 624–630, Elsevier. URL http://linguistics.ucdavis.edu/People/fzojeda/linguistics-papers/discontinuous_constituents.pdf. Dostęp 30.04.2008.
- Øvrelid L., Nivre J. (2007), *When word order and part-of-speech tags are not enough – Swedish dependency parsing with rich linguistic features*, [w:] Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP), s. 447–451.
- Partee B. H., ter Meulen A., Wall R. (1990), *Mathematical Methods in Linguistics*. Dordrecht, Boston–London: Kluwer Academic Publishers.
- Pike K. (1943), *Taxemes and immediate constituents*, „Language”, t. 19/2, s. 65–82.
- Postal P. (1964), *Constituent Structure: a Study of Contemporary Models of Syntactic Description*, Bloomington.
- Rambow O., Joshi A. (1997), *A formal look at dependency grammars and phrase-structure grammars, with special consideration of word-order phenomena*, [w:] *Recent Issues in Meaning-Text Theory*, [pod red.] L. Wanner, London: Pinter, s. 167–191.
- Saloni Z., Świdziński M. (2001), *Składnia współczesnego języka polskiego*. Warszawa.
- Sampson G. (1997), *Depth in english grammar*, „Journal of Linguistics”, t. 33 (1), s. 131–151.
- Schlesewsky M., Bornkessel I., Frisch S., (2003), *The neurophysiological basis of word order variations in German*, „Brain and Language”, t. 86, s. 116–128.
- Sweeney D., Love T. (1995), *The processing of discontinuous dependencies in lanugage and music*, „Music Perception”, t. 16 (1), s. 63–78.
- Szpakowicz S. (1978), *Automatyczna analiza polskich zdań pisanych*, [rozprawa doktorska], Instytut Informatyki UW.
- Szpakowicz S., Świdziński M. (1992), *Discontinuities*, [w:] *Le mot, le mots, les bons mots. Word, words, witty words*, Le Presses de l'Universite de Montreal.
- Świdziński M. (1992), *Gramatyka formalna języka polskiego*, Warszawa.
- Tesnière L. (1959), *Elements de syntaxe structurale*, Paris.
- Wells R. S. (1947), *Immediate constituents*, „Language”, t. 23, s. 81–117.
- Wiśniewski M. (1994), *Strukturalna charakterystyka polskich wypowiedzi niezdanowych*, Toruń.
- Woliński M. (2005), *Komputerowa weryfikacja gramatyki Świdzińskiego*, [rozprawa doktorska], Instytut Podstaw Informatyki PAN.
- Yngve V. H. (1960), *A model and a hypothesis for language structure*. Rap. tech. 369, Massachusetts Institute of Technology. (*Model i hipoteza struktury języka* [w:] *Z zagadnień językoznawstwa współczesnego: gramatyka transformacyjna, teoria informacji: wybór materiałów*, wybór M. Jurkowski et al., Warszawa 1966: Państwowe Wydawnictwo Naukowe, s. 70–115).
- (1998), *Clues from the depth hypothesis*, „Computational Linguistics”, t. 24 (4), s. 633–640, URL <http://www.aclweb.org/anthology-new/J/J987J98-4005.pdf>. Dostęp 30.04.2008.