

1 Twierdzenie Herbranda

1.1 Sylwetka Herbranda

Jacques Herbrand urodził się 12 lutego 1908 roku. W latach 1928-29 przygotował rozprawę doktorską opublikowaną w Polsce w 1930 roku. Zawiera ona m.in. dowód twierdzenia o dedukcji i konstruktywny dowód twierdzenia Gödla o pełności. Twierdzenie o pełności zostało wyprowadzone z tzw. twierdzenia Herbranda, którego dowód zawiera algorytm semirozstrzygający, czy dane zdanie jest tautologią rachunku kwantyfikatorów. Wiosną 1931 roku Herbrand zasugerował Gödlowi w prywatnym liście rozszerzenie definicji (pochodzącej z pracy Gödla) klasy funkcji rekurencyjnych po to, by opisać pojęcie obliczalności. Latem 1931 roku, 27 lipca, zginął w Alpach podczas wspinaczki.

Koncepcja funkcji rekurencyjnych zasugerowana przez Herbranda została przedstawiona przez Gödla podczas wykładów Institute for Advanced Study w 1934 roku i opublikowana w notatkach z wykładu sporządzonych przez J.B. Rossera i S.C. Kleene'ego, doktorantów Alonzo Churcha.

1.2 Skolemizacja

O formule mówimy, że jest w postaci normalnej, jeżeli zaczyna się blokiem kwantyfikatorów, po którym znajduje się formuła bezkwantyfikatorowa. Dowodzi się, że każdą formułę można sprowadzić do postaci normalnej, a dokładniej

Lemat 1.1 *Dla każdej formuły φ istnieje formuła ψ w postaci normalnej taka, że równoważność $\varphi \leftrightarrow \psi$ jest prawem rachunku kwantyfikatorów (oraz tautologią). $\square$*

Będziemy teraz rozważać ustalony język pierwszego rzędu pozwalający na zapisywanie formuł. Każdej formule φ tego języka przypiszemy pewną formułę φ^* . Jeżeli φ jest formułą w postaci normalnej bez kwantyfikatorów egzystencjalnych, to $\varphi^* = \varphi$. Ponadto przyjmujemy, że

$$(\forall x_1 \forall x_2 \dots \forall x_n \exists y \varphi)^* = (\forall x_1 \forall x_2 \dots \forall x_n \varphi[y \leftarrow f(x_1, x_2, \dots, x_n)])^*,$$

gdzie f jest nowym symbolem funkcyjnym. Analogiczną definicję przyjmujemy w przypadku, gdy $n = 0$. Wtedy zamiast termu z symbolem funkcyjnym używamy stałej. Podane równości pozwalają wyliczyć φ^* dla dowolnej formuły w postaci normalnej.

Jeżeli ψ nie ma postaci normalnej, to aby wyliczyć ψ^* formułę ψ sprowadzamy do postaci normalnej φ i przyjmujemy, że $\psi^* = \varphi^*$.

Formułę ψ^* nazywamy postacią skolemowską formuły ψ .

1.3 Własność skolemizacji

Lemat 1.2 *Formuła φ jest spełniona w strukturze $\mathcal{A}$ przy wartościowaniu h wtedy i tylko wtedy, gdy w pewnym rozszerzeniu $\mathcal{A}'$ struktury $\mathcal{A}$ o interpretacje dodanych symboli funkcyjnych i stałych jest spełniona przy wartościowaniu h formuła φ^* .*

Wniosek 1.3 *Zdanie φ jest sprzeczne wtedy i tylko wtedy, gdy φ^* jest sprzeczne.*

Dowód. Przypuśćmy, że φ jest zdaniem sprzecznym, a φ^* nie jest sprzeczne. Tak więc $\mathcal{A}' \models \varphi^*$. Z poprzedniego lematu wynika, że po odrzuceniu niepotrzebnych symboli z modelu $\mathcal{A}'$ otrzymujemy model $\mathcal{A}$, w którym jest spełniona formuła φ . Zdanie φ nie jest więc sprzeczne.

Jeżeli φ^* jest sprzeczne, a φ – nie, to zgodnie z poprzednim lematem możemy przekształcić model, w którym jest spełniona formuła φ , w model, w którym jest spełniona formuła φ^* . Tak więc φ^* nie może być sprzeczna. $\square$

Wniosek 1.4 *Zdanie φ jest tautologią wtedy i tylko wtedy, gdy $\neg(\neg\varphi)^*$ jest tautologią.* $\square$

Lemat 1.5 *Formuła φ^* jest klasy Π_1 , a więc zaczyna się blokiem kwantyfikatorów ogólnych, po którym znajduje się formuła bez kwantyfikatorów.* $\square$

1.4 Modele (struktury) Herbranda

Teraz musimy mieć język z przynajmniej jedną stałą. Niech $\mathcal{T}_c$ oznacza zbiór termów stałych w tym języku. Będziemy rozważać modele z uniwersum równym $\mathcal{T}_c$.

W takich modelach symbole funkcyjne interpretujemy w specyficzny sposób, a mianowicie k -argumentowy symbol f oznacza funkcję $F : \mathcal{T}_c^k \rightarrow \mathcal{T}_c$ taką, że

$$F(t_1, \dots, t_k) = f(t_1, \dots, t_k).$$

Podobnie, stała c w takim uniwersum jest interpretowana jako term c .

Modele o uniwersum $\mathcal{T}_c$, w których stałe i symbole funkcyjne interpretujemy w podany sposób, nazywamy modelami Herbranda.

Lemat 1.6 *W dowolnym modelu Herbranda wartość termu t przy wartościowaniu h jest równa*

$$t[h] = t[x \leftarrow h(x)][y \leftarrow h(y)] \dots$$

W szczególności, wartością termu stałego t przy dowolnym wartościowaniu w modelu Herbranda jest term t . Jeżeli w jest funkcją przyporządkowującą termowi $t \in \mathcal{T}_c$ jego wartość w strukturze $\mathcal{A}$, to

$$w(t[h]) = t[w \circ h]. \quad \square$$

Przypuśćmy, że $\mathcal{A}$ jest modelem (strukturą) dla rozważanego języka. Taki model wyznacza pewien model Herbranda $\mathcal{H}_{\mathcal{A}}$. W modelu $\mathcal{H}_{\mathcal{A}}$ symbol relacyjny r jest interpretowany¹ jako relacja R taka, że

$$\langle t_1, \dots, t_k \rangle \in R \iff \mathcal{A} \models r(t_1 \dots, t_k).$$

Lemat 1.7 *Jeżeli φ jest formułą bez kwantyfikatorów, a h jest wartościowaniem zmiennych w strukturze $\mathcal{H}_{\mathcal{A}}$ (w zbiorze $\mathcal{T}_c$), to*

$$\mathcal{A} \models \varphi[w \circ h] \iff \mathcal{A} \models \varphi[x \leftarrow h(x)][y \leftarrow h(y)] \dots \iff \mathcal{H}_{\mathcal{A}} \models \varphi[h].$$

Wniosek 1.8 *Jeżeli φ jest formułą klasy Σ_1 (czyli zaczyna się blokiem kwantyfikatorów egzystencjalnych, po którym mamy formułę bez kwantyfikatorów) oraz $\mathcal{H}_{\mathcal{A}} \models \varphi$, to $\mathcal{A} \models \varphi$.*

Wniosek 1.9 *Zdanie $\neg\varphi^*$ jest tautologią wtedy i tylko wtedy, gdy zdanie $\neg\varphi^*$ jest spełniona w każdym modelu Herbranda.*

Twierdzenie 1.10 *Zdanie φ jest tautologią wtedy i tylko wtedy, gdy zdanie $\neg(\neg\varphi)^*$ jest spełnione w każdym modelu Herbranda.*

¹Taki sposób definiowania interpretacji relacji równości nie gwarantuje spełniania aksjomatów równości. Kwestie związane tymi aksjomatami są w tutaj pomijane. Dowody są słuszne dla teorii bez równości, a więc przy założeniu, że symbol równości nie jest w żaden sposób wyróżniony.

1.5 Kolejny krok

Przyjmijmy, że φ jest formułą ze zmiennymi $x_1, \dots, x_n$, a $\vec{t} = (t_1, \dots, t_n)$ jest ciągiem termów. Wtedy symbolem $\varphi(\vec{t})$ będziemy oznaczać formułę

$$\varphi[x_1 \leftarrow t_1] \dots [x_n \leftarrow t_n],$$

a więc formułę φ , w której za kolejne zmienne zostały podstawione kolejne termy z ciągu $\vec{t}$.

Twierdzenie 1.11 *Niech φ będzie formułą otwartą, a więc bez kwantyfikatorów. Zdanie $\psi = \exists x_1 \dots \exists x_n \varphi$ jest spełnione w każdym modelu Herbranda wtedy i tylko wtedy, gdy istnieją układy termów $\vec{t}_1, \dots, \vec{t}_m \in \mathcal{T}_c^n$ takie, że formuła*

$$\varphi(\vec{t}_1) \vee \dots \vee \varphi(\vec{t}_m)$$

jest tautologią (w sensie rachunku zdań, różne formuły atomowe uważamy za różne zmienne zdaniowe).

Dowód. Zauważmy, że formuła ψ wynika z podanej alternatywy. Jeżeli ta alternatywa jest tautologią, to jest spełniona w każdym modelu Herbranda, w każdym takim modelu są spełnione też wszystkie wnioski z tej alternatywy i ostatecznie, w każdym modelu Herbranda jest spełniona formuła ψ . Dowód w drugą stronę jest trudniejszy.

Ustawmy wszystkie zdania atomowe w różnowartościowy ciąg. Niech ζ_n oznacza n -te zdanie atomowe. Będziemy teraz rozważać skończone i nieskończone ciągi zero-jedynkowe. Mając nieskończony ciąg zero-jedynkowy $\alpha \in 2^N$ możemy zdefiniować model Herbranda $\mathcal{H}_\alpha$. Aby zdefiniować taki model wystarczy podać, które zdania atomowe są w nim spełnione. Przyjmujemy, że

$$\mathcal{H}_\alpha \models \zeta_n \iff \alpha(n) = 1.$$

Przy okazji zauważmy, że taki ciąg α możemy uważać za wartościowanie: faktycznie – zdań atomowych, ale także obiektów, które możemy uważać za zmienne zdaniowe i z których możemy budować formuły rachunku zdań. Zauważmy też, że w podany sposób można opisać wszystkie modele Herbranda.

Każdemu $\alpha \in 2^N$ przypiszemy teraz pewną liczbę

$$m(\alpha) = \min\{k \in N : \exists \vec{t} \in \mathcal{T}_c^n \mathcal{H}_\alpha \models \varphi(\vec{t}) \text{ oraz } \forall l > k \zeta_l \text{ nie jest podformułą } \varphi(\vec{t})\}.$$

Z założenia, że w modelu $\mathcal{H}_\alpha$ jest spełnione zdanie ψ wynika, że liczba $m(\alpha)$ jest dobrze zdefiniowana.

Liczby $m(\alpha)$ mają kilka ciekawych własności. Na przykład, jeżeli $\alpha, \beta \in 2^N$ oraz $\alpha(i) = \beta(i)$ dla wszystkich $i \leq m(\alpha)$, to $m(\alpha) = m(\beta)$. Teraz będziemy zajmować się wartością

$$\max\{m(\alpha) : \alpha \in 2^N\}$$

i – aby zbadać tę wartość – będziemy rozważać zbiór

$$X = \{(\alpha(0), \dots, \alpha(m(\alpha))) : \alpha \in 2^N\}$$

skończonych ciągów zero-jedynkowych.

Przypuśćmy, że wspomniane maksimum nie jest określone. Wtedy zbiór X zawiera dowolnie długie ciągi. Zauważmy też, że jeżeli pewien ciąg można przedłużyć do elementu X na nieskończenie wiele sposobów, to tę samą własność ma jeden z ciągów długości o 1 większej.

W ten sposób, jeżeli maksimum nie jest określone, można zdefiniować nieskończony ciąg α , którego każdy prefiks da się wydłużyć do elementu X (nawet na nieskończenie wiele sposobów). Ciąg $(\alpha(0), \dots, \alpha(m(\alpha)))$ można więc istotnie wydłużyć do ciągu $(\beta(0), \dots, \beta(m(\beta)))$ dla pewnego β . Wtedy zachodzi nierówność $m(\alpha) < m(\beta)$. Ze wspomnianej wyżej własności wynika jednak, że $m(\alpha) = m(\beta)$. Uzyskana sprzeczność dowodzi, że wspomniane maksimum jest liczbą naturalną i – co więcej – że zbiór X jest skończony.

Każdy ciąg $a \in X$ wyznacza pewien ciąg termów $\vec{t}_a \in \mathcal{T}_c^n$. Aby skonstruować ten ciąg, znajdujemy najpierw $\alpha \in 2^N$ takie, że $a = (\alpha(0), \dots, \alpha(m(\alpha)))$, a następnie ciąg $\vec{t}$ taki, że

- 1) $\mathcal{H}_\alpha \models \varphi(\vec{t})$,
- 2) formuła $\varphi(\vec{t})$ jest zbudowana ze zdań atomowych $\zeta_0, \dots, \zeta_{m(\alpha)}$.

Istnienie takiego ciągu $\vec{t}$ wynika z definicji liczby $m(\alpha)$. Przyjmujemy, że ciąg $\vec{t}_a$ jest jednym z takich ciągów, a więc ma następujące własności:

- 1) $\mathcal{H}_\alpha \models \varphi(\vec{t}_a)$,
- 2) formuła $\varphi(\vec{t}_a)$ jest zbudowana ze zdań atomowych $\zeta_0, \dots, \zeta_{m(\alpha)}$.

Mając wybrane w ten sposób ciągi termów tworzymy potrzebną nam alternatywę

$$\bigvee_{a \in X} \varphi(\vec{t}_a).$$

Pozostało dowieść, że ta alternatywa jest tautologią.

Przyjmijmy, że $k = \max\{m(\alpha) : \alpha \in 2^N\}$. Interesująca nas alternatywa jest zbudowana ze zdań atomowych $\zeta_0, \dots, \zeta_k$. Weźmy dowolne wartościowanie tych zdań. Możemy je utożsamiać z zerojedynkowym ciągiem $\alpha_0, \dots, \alpha_k$ przyjmując, że α_i jest wartością logiczną zdania ζ_i . Ten skończony ciąg rozszerzamy (jakkolwiek) do nieskończonego ciągu $\alpha \in 2^N$ (spełniającego zatem dla $i \leq k$ równości $\alpha(i) = \alpha_i$). Niech a oznacza ciąg $(\alpha(0), \dots, \alpha(m(\alpha))) \in X$. Zauważmy, że $\mathcal{H}_\alpha \models \varphi(\vec{t}_a)$. Tym samym formuła $\varphi(\vec{t}_a)$ ma przy wartościowaniu $(\alpha_0, \dots, \alpha_k)$ wartość logiczną 1. Tym bardziej rozważana alternatywa ma przy tym wartościowaniu wartość 1. Ponieważ wartościowanie zostało dowolnie wybrane, więc podana alternatywa jest tautologią. $\square$

Twierdzenie 1.12 (Herbrand) *Przypuśćmy, że φ jest zdaniem i zdanie $\neg(\neg\varphi)^*$ jest równoważne ze zdaniem $\exists x_1 \dots \exists x_n \psi$, gdzie ψ jest formułą otwartą. Zdanie φ jest tautologią wtedy i tylko wtedy, gdy istnieją układy termów $\vec{t}_1, \dots, \vec{t}_m \in \mathcal{T}_c^n$ takie, że formuła*

$$\psi(\vec{t}_1) \vee \dots \vee \psi(\vec{t}_m)$$

jest tautologią. Dokładniej, będzie ona prawdziwa bez względu na to, jakie są wartości logiczne formuł atomowych, z których jest zbudowana.

Wniosek 1.13 *Jeżeli zdanie φ jest tautologią, to zdanie $\neg(\neg\varphi)^*$ jest twierdzeniem rachunku kwantyfikatorów.*

Dowód. Jest to wniosek z twierdzenia Herbranda i twierdzenia o pełności dla rachunku zdań.

Mniej więcej tak dowodzi się twierdzenie o pełności w rachunku kwantyfikatorów, choć dowód tego twierdzenia musi być jednak subtelniejszy. $\square$

1.6 Algorytm Herbranda

Naturalny algorytm Herbranda ma następującą postać:

Dane: zdanie Ψ .

Wynik: informacja, czy zdanie Ψ jest tautologią (prawem rachunku kwantyfikatorów).

1. Daną formułę negujemy, sprowadzamy do postaci normalnej, skolemizujemy, ponownie negujemy i sprowadzamy do postaci normalnej. Przypuśćmy, że po tych przekształceniach otrzymujemy zdanie $\psi = \exists x_1 \dots \exists x_n \varphi$ (w φ nie ma kwantyfikatorów).
2. $\Phi := \text{false}$.
3. **while** Φ nie jest tautologią **do**
 - wygeneruj kolejny ciąg termów $\vec{t}$
 - $\Phi := \Phi \vee \varphi(\vec{t})$
4. przekaz informację, że Ψ jest tautologią.

Pierwsza próba implementacji tego algorytmu pochodzi z 1959 r. Paul Gilmore z IBM Research Center eksperymentował z niemal dokładnie takim algorytmem. Stosował badanie tautologii polegające na sprowadzeniu formuły Φ do koniunkcyjnej postaci normalnej i korzystaniu z faktu, że koniunkcyjna postać normalna tautologii jest koniunkcją alternatyw, w których pewna zmienna występuje dwukrotnie: jako zmienna bez negacji i zanegowana.

W 1960 roku Martin Davis i Hilary Putnam zaproponowali inny sposób badania tautologii, bardziej dostosowany do algorytmu Herbranda. Zaproponowali mianowicie, aby formułę φ nadać alternatywną postać normalną, a sprawdzanie, czy Φ jest tautologią zastąpić wyprowadzaniem sprzeczności z negacji Φ , czyli dowodzeniem Φ metodą nie wprost. Jeżeli φ ma alternatywną postać normalną, to bardzo łatwo znajdujemy koniunkcyjną postać normalną negacji Φ i możemy wyprowadzać sprzeczność z wielu założeń będących alternatywami zmiennych zdaniowych i negacji zmiennych zdaniowych. Do tego w rachunku zdań w zasadzie wystarcza reguła rezolucji

$$\frac{\alpha \vee \beta, \neg\alpha \vee \gamma}{\beta \vee \gamma}.$$

Naturalny algorytm Herbranda, nawet zaimplementowany zgodnie z propozycją Davisa i Putnama, wymaga dwóch pętli. W jednej szukamy odpowiednich podstawień, w drugiej – na przykład wyprowadzenia sprzeczności. John Alan Robinson w 1965 roku zaproponował jednocześnie szukanie potrzebnych podstawień i dowodu sprzeczności poprzez używanie specjalnej, dostosowanej do rachunku kwantyfikatorów, reguły rezolucji

$$\frac{\alpha \vee \beta, \neg\alpha' \vee \gamma}{(\beta \vee \gamma)\theta},$$

gdzie θ jest najogólniejszym unifikatorem formuł α i α' . Założeniami dowodu sprzeczności w tym przypadku są człony koniunkcji w koniunkcyjnej postaci normalnej formuły $\neg\varphi$.

Prace zmierzające do efektywnej implementacji algorytmu Herbranda zostały wykorzystane podczas tworzenia interpreterów Prologu. Interpretery te realizują obliczenia wymagane podczas dowodzenia twierdzeń metodą Herbranda, ale wymagają dodatkowego założenia, że formuła $\neg\varphi$ jest hornowska, czyli w koniunkcyjnej postaci normalnej tej formuły, każda alternatywa zawiera najwyżej jedną niezanegowaną formułę atomową.

1.7 Twierdzenie o pełności raz jeszcze

Twierdzenie 1.14 (o stałej) *Jeżeli stała c nie występuje w aksjomatach teorii T , ani w formule φ oraz $T \vdash \varphi[x \leftarrow c]$, to $T \vdash \forall x \varphi$.*

Wniosek 1.15 *Jeżeli T jest teorią, φ jest formułą w języku L , a c jest stałą spoza L , to φ można dowieść w teorii T posługując się formułami języka $L \cup \{c\}$ wtedy i tylko wtedy, gdy φ można dowieść w teorii T posługując się formułami języka L .*

Niech T będzie teorią w języku L . Dodamy do L pewien zbiór stałych i rozszerzymy zbiór aksjomatów T tworząc w ten sposób teorię T_c . Dla każdego zdania postaci $\exists x \varphi$ dodajemy specjalną stałą c i nowy aksjomat $(\exists x \varphi) \rightarrow \varphi[x \leftarrow c]$. Teorię T_c konstruujemy tak długo aż dla każdej takiej formuły będzie istniała w języku odpowiednia stała, a w teorii znajdzie się odpowiedni aksjomat. Czasem taką teorię nazywa się teorią Henkina.

Twierdzenie 1.16 *Jeżeli φ jest formułą zapisaną w języku L i $T_c \vdash \varphi$, to także $T \vdash \varphi$.*

Dowód. Twierdzenie to mówi, że w dowodach mamy prawo posługiwać się regułą opuszczania kwantyfikatora egzystencjalnego, przynajmniej w szczególnych przypadkach, gdy przesłanka jest zdaniem.

Dowodząc zdanie φ w teorii T_c korzystamy ze skończonej liczby specjalnych aksjomatów. Niech $(\exists x \psi) \rightarrow \psi[x \leftarrow c]$ będzie najpóźniej dodanym, wykorzystywanym w dowodzie, specjalnym aksjomatem, a Ψ – koniunkcją pozostałych. Na mocy twierdzenia o dedukcji

$$T \vdash ((\exists x \psi) \rightarrow \psi[x \leftarrow c]) \rightarrow (\Psi \rightarrow \varphi).$$

Możemy też założyć, że zmienna x nie występuje w φ . Wtedy x występuje najwyżej w ψ oraz

$$((\exists x \psi) \rightarrow \psi[x \leftarrow c]) \rightarrow (\Psi \rightarrow \varphi) = (((\exists x \psi) \rightarrow \psi) \rightarrow (\Psi \rightarrow \varphi))[x \leftarrow c].$$

Stąd na mocy twierdzenia o stałej otrzymujemy, że

$$T \vdash \forall x (((\exists x \psi) \rightarrow \psi) \rightarrow (\Psi \rightarrow \varphi)).$$

Korzystając jak zwykle z praw logiki możemy zauważyć, że także

$$T \vdash ((\exists x \psi) \rightarrow \exists x \psi) \rightarrow (\Psi \rightarrow \varphi)$$

i w konsekwencji,

$$T \vdash \Psi \rightarrow \varphi.$$

Na mocy zasady indukcji ze względu na liczbę potrzebnych w dowodzie aksjomatów specjalnych otrzymujemy, że

$$T \vdash \varphi. \quad \square$$

Teoria T_c ma kilka ciekawych własności. Na przykład, jeżeli $\exists x \neg \varphi$ jest zdaniem i c jest stałą specjalną dla tego zdania, to

$$T_c \vdash \varphi[x \leftarrow c] \rightarrow \forall x \varphi.$$

Tego typu własność można skomplikować. Przypuśćmy, że mamy formułę

$$\exists x \forall y \exists z \forall w \psi(x, y, z, w),$$

a a jest termem stałym. Dla tego termu istnieje stała specjalna $\tau(a)$ dla zdania

$$\exists y \neg(\exists z \forall w \psi[x \leftarrow a]).$$

Z przytoczonego spostrzeżenia otrzymujemy teraz

$$T_c \vdash \exists z \forall w \psi[x \leftarrow a][y \leftarrow \tau(a)] \rightarrow \forall y \exists z \forall w \psi[x \leftarrow a],$$

a także

$$T_c \vdash \exists z \forall w \psi[x \leftarrow a][y \leftarrow \tau(a)] \rightarrow \exists x \forall y \exists z \forall w \psi.$$

Powtórzmy to rozumowanie raz jeszcze. Dla termu stałego b (i termu a) bierzemy stałą specjalną $\sigma(a, b)$ dla zdania

$$\exists w \neg\psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b].$$

Mamy więc

$$T_c \vdash \psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b][w \leftarrow \sigma(a, b)] \rightarrow \forall w \psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b]$$

oraz

$$T_c \vdash \psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b][w \leftarrow \sigma(a, b)] \rightarrow \exists z \forall w \psi[x \leftarrow a][y \leftarrow \tau(a)].$$

Łącząc dwa udowodnione twierdzenia otrzymujemy, że

$$T_c \vdash \psi[x \leftarrow a][y \leftarrow \tau(a)][z \leftarrow b][w \leftarrow \sigma(a, b)] \rightarrow \exists x \forall y \exists z \forall w \psi.$$

Przedstawione rozumowanie, które może zostać przeprowadzone dla dowolnych formuł w postaci preneksowej klasy Σ_{2k} zostanie wykorzystane do dowodu kolejnego twierdzenia Herbranda.

Twierdzenie 1.17 (Herbrand) *Przypuśćmy, że φ jest formułą i formuła $\neg(\neg\varphi)^*$ jest równoważna formule $\exists x_1 \dots \exists x_n \psi$, gdzie ψ jest formułą otwartą (bez kwantyfikatorów). Jeżeli istnieją układy termów $\vec{t}_1, \dots, \vec{t}_m \in \mathcal{T}_c^n$ takie, że formuła*

$$\psi(\vec{t}_1) \vee \dots \vee \psi(\vec{t}_m)$$

jest tautologią, to φ jest prawem rachunku kwantyfikatorów.

Dowód. Dowód przedstawimy na przykładzie formuły klasy Σ_4 . Niech więc

$$\varphi = \exists x \forall y \exists z \forall w \psi(x, y, z, w).$$

Wtedy formuła $\neg(\neg\varphi)^*$ ma postać

$$\exists x \exists z \psi(x, f(x), z, g(x, z))$$

dla pewnych nowych symboli funkcyjnych f i g . Zakładamy więc, że istnieją układy termów stałych t_i oraz s_i takie, że formuła

$$\Phi = \bigvee_i \psi(t_i, f(t_i), s_i, g(t_i, s_i))$$

jest tautologią. Niech teraz T oznacza rachunek kwantyfikatorów, czyli teorię pozbawioną aksjomatów. Dla uproszczenia zakładamy także, że T nie zawiera nawet aksjomatów równości (dla teorii z równością są potrzebne dodatkowe konstrukcje, które nie zostaną przedstawione).

Utwórzmy teorię T_c i weźmy zdefiniowane wyżej operacje τ i σ przyporządkowujące termom stałym specjalne stałe. Formułę Φ będziemy przekształcać wielokrotnie wykonując przekształcenie następującej postaci: sprawdzamy, czy w Φ występuje podterm postaci $f(a)$ dla pewnego termu stałego a nie zawierającego ani f , ani g , i jeżeli taki podterm znajdziemy, to zastępujemy go we wszystkich miejscach stałą $\tau(a)$. Analogicznie, wszystkie wystąpienia podtermów postaci $g(a, b)$ zastępujemy stałą $\sigma(a, b)$.

Zauważmy, że przekształcenia te zachowują „bycie tautologią”. Przypuśćmy, że po wykonaniu wszystkich możliwych takich przekształceń otrzymaliśmy formułę Φ^* . Tak więc Φ^* jest tautologią.

Podczas wykonywania opisanych przekształceń w jakiś sposób zmieniały się termy t_i oraz s_i . Przyjmimy, że zostały przekształcone w a_i oraz b_i odpowiednio. Nietrudno zauważyć, że formuła Φ^* ma postać

$$\Phi^* = \bigvee_i \psi(a_i, \tau(a_i), b_i, \sigma(a_i, b_i)).$$

Zgodnie z wcześniejszymi uwagami, człony tej alternatywy implikują zdanie φ . Wobec tego, także mamy

$$T_c \vdash \Phi^* \Rightarrow \varphi.$$

Na mocy twierdzenia o pełności dla rachunku zdań dowód ma także formuła Φ^* . Tak więc φ jest twierdzeniem teorii T_c , a ponieważ nie zawiera stałych specjalnych, jest też twierdzeniem rachunku kwantyfikatorów. $\square$