6. INTERPRETABILITY

Let S and S′ be arbitrary theories. S′ is interpretable in S if, roughly speaking, the primitive concepts and the range of the variables of S′ are definable in S in such a way as to turn every theorem of S′ into a theorem of S. If, in addition every nontheorem of S′ is transformed into a nontheorem of S, then S′ is faithfully interpretable in S.

In this chapter, we assume that PA$\dashv$ T. Thus, T is essentially reflexive.

§1. Interpretability. Let S and S′ be arbitrary theories. By a *translation* (of the language of S′ into the language of S) we understand a function t on the set of formulas (of S′) into the set of formulas (of S) for which there are formulas $\eta_0(x)$, $\eta_S(x,y)$, $\eta_+(x,y,z)$, $\eta_\times(x,y,z)$ and a formula $\mu_t(x)$ such that t satisfies the following conditions for all formulas φ, ψ, $\xi(x)$:

(*) $t(x = y) := x = y,$
$t(x = 0) := \eta_0(x),$
$t(Sx = y) := \eta_S(x,y),$
$t(x + y = z) := \eta_+(x,y,z),$
$t(x \times y = z) := \eta_\times(x,y,z),$
$t(\neg\varphi) := \neg t(\varphi),$
$t(\varphi \wedge \psi) := t(\varphi) \wedge t(\psi),$
$t(\exists x\xi(x)) := \exists x(\mu_t(x) \wedge t(\xi(x))).$

(Here x, y, z are arbitrary variables.) We assume that $\forall$ and the connectives $\vee$, $\rightarrow$, $\leftrightarrow$ are defined in terms of $\exists$, $\neg$, $\wedge$. Note that t, on the formulas for which it is defined by the above conditions, is uniquely determined by its values on atomic formulas together with the formula $\mu_t(x)$.

So far $t(\varphi)$ is only defined provided that φ is written in a certain "normal form". For example, t is not defined on the formula $x + 0 = y$. But this formula is equivalent to $\exists z(z = 0 \wedge x + z = y)$ and t is defined on this formula so we can set $t(x + 0 = y) := t(\exists z(z = 0 \wedge x + z = y))$. Similarly, for any formula φ not already on "normal form", replace φ in some canonical way by φ^* on "normal form" (logically equivalent to φ) and set $t(\varphi) := t(\varphi^*)$. It follows, for example, that $t(\forall x\xi(x))$ is equivalent to $\forall x(\delta(x) \rightarrow t(\xi(x)))$. Clearly t is a primitive recursive function.

The translation t is an *interpretation in* S iff

(**) $S\vdash \exists x\mu_t(x),$
$S\vdash \exists x(\mu_t(x) \wedge \forall y(\mu_t(y) \rightarrow (\eta_0(y) \leftrightarrow y = x))),$
$S\vdash \forall x\big(\mu_t(x) \rightarrow \exists y(\mu_t(y) \wedge \forall z(\mu_t(z) \rightarrow (\eta_S(x,z) \leftrightarrow z = y)))\big),$
$S\vdash \forall xy\big(\mu_t(x) \wedge \mu_t(y) \rightarrow \exists z\big(\mu_t(z) \wedge \forall u(\mu_t(u) \rightarrow (\eta_*(x,y,u) \leftrightarrow u = z))\big)\big)$, $* = +, \times$.

Thus, t is an interpretation in S iff $S\vdash t(\varphi)$ for every logically valid sentence φ.

t is an *interpretation of* S′ *in* S, t: S′$\leq$ S, iff $S\vdash t(\varphi)$ for every φ such that $S'\vdash \varphi$. S′