Universes over Frege Structures

Reinhard Kahle

Abstract

We investigate *universes* axiomatized as sets with natural closure conditions over Frege structures. In the presence of a natural form of induction, we obtain a theory of proof-theoretic strength Γ_0.

1 Introduction

Frege structures were introduced by Aczel in [Acz80] as a semantical concept to introduce a notion of sets by means of a *partial truth predicate*. This approach is closely related to prior work of Scott [Sco75] and was originally developed for questions around Martin-Löf's type theory. In [Bee85, Ch. XVII] Beeson gave a formalization of Frege structures as a truth theory over applicative theories.

Applicative theories go back to Feferman's systems of explicit mathematics introduced in [Fef75, Fef79]. These systems provide a logical basis for functional programming. The basic theory for which Frege structures are defined is the basic theory of operations and numbers TON, introduced and studied in [JS95]. It comprises total combinatorial logic and arithmetic.

The notion of Frege structures is formalized in slightly different ways (cf. e.g. [FM87, Tur90, Can93, HK95]). The most comprehensive exposition can be found in the forthcoming book *Logical Frameworks for Truth and Abstraction* [Can9x] of Cantini where he presents also many extensions and applications.

In this paper we discuss the extension of Frege structures by *universes*. Universes are well-known from Martin-Löf's type theory as a natural extension of formal systems. They are introduced as sets with canonical closure conditions (cf. [ML84]). Proof-theoretically they are investigated in [Acz77, Fef82] by making use of fixed point theories. In the context of applicative theories, universes were first introduced by Marzetta in [Mar93]. Starting from a two sorted theory of types and names, he defined the theory UTN. In UTN universes are types closed under set-theoretical operations, like elementary comprehension and join.

Here we start with the concept of Frege structures and introduce the theory FSU by defining universes as propositional functions which are closed under formation of formulae. The closure conditions are similar to those of truth levels in the theory

TLR of Cantini [Can9x]. This theory is also based on Frege structures, but needs a new sort of objects (levels) outside the applicative basis in order to formalize a hierarchy of truth predicates. With our notion of universes over Frege structures we stay in the applicative framework without giving up the expressive power of TLR, in particular, the possibility to define fixed points of positive arithmetical operators. In contrast to UTN, this allows an easy embedding of the above fixed point theories into FSU.

In the presence of a natural form of induction FSU has the proof-theoretical ordinal Γ_0, the so-called limit of predicativity. The lower bound is presented by embedding the fixed point theory FP_0 of Avigad [Avi9x] into FSU, while the upper bound follows from a translation in Cantini's theory TLR.

The paper is organized as follows: in the next section we introduce the basic theory TON. In the third section we define Frege structures, following the axiomatization of Cantini. In section 4 we give the axiomatization of universes which are proof-theoretically studied in the last section.

2 The theory of total operations and numbers

As introduced in [JS95] the basic theory of total operations and numbers (TON) is formulated in L. L comprises individual variables $x, y, z, u, v, f, g, h, \ldots$ (possibly with subscripts), individual constants $0, \mathsf{k}, \mathsf{s}, \mathsf{p}, \mathsf{p}_0, \mathsf{p}_1, \mathsf{s}_\mathsf{N}, \mathsf{p}_\mathsf{N}, \mathsf{d}_\mathsf{N}$, a binary function symbol $\cdot$ for term application, and the relation symbols $=$ and N.

Terms $(r, s, t, r_1, s_1, t_1, \ldots)$ and formulae $(\varphi, \psi, \ldots)$ are defined by $\neg$, $\wedge$ and $\forall$ as usual, starting from the atomic formulae $t = s$, and $\mathsf{N}(t)$.

In the following we write $s\,t$ for $(s \cdot t)$ with the convention of association to the left, $\vee$, $\rightarrow$ and $\exists$ are defined as usual and we set t' as $\mathsf{s}_\mathsf{N}\, t$.

The logic of TON is classical first order predicate logic with equality. The non-logical axioms of TON include:

I. **Partial combinatory algebra.**

(1) $\mathsf{k}\, x\, y = x$,

(2) $\mathsf{s}\, x\, y\, z = x\, z\, (y\, z)$.

II. **Pairing and projection.**

(3) $\mathsf{p}_0\, (\mathsf{p}\, x\, y) = x \wedge \mathsf{p}_1\, (\mathsf{p}\, x\, y) = y$,

III. **Natural numbers.**

(4) $\mathsf{N}(0) \wedge \forall x.\mathsf{N}(x) \rightarrow \mathsf{N}(x')$,

(5) $\forall x.\mathsf{N}(x) \rightarrow x' \neq 0 \wedge \mathsf{p}_\mathsf{N}\, (x') = x$,

(6) $\forall x. \mathsf{N}(x) \wedge x \neq 0 \to \mathsf{N}(\mathsf{p_N}\, x) \wedge (\mathsf{p_N}\, x)' = x$.

IV. **Definition by cases on N.**

(7) $\mathsf{N}(v) \wedge \mathsf{N}(w) \wedge v = w \to \mathsf{d_N}\, x\, y\, v\, w = x$,

(8) $\mathsf{N}(v) \wedge \mathsf{N}(w) \wedge v \neq w \to \mathsf{d_N}\, x\, y\, v\, w = y$.

It is well-known that we can introduce in TON a notion of λ-abstraction and prove the recursion theorem (cf. [Bar85]):

Proposition 1 For every variable x and every term t of L there exists a term $\lambda x.t$ of L whose free variables are those of t, excluding x, so that

$$\mathrm{TON} \vdash (\lambda x.t)\, x = t.$$

Proposition 2 There is a term rec of L so that

$$\mathrm{TON} \vdash \mathsf{rec}\, f = f\, (\mathsf{rec}\, f).$$

Remark 3 TON is equivalent to the partial theory BON, introduced in [FJ93], plus the axiom of totality (Tot): $\forall x, y.\, x\, y \downarrow$.

3 Frege structures over TON

Syntactically Frege structures can be considered as a truth theory over applicative theories. Among the several, slightly different axiomatizations (e.g. [Bee85, HK95]) we follow the formalization of Cantini, presented as the theory NMT in [Can93] or MF^- in [Can9x].

The theory FON, Frege structures for TON, is formulated in the language $\mathcal{L}_F$. $\mathcal{L}_F$ is the language L of TON extended by the new relation symbol T and new individual constants $\dot{=}$, $\dot{\mathsf{N}}$, $\dot{\neg}$, $\dot{\wedge}$, $\dot{\forall}$ and $\dot{\mathsf{T}}$.[1]

The axioms of FON are the axioms of TON extended to the new language plus the following axioms:

I. **Closure under prime formulae of TON**

(1) $x = y \leftrightarrow \mathsf{T}(\dot{=}\, x\, y)$

(2) $\neg x = y \leftrightarrow \mathsf{T}(\dot{\neg}\, (\dot{=}\, x\, y))$

(3) $\mathsf{N}(x) \leftrightarrow \mathsf{T}(\dot{\mathsf{N}}\, x)$

[1] Cantini does not use new individual constants like $\dot{=}$, but introduces them via encoding. This guarantees some independence properties which are missing here, but are not necessary for the lower bounds of our theories.

(4) $\neg\mathsf{N}(x) \leftrightarrow \mathsf{T}(\dot{\neg}\,(\dot{\mathsf{N}}\,x))$

II. **Closure under composed formulae**

(5) $\mathsf{T}(x) \leftrightarrow \mathsf{T}(\dot{\neg}\,(\dot{\neg}\,x))$

(6) $\mathsf{T}(x) \wedge \mathsf{T}(y) \leftrightarrow \mathsf{T}(\dot{\wedge}\,x\,y)$

(7) $\mathsf{T}(\dot{\neg}\,x) \vee \mathsf{T}(\dot{\neg}\,y) \leftrightarrow \mathsf{T}(\dot{\neg}\,(\dot{\wedge}\,x\,y))$

(8) $(\forall x.\mathsf{T}(f\,x)) \leftrightarrow \mathsf{T}(\dot{\forall}\,f)$

(9) $(\exists x.\mathsf{T}(\dot{\neg}\,(f\,x))) \leftrightarrow \mathsf{T}(\dot{\neg}\,(\dot{\forall}\,f))$

III. **Self-reference**

(10) $\mathsf{T}(x) \leftrightarrow \mathsf{T}(\dot{\mathsf{T}}\,x)$

(11) $\mathsf{T}(\dot{\neg}\,x) \leftrightarrow \mathsf{T}(\dot{\neg}\,(\dot{\mathsf{T}}\,x))$

IV. **Consistency**

(12) $\neg(\mathsf{T}(x) \wedge \mathsf{T}(\dot{\neg}\,x))$

In straightforward manner we define for every $\mathcal{L}_F$-formula φ a term $\dot{\varphi}$ by replacing $=, \mathsf{N}, \mathsf{T}, \neg, \wedge$, and $\forall$ by the corresponding individual constants $\dot{=}, \dot{\mathsf{N}}, \dot{\mathsf{T}}, \dot{\neg}, \dot{\wedge}$, and $\dot{\forall}$, respectively (using prefix notation). This definition is compatible with term substitution, i.e. we have $\dot{\varphi}[t/x] \equiv \dot{\overbrace{\varphi[t/x]}}$.

Since we cannot put the negation outside the scope of T in the axioms of self-reference, the formulae which are T-positive become of special interest. We define T-positive and T-negative formulae simultaneously:

Definition 4

1. $t = s$, $\mathsf{N}(t)$, $\neg t = s$ and $\neg\mathsf{N}(t)$ are T-positive as well as T-negative.
2. $\mathsf{T}(t)$ is T-positive; $\neg\mathsf{T}(t)$ is T-negative.
3. If φ is T-positive (T-negative), so $\neg\varphi$ is T-negative (T-positive).
4. If φ and ψ are T-positive (T-negative), so also $\varphi \wedge \psi$.
5. If φ is T-positive (T-negative), so also $\forall x.\varphi$.

By induction on the formation of formulae we get the following main result for Frege structures [Can9x, Th. 8.8.(ii)]:

Proposition 5 For T-positive φ we have

$$\mathrm{FON} \vdash \mathsf{T}(\dot{\varphi}) \leftrightarrow \varphi.$$

We say "t is true" for $\mathsf{T}(t)$ and "t is false" for $\mathsf{T}(\dot\neg\, t)$, and introduce *propositions* in the form of objects which are true or false, and *propositional functions* (also called *sets* or *classes*) which return propositions for any argument:

$$\mathsf{P}(x) :\Leftrightarrow \mathsf{T}(x) \vee \mathsf{T}(\dot\neg\, x) \quad \text{and} \quad \mathsf{Pf}(f) :\Leftrightarrow \forall x.\mathsf{P}(f\,x).$$

As an easy consequence we get by diagonalizing $\dot\neg$:

Lemma 6 FON $\vdash \neg\forall x.\mathsf{P}(x)$.

Now we introduce a notion of abstraction which, by the previous proposition, is useful for T-positive φ:

$$\{x|\varphi\} :\equiv \lambda x.\dot\varphi$$

The element relation is defined as follows:

$$t \in s :\Leftrightarrow \mathsf{T}(s\,t).$$

Moreover, we use $t \notin s$ for $\neg\mathsf{T}(s\,t)$. In general this is weaker than the statement $\mathsf{T}(\dot\neg\,(s\,t))$, but for propositional functions s they are equivalent.

As a natural form of complete induction on natural numbers we study the so-called *proposition induction*:

$$(\mathsf{P}\text{-}\mathsf{I}_{\mathsf{N}}) \qquad \mathsf{Pf}(f) \wedge \mathsf{T}(f\,0) \wedge (\forall x.\mathsf{N}(x) \wedge \mathsf{T}(f\,x) \rightarrow \mathsf{T}(f\,(x'))) \rightarrow \forall x.\mathsf{N}(x) \rightarrow \mathsf{T}(f\,x)$$

4 Universes over Frege structures

Universes as introduced in Martin-Löf's type theory or in theories of types and names are objects with natural set-theoretic closure condition. In our context we have to close universes under formation of formulae. This approach is closely related to Cantini's theory of truth level where he introduces predicates with similar closure conditions. This system is presented below to estimate the upper bound of our theory.

Before we start to axiomatize universes, we introduce a binary relation which is used as an order structure on universes. It expresses in some sense that the *course-of-T-value* of a function is contained in another.[2]

Definition 7 $f \sqsubset g :\Leftrightarrow \forall x.(x \in f \rightarrow f\,x \in g) \wedge (x \notin f \rightarrow \dot\neg\,(f\,x) \in g)$.

The power of this relation is that, if we have $f \sqsubset g$, we can translate a T-negative statement about f in a T-positive one about g. If f is a propositional function and g a universe in the sense axiomatized below, we say g is *above* f.

The language of FSU is the language $\mathcal{L}_F$ of FON extended by the new relation symbol U.

The axioms of FSU are the axioms of FON extended to the new language plus the following axioms:

[2]The relation is a counterpart of the relation $\dot\in$ of UTN and the level ordering $\prec$ of TLR.

I. **Basics**

(1) $\mathsf{U}(u) \to \mathsf{Pf}(u)$

(2) $\mathsf{U}(u) \to \forall x.x \in u \to \mathsf{T}(x)$

II. **Closure under prime formulae of TON**

(3) $\mathsf{U}(u) \to \forall x.\mathsf{N}(x) \leftrightarrow \dot{\mathsf{N}}\, x \in u$

(4) $\mathsf{U}(u) \to \forall x.\neg\mathsf{N}(x) \leftrightarrow \dot{\neg}\,(\dot{\mathsf{N}}\, x) \in u$

(5) $\mathsf{U}(u) \to \forall x, y.x = y \leftrightarrow \dot{=}\, x\, y \in u$

(6) $\mathsf{U}(u) \to \forall x, y.x \neq y \leftrightarrow \dot{\neg}\,(\dot{=}\, x\, y) \in u$

III. **Closure under composed formulae**

(7) $\mathsf{U}(u) \to \forall x, y.x \in u \wedge y \in u \leftrightarrow \dot{\wedge}\, x\, y \in u$

(8) $\mathsf{U}(u) \to \forall x, y.\dot{\neg}\, x \in u \vee \dot{\neg}\, y \in u \leftrightarrow \dot{\neg}\,(\dot{\wedge}\, x\, y) \in u$

(9) $\mathsf{U}(u) \to \forall x.x \in u \leftrightarrow \dot{\neg}\,(\dot{\neg}\, x) \in u$

(10) $\mathsf{U}(u) \to ((\forall x.f\, x \in u) \leftrightarrow \dot{\forall}\, f \in u)$

(11) $\mathsf{U}(u) \to ((\exists x.\dot{\neg}\,(f\, x) \in u) \leftrightarrow \dot{\neg}\,(\dot{\forall}\, f) \in u)$

IV. **Self-reference**

(12) $\mathsf{U}(u) \to \forall x.x \in u \leftrightarrow u\, x \in u$

(13) $\mathsf{U}(u) \to \forall x.\dot{\neg}\, x \in u \leftrightarrow \dot{\neg}\,(u\, x) \in u$

(14) $\mathsf{U}(u) \to \forall x.x \in u \leftrightarrow \dot{\mathsf{T}}\, x \in u$

(15) $\mathsf{U}(u) \to \forall x.\dot{\neg}\, x \in u \leftrightarrow \dot{\neg}\,(\dot{\mathsf{T}}\, x) \in u$

V. **Order structure**

(16) $\mathsf{U}(u) \wedge \mathsf{U}(v) \wedge u \sqsubset v \to \forall x.x \in u \to x \in v$

(17) $\mathsf{U}(u) \wedge \mathsf{U}(v) \to u \sqsubset v \vee u = v \vee v \sqsubset u$

VI. **Limit**

(18) $\forall f.\mathsf{Pf}(f) \to \exists u.\mathsf{U}(u) \wedge f \sqsubset u$

Let us briefly give a informal discussion of the axioms:
I.(1) expresses that universes are propositional functions, i.e. sets in the sense of Frege Structures. I.(2) means that only *true* objects are elements of a universe.
The groups II. and III. axiomatize the closure of universes under positive connectives and quantifiers starting from (positive and negative) equality and N. But a universe is not closed under negated truth.

A weaker closure condition for dotted negation is contained in the self-reference axioms. Self-reference comes in two forms, one for the universes itself, one for the truth predicate T (axioms IV.).
V. deals with the relation $\sqsubset$ for universes and states that the elements of a lower universe are contained in the upper one, and that the order is total.
The limit axiom is the crucial one of the theory. It guarantees the existence of some universe (in fact finitely many).

Now we state two basic propositions in FSU. First it follows from consistency and I.(2) that universes are "consistent". Second, we prove that universes are linearly ordered by $\sqsubset$.

Lemma 8 FSU $\vdash \mathsf{U}(u) \to \forall x.\neg(x \in u \wedge \dot{\neg}\, x \in u)$

Lemma 9 Universes are linearly ordered by $\sqsubset$.

Proof: Totality follows by V.(17), transitivity by the definition of $\sqsubset$ and V.(16). It holds also in a little more general form, which we will need later:

$$\mathsf{U}(v) \wedge \mathsf{U}(u) \wedge \mathsf{Pf}(f) \wedge u \sqsubset v \wedge f \sqsubset u \to f \sqsubset v. \tag{trans}$$

Irreflexivity ($\mathsf{U}(u) \to \neg u \sqsubset u$) is proved as follows: Assume $\mathsf{U}(u)$ and $u \sqsubset u$. Then we have

$$\begin{array}{lll} & \mathsf{U}(u) & \\ \to & \mathsf{Pf}(u) & \\ \to & \forall x.\mathsf{T}(u\,x) \vee \mathsf{T}(\dot{\neg}\,(u\,x)) & \\ \to & \forall x.\mathsf{T}(u\,x) \vee \mathsf{T}(u\,(\dot{\neg}\,(u\,x))) & \text{Assumption} \\ \to & \forall x.\mathsf{T}(u\,x) \vee \mathsf{T}(u\,(\dot{\neg}\,x)) & \text{IV.(13)} \\ \to & \forall x.\mathsf{T}(x) \vee \mathsf{T}(\dot{\neg}\,x) & \text{I.(2)} \end{array}$$

But this contradicts lemma 6. □

5 Proof-theoretic investigations

We show that FSU+(P-$\mathsf{I_N}$) has the proof-theoretical ordinal Γ_0, the so-called limit of predicativity. For the lower bound[3] we use a fixed point theory introduced by Avigad, where a fixed point corresponds to a universe. The upper bound is determined by Cantini's theory TLR, where our universes are interpreted as truth levels.

[3] For the notion of proof-theoretic reduction used here we refer to [Fef88].

5.1 The lower bound

We determine the lower bound of FSU plus (P-I_{N}) using a fixed point theory which is introduced by Avigad in [Avi9x] to study the relation between ATR_0 and $\widehat{\mathrm{ID}}_{<\omega}$. This theory, called FP_0, is formulated in the language L_2 of second order arithmetic and comprises Peano Arithmetic PA, where induction is restricted to sets:[4]

(Set-Ind) $$0 \in X \land (\forall x.x \in X \to x' \in X) \to \forall x.x \in X.$$

The crucial axiom for FP_0 is the following fixed point axiom for a positive arithmetic operators φ, i.e. it is an arithmetic formula in which Y occurs only positively:

(FP) $$\exists Y.\forall x.x \in Y \leftrightarrow \varphi(x, Y),$$

Hereby it is important that φ can have set parameters. If φ has no set parameter (FP) is exactly arithmetical comprehension.

Avigad [Avi9x, Th. 3.1] has proved that (FP) is equivalent to the axiom (ATR) of arithmetical transfinite recursion of ATR_0:

Lemma 10 $\mathrm{ATR}_0 \equiv \mathrm{FP}_0$.

Now we interpret FP_0 in FSU plus proposition induction. The crucial point is hereby that a fixed point of FP_0 correspond to a universe of FSU.

Proposition 11 $\mathrm{FP}_0 \trianglelefteq \mathrm{FSU} + (\mathsf{P}\text{-}\mathrm{I}_{\mathsf{N}})$.

Proof: The first order part is interpreted by letting range the object variables over N. This way the verification of the axioms of PA (without induction) is straightforward (cf. [FJ93]).
We interpret second order variables by propositional functions and the element relation of L_2 by the one defined in FSU, i.e. $t \in Z$ is translated by $\mathsf{Pf}(z) \to \mathsf{T}(z\,t)$. So (Set-Ind) becomes an instance of (P-I_{N}). It remains the fixed point axiom (FP):

$$\exists Y.\forall x.x \in Y \leftrightarrow \varphi(x, Y, Z_1, \ldots, Z_n),$$

where φ is a positive arithmetical operator with free variables x, Y and set parameters $Z_1, \ldots, Z_n$, $n \geq 0$.
To verify (FP) we generalize the interpretation of $\widehat{\mathrm{ID}}_1$ in FON, as it is presented in [Can9x, § 10A]: For the required fixed point we choose essentially the recursion theoretic fixed point of $\dot{\varphi}$[5].
Since φ is in general not T-positive (it can contain subformulae $\neg t \in Z_i$), we do not have $\mathsf{T}(\dot{\varphi}) \leftrightarrow \varphi$. Here the use of universes is crucial in order to capture negative occurrences of T in φ.

[4] Avigad allows induction for arbitrary Σ^0_1 formulae, but this form is reduceable by the following fixed point axiom to the more natural form of set induction.

[5] In the following we do not distinguish explicitly between a L_2-formula φ and its translation.

Let the propositional functions y and z_i be the translations of Y and Z_i, resp. Now $n+1$ applications of the limit axiom for y and z_i yield universes $u_0, u_1, \ldots, u_n$, so that we have $y \sqsubset u_0$ and $z_i \sqsubset u_i$ for all $1 \leq i \leq n$. By lemma 9 we know that universes are linearly ordered by $\sqsubset$. So the limit axiom applied to the "highest" universe u_j, $0 \leq j \leq n$, yields a further universe v so that $u_i \sqsubset v$ for all $0 \leq i \leq n$. Moreover we have by (trans) (from the proof of lemma 9) or all $1 \leq i \leq n$

$$y \sqsubset v \quad \text{and} \quad z_i \sqsubset v.$$

Now we set p_φ as $\mathsf{rec}\,(\lambda y, x.\dot{\varphi})$ and choose for the required fixed point of φ the term $\lambda x.v\,(p_\varphi\, x)$. This is a propositional function since v, being a universe, is a propositional function. It remains to show that it has the fixed point property:

$$\begin{array}{rcl} x \in \lambda x.v\,(p_\varphi\, x) & \leftrightarrow & \mathsf{T}(v\,(p_\varphi\, x)) \\ & \leftrightarrow & \mathsf{T}(v\,((\lambda y, x.\dot{\varphi})\, p_\varphi\, x)) \\ & \leftrightarrow & \mathsf{T}(v\,(\dot{\varphi}[p_\varphi/y])) \\ & \leftrightarrow & \mathsf{T}(v\,(\overset{\cdot}{\overbrace{\varphi[p_\varphi/y]}})) \end{array}$$

Now we prove by induction on the formation of φ,

$$\mathsf{T}(v\,(\overset{\cdot}{\overbrace{\varphi[p_\varphi/y]}})) \leftrightarrow \varphi[\lambda x.v\,(p_\varphi\, x)/y]$$

which yields the required equivalence:

$\varphi \equiv t = s, \neg t = s, \mathsf{N}(t)$ or $\neg\mathsf{N}(t)$. Here y does not occur in φ and we get the equivalence from the closure of universes under prim formulae of TON.

$\varphi \equiv \mathsf{T}(y\, t)$. So $\varphi[\lambda x.v\,(p_\varphi\, x)/y]$ is $\mathsf{T}(v\,(p_\varphi\, t))$ and the equivalence to $\mathsf{T}(v\,(\dot{\mathsf{T}}\,(p_\varphi\, t)))$ follows from the self-reference axiom IV.(14).

$\varphi \equiv \mathsf{T}(z_i\, t)$. $\mathsf{T}(z_i\, t) \leftrightarrow \mathsf{T}(v\,(\dot{\mathsf{T}}\,(z_i\, t)))$ follows with self-reference from $z_i \sqsubset v$.

$\varphi \equiv \neg\mathsf{T}(z_i\, t)$. $\neg\mathsf{T}(z_i\, t) \leftrightarrow \mathsf{T}(v\,(\dot{\neg}\,(\dot{\mathsf{T}}\,(z_i\, t))))$ follows again from $z_i \sqsubset v$.

φ is a composed formula. The equivalence follows from the induction hypothesis by closure of universes under composed formulae. $\square$

Remark 12 The above proposition yields Γ_0, the proof-theoretic ordinal of ATR_0, as lower bound of $\mathrm{FSU}+(\mathsf{P}\text{-}\mathrm{I}_{\mathsf{N}})$. If we define FP in analogy to ATR as the theory FP_0 with full induction, we get $\mathrm{ATR} \equiv \mathrm{FP}$, and the proof of proposition 11 yields also Γ_{ε_0}, the proof-theoretic ordinal of ATR, as a lower bound for FSU with induction for arbitrary formulae.

5.2 The upper bound

We determine the proof-theoretic upper bound of our theory by interpreting it in Cantini's theory TLR of *truth with levels and reflection*. It has been introduced and studied in [Can9x, Ch. VIII]. It expands TON by a notion of truth levels, which approximate the truth predicate T.

The language $\mathcal{L}_V$ of TLR is the language $\mathcal{L}_F$ of FON extended by

1. a new sort of variables for levels $i_0, i_1, \ldots$ (L-variables);
2. a new unary function symbol LT;
3. three new binary predicates $\preceq$, $=_l$, and V (for level ordering, level identity and local truth resp.).

L-terms are the L-variables, $\mathcal{L}_V$-terms are defined as the least set closed under the following clauses:

1. Every individual variable and individual constant is a $\mathcal{L}_V$-term,
2. if j is a L-term, so $\mathsf{LT}(j)$ is a $\mathcal{L}_V$-term,
3. if t and s are $\mathcal{L}_V$-terms, so is $(t \cdot s)$ a $\mathcal{L}_V$-term.

As new prime formulae we have $i \preceq j$, $i =_l j$ and $V(i,t)$, where i, j are L-terms and t a $\mathcal{L}_V$-term. In addition, for building composed formulae we have the closure under quantification over levels, i.e. $\forall j.\varphi$ is a formula, if j is a L-variable and φ a formula.

As abbreviations we use $\mathsf{T}_i(t)$ for $V(i,t)$ and $\dot{\mathsf{T}}_i$ for $\mathsf{LT}(i)$. We drop the index for the level equality and write $i \prec j$ for $\neg i = j \wedge i \preceq j$.

The theory TL^- of truth with levels is defined over a two sorted classical predicate calculus, where the levels are the second sort. The non-logical axioms comprise those of TON extended to the new $\mathcal{L}_V$-terms and formulae and the following ones:

I. **Projectibility**

(1) $\forall i.\forall j.\mathsf{LT}(i) = \mathsf{LT}(j) \rightarrow i = j$

II. **Local truth axioms**

(2) $x = y \leftrightarrow \mathsf{T}_i(\dot{=}\, x\, y)$
(3) $\neg x = y \leftrightarrow \mathsf{T}_i(\dot{\neg}\, (\dot{=}\, x\, y))$
(4) $\mathsf{N}(x) \leftrightarrow \mathsf{T}_i(\dot{\mathsf{N}}\, x)$
(5) $\neg \mathsf{N}(x) \leftrightarrow \mathsf{T}_i(\dot{\neg}\, (\dot{\mathsf{N}}\, x))$
(6) $\mathsf{T}_i(x) \rightarrow \mathsf{T}_i(\dot{\mathsf{T}}_i\, x)$
(7) $\mathsf{T}_i(\dot{\neg}\, x) \rightarrow \mathsf{T}_i(\dot{\neg}\, (\dot{\mathsf{T}}_i\, x))$

(8) $\mathsf{T}_i(x) \leftrightarrow \mathsf{T}_i(\dot\neg\,(\dot\neg\,x))$

(9) $\mathsf{T}_i(x) \wedge \mathsf{T}_i(y) \leftrightarrow \mathsf{T}_i(\dot\wedge\,x\,y)$

(10) $\mathsf{T}_i(\dot\neg\,x) \vee \mathsf{T}_i(\dot\neg\,y) \leftrightarrow \mathsf{T}_i(\dot\neg\,(\dot\wedge\,x\,y))$

(11) $(\forall x.\mathsf{T}_i(f\,x)) \leftrightarrow \mathsf{T}_i(\dot\forall\,f)$

(12) $(\exists x.\mathsf{T}_i(\dot\neg\,(f\,x))) \leftrightarrow \mathsf{T}_i(\dot\neg\,(\dot\forall\,f))$

(13) $\neg(\mathsf{T}_i(x) \wedge \mathsf{T}_i(\dot\neg\,x))$ Local consistency

III. **Level axioms**[6]

(14) $\forall i.\forall j.\forall k.i \preceq i \wedge (i \preceq j \wedge j \preceq k \rightarrow i \preceq k) \wedge (i \preceq j \wedge j \preceq i \rightarrow i = j)$

(15) $\forall i.\forall j.i \prec j \vee i = j \vee j \prec i$

(16) $\forall i.\exists j.i \prec j$

IV. **Connection axioms**

(17) $\mathsf{T}(x) \rightarrow \exists i.\mathsf{T}_i(x)$ Limit.1

(18) $\mathsf{T}_i(x) \rightarrow \mathsf{T}(x)$ Limit.2

(19) $i \preceq j \wedge \mathsf{T}_i(x) \rightarrow \mathsf{T}_j(x)$ Persistence

(20) $\mathsf{T}_i(\dot{\mathsf{T}}\,x) \leftrightarrow \mathsf{T}_i(x)$ Localization.1

(21) $\mathsf{T}_i(\dot\neg\,(\dot{\mathsf{T}}\,x)) \leftrightarrow \mathsf{T}_i(\dot\neg\,x)$ Localization.2

(22) $i \prec j \rightarrow (\mathsf{T}_j(\dot{\mathsf{T}}_i\,x) \vee \mathsf{T}_j(\dot\neg\,(\dot{\mathsf{T}}_i\,x)))$ Potential Completeness

(23) $\mathsf{T}_j(\dot{\mathsf{T}}_i\,x) \rightarrow i \preceq j \wedge \mathsf{T}_i(x)$ Positive soundness

(24) $\mathsf{T}_j(\dot\neg\,(\dot{\mathsf{T}}_i\,x)) \rightarrow (i = j \wedge \mathsf{T}_i(\dot\neg\,x)) \vee (i \prec j \wedge \neg\mathsf{T}_i(x))$ Negative soundness

The theory TLR^- is TL^- plus the reflection principle:

(REF) $$(\forall x.\exists i.\mathsf{T}_i(y\,x)) \rightarrow \exists i.\forall x.\mathsf{T}_i(y\,x).$$

Like in FON we can introduce propositions and propositional functions (classes in terms of Cantini), but now at every level:

$$\mathsf{P}_i(x) :\Leftrightarrow \mathsf{T}_i(x) \vee \mathsf{T}_i(\dot\neg\,x) \quad \text{and} \quad \mathsf{Pf}_i(f) :\Leftrightarrow \forall x.\mathsf{P}_i(f\,x).$$

As for FON we can prove for every level i:

Lemma 13 $\mathrm{TLR}^- \vdash \neg\forall x.\mathsf{P}_i(x)$.

[6]In the original version of TLR Cantini demands that $\preceq$ is only a partial order. But it is easy to check that all his arguments for the upper bound of TLR also work for the linear order, axiomatized here (cf. [Can9x, Fact 39.2. and Ch. X]). See also [Mar94, p. 74].

We get the full theory TLR by adding *local number theoretic induction* (LIND) to TLR^-:

$$\mathsf{Pf}_i(f) \wedge \mathsf{T}_i(f\,0) \wedge (\forall x.\mathsf{N}(x) \wedge \mathsf{T}_i(f\,x) \rightarrow \mathsf{T}_i(f\,(x'))) \rightarrow \forall x.\mathsf{N}(x) \rightarrow \mathsf{T}_i(f\,x)$$

In contrast to FSU which is defined over FON, TLR is defined over TON and the axioms of FON are provable in TLR [Can9x, lemma 37.3 and 37.8, th. 37.9]:

Proposition 14 For every $\mathcal{L}_F$-formula φ we have:

$$\mathrm{FON} \vdash \varphi \quad \Rightarrow \quad \mathrm{TLR}^- \vdash \varphi,$$
$$\mathrm{FON} + (\mathsf{P}\text{-}\mathsf{I}_\mathsf{N}) \vdash \varphi \quad \Rightarrow \quad \mathrm{TLR} \vdash \varphi.$$

Proof: The statements follows by the connection axioms. Merely for the verification of closure under quantification (REF) is necessary, thus FON is not yet contained in TL^-. For the interpretation of $(\mathsf{P}\text{-}\mathsf{I}_\mathsf{N})$, we show that $\mathsf{Pf}(f)$ implies $\mathsf{Pf}_i(f)$ for some level i [Can9x, lemma 37.8.(iv)]:
Let us assume $\mathsf{Pf}(f)$, i.e. $\forall x.\mathsf{T}(f\,x) \vee \mathsf{T}(\dot{\neg}\,(f\,x))$. By Limit.1 of TLR^- we get

$$\forall x.\exists i.\mathsf{T}_i(f\,x) \vee \mathsf{T}_i(\dot{\neg}\,(f\,x)).$$

Now (REF) (with $y := \lambda z.\dot{\neg}\,(\dot{\wedge}\,(f\,z)\,(\dot{\neg}\,(f\,z)))$) and the local truth axioms yield

$$\exists i.\forall x.\mathsf{T}_i(f\,x) \vee \mathsf{T}_i(\dot{\neg}\,(f\,x)).$$

That is $\exists i.\mathsf{Pf}_i(f)$. So $(\mathsf{P}\text{-}\mathsf{I}_\mathsf{N})$ is provable by (LIND). $\square$

For the proof-theoretic ordinal of TLR we have [Can9x, th. 58.1.(i)]:

Theorem 15 $|\mathrm{TLR}| = \Gamma_0$.

For the interpretation of FSU+$(\mathsf{P}\text{-}\mathsf{I}_\mathsf{N})$ in TLR we need the following technical lemma, which follows straightforward by the limit axioms:

Lemma 16 $\mathrm{TL}^- \vdash \mathsf{T}(\mathsf{T}_i\,x) \leftrightarrow \mathsf{T}_i\,x$

Now the translation works by identifying universes with truth levels:

Proposition 17 $\mathrm{FSU} + (\mathsf{P}\text{-}\mathsf{I}_\mathsf{N}) \trianglelefteq \mathrm{TLR}$

Proof: We interpret $\mathsf{U}(u)$ by $\exists i.u = \dot{\mathsf{T}}_i$.
The axioms of FON are already contained in TLR^- by prop. 14.
The translation of axiom I.(1) is provable by potential completeness and Limit.2, I.(2) by lemma 16 and Limit.2.
The groups II. and III. of axioms of FSU follow from the local truth axioms of TLR^- using lemma 16. For the self-reference IV. we need also soundness IV.(23) and (24), and localization IV.(20) and (21).

The relation $\sqsubset$ in FSU corresponds to $\prec$ in TLR^- in the sense, that we can prove in TLR^- using positive and negative soundness, and lemma 13:

$$u = \mathsf{T}_i \wedge v = \mathsf{T}_j \wedge u \sqsubset v \to i \prec j.$$

So the axioms V follow with persistence and the level axiom III.(15).
We present the verification of the limit axiom of FSU in more detail: Like in the proof of prop. 14, $\mathsf{Pf}(f)$ implies $\exists i.\mathsf{Pf}_i(f)$, i.e. $\exists i.\mathsf{T}_i(f\,x) \vee \mathsf{T}_i(\dot{\neg}\,(f\,x))$. It follows by local consistency that $\mathsf{T}_i(f\,x)$ is a consequence of $\mathsf{T}(f\,x)$ and $\mathsf{T}_i(\dot{\neg}\,(f\,x))$ of $\mathsf{T}(\dot{\neg}\,(f\,x))$. So we have:

$$\exists i.\forall x.(\mathsf{T}(f\,x) \to \mathsf{T}_i(f\,x)) \wedge (\mathsf{T}(\dot{\neg}\,(f\,x)) \to \mathsf{T}_i(\dot{\neg}\,(f\,x))).$$

With lemma 16 we can replace T_i by "a universe" u and get:

$$\exists u.\exists i.u = \mathsf{T}_i \wedge \forall x.(\mathsf{T}(f\,x) \to \mathsf{T}(u\,(f\,x))) \wedge (\mathsf{T}(\dot{\neg}\,(f\,x)) \to \mathsf{T}(u\,(\dot{\neg}\,(f\,x)))).$$

This is exactly the translation of $\exists u.\mathsf{U}(u) \wedge f \sqsubset u$.
So we have proved UFON $\trianglelefteq$ TLR^-.
The remaining induction principle is verifiable in the same way as in prop. 14. $\square$

References

[Acz77] Peter Aczel. The strength of Martin Löf's intuitionistic Type Theory with one universe. In Seppo Miettinen and Jonko Väänänen, editors, *Proceedings of the Symposiums on Mathematical Logic in Oulo 1974 and in Helsinki 1975*. Report No. 2 from the Dept. of Philosophy, University of Helsinki, 1977.

[Acz80] Peter Aczel. Frege Structures and the Notion of Proposition, Truth and Set. In J. Barwise, H. Keisler, and K. Kunen, editors, *The Kleene Symposium*, pages 31–59. North-Holland, 1980.

[Avi9x] Jeremy Avigad. On the Relationship between ATR_0 and $\widehat{ID}_{<\omega}$. 199x. Preprint.

[Bar85] Hendrik Barendregt. *The Lambda Calculus*. North Holland, Amsterdam, revised edition, 1985.

[Bee85] Michael Beeson. *Foundations of Constructive Mathematics*. Ergebnisse der Mathematik und ihrer Grenzgebiete; 3.Folge, Bd. 6. Springer, Berlin, 1985.

[Can93] Andrea Cantini. Extending the first-order Theory of Combinators with Self-Referential Truth. *Journal of Symbolic Logic*, 58(2), June 1993.

[Can9x] Andrea Cantini. *Logical Frameworks for Truth and Abstraction*. in preparation, 199x. Preprint September, 1993.

[Fef75] Solomon Feferman. A Language and Axioms for explicit Mathematics. In J. Crossley, editor, *Algebra and Logic*, volume 450 of *Lecture Notes in Mathematics*, pages 87–139. Springer, 1975.

[Fef79] Solomon Feferman. Constructive Theories of Functions and Classes. In M. Boffa, D. van Dalen, and K. McAloon, editors, *Logic Colloquium 78*, pages 159–224. North–Holland, Amsterdam, 1979.

[Fef82] Solomon Feferman. Iterated Inductive Fixed-Point Theories: Application to Handcock's Conjecture. In G. Metakides, editor, *Patras Logic Symposion*, pages 171–196. North–Holland, Amsterdam, 1982.

[Fef88] Solomon Feferman. Hilbert's Program relativized: Proof-theoretical and foundational Reductions. *Journal of Symbolic Logic*, 53:364–384, June 1988.

[FJ93] Solomon Feferman and Gerhard Jäger. Systems of explicit mathematics with non-constructive μ-operator. Part I. *Annals of pure and applied logic*, 65(3):243–263, December 1993.

[FM87] Robert Flagg and John Myhill. Implication and Analysis in Classical Frege Structures. *Annals of pure and applied logic*, 34:33–85, 1987.

[HK95] Susumu Hayashi and Satoshi Kobayashi. A New Formulation of Feferman's System of Functions and Classes and its Relation to Frege Structures. *International Journal of Foundations of Computer Science*, 6(3):187–202, September 1995.

[JS95] Gerhard Jäger and Thomas Strahm. Totality in applicative theories. *Annals of Pure and Applied Logic*, 74:105–120, 1995.

[Mar93] M. Marzetta. Universes in the theory of types and names. In E. Börger et al., editor, *Computer Science Logic '92*, volume 702 of *Lecture Notes in Computer Science*, pages 340–351. Springer, Berlin, 1993.

[Mar94] Markus Marzetta. *Predicative Theories of Types and Names*. PhD thesis, Universtiät Bern, Institut für Informatik und angewandte Mathematik, 1994.

[ML84] Per Martin-Löf. *Intuitionistic Type Theory*. Bibliopolis, Napoli, 1984.

[Sco75] Dana Scott. Combinators and Classes. In C. Böhm, editor, *λ-Calculus and Computer Science Theory*, volume 37 of *Lecture Notes in Computer Science*, pages 1–26. Springer, 1975.

[Tur90] Raymond Turner. *Truth and Modality for Knowledge Representation*. Pitman, 1990.