

**Guilherme Araújo Cardoso**

**REPERCUSSÕES CONTEMPORÂNEAS DO PARADOXO DO MENTIROSO:**

**Tarski e Kripke**

**Belo Horizonte**

**Faculdade de Filosofia e Ciências Humanas**

**2010**

**Guilherme Araújo Cardoso**

**REPERCUSSÕES CONTEMPORÂNEAS DO PARADOXO DO MENTIROSO:  
Tarski e Kripke**

Dissertação apresentada ao curso de Mestrado do Departamento de Filosofia da Faculdade de Filosofia e Ciências Humanas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Filosofia.

Linha de pesquisa: Lógica e Filosofia da Ciência

Orientador: Túlio Roberto Xavier de Aguiar

**Belo Horizonte**  
**Faculdade de Filosofia e Ciências Humanas**  
**2010**

## **Folha de aprovação**

*Para Chris*

## **Agradecimentos**

Ao professor Túlio Aguiar, pela imprescindível orientação, pela disponibilidade, e pela confiança.

Aos professores Ernesto Perini, Abílio Rodrigues e Elaine Pimentel pelos cursos e pelos grupos de estudos que me instruíram e motivaram.

Ao departamento de Filosofia da UFMG e a todos os seus funcionários.

Ao CNPQ, pelo apoio financeiro.

Aos meus amigos e companheiros de filosofia: Antônio, Daniel De Luca, Eduardo Dayrell, Felipe, Henrique, Marco Aurélio, Marco Antônio, Paulinho, Roberta, Stener, Thiago Galery.

A Alex Lara, pelo companheirismo, pela amizade, e por me tornar parte de sua família.

A Rodrigo Barros, pelas conversas fundamentais e sugestões enriquecedoras, e por acompanhar com sincero interesse o meu trabalho.

A Thiago Chaves, grande amigo, caro interlocutor, filósofo.

A todos da minha família, especialmente minhas tias Elizabete e Marta.

A Ana Lúcia, minha mãe, por toda dedicação, carinho e apoio.

A Chris, por me encantar com o mais fascinante de todos os enigmas, e por decifrar para mim todos os outros.

*In Memoriam:* a Ademir Ferreira Cardoso, Miro, meu pai, pelas eternas boas lembranças.

*The point of philosophy is to start with something so simple as not to seem worth stating, and to end with something so paradoxical that no one will believe it.* (Russel, B. "The Philosophy of Logical Atomism" 1918, p.193).

## **Resumo**

O objetivo desta dissertação é discutir os problemas levantados pelo Paradoxo do Mentiroso de uma maneira geral. Vamos apresentar e comparar as principais vias de solução para este paradoxo: a Solução Ortodoxa e a Solução Alternativa. Ambas partem do diagnóstico de Tarski. Segundo este diagnóstico, o paradoxo é uma consequência de duas características essenciais: linguagens semanticamente fechadas e a preservação da lógica clássica. A Solução Ortodoxa mantém a lógica clássica, mas precisa abrir a semântica da linguagem. A Solução Alternativa mantém as linguagens semanticamente fechadas, mesmo que isto implique alguma espécie de restrição à lógica clássica. Tentaremos mostrar que, de ambos os lados, o que temos são resultados que podem ser pensados no máximo como parcialmente satisfatórios. Mostraremos um quadro geral das soluções, mas levando em conta principalmente as teorias de Tarski e Kripke.

**Palavras Chave: Paradoxo do Mentiroso, Linguagens Semanticamente Fechadas, Princípio de Bivalência, Princípio do Terceiro Excluído.**

## **Abstract**

The purpose of this dissertation is to discuss the problems raised by the Liar Paradox in general. We are going to present and compare the two main ways of solution to this paradox: The Orthodox Solution and the Alternative Solution. Both of them start from Tarski's diagnosis. According to this diagnosis the paradox is a consequence of two essential features: The semantically closed languages and the preservation of classical logic. The Orthodox Solution keeps classical logic but it has to open the semantic of the language. The Alternative Solution keeps the language semantically closed even if it implies some kind of restriction on classical logic. We will try to show that from both these sides what we have are just results that can, at most, be thought as partially satisfactory. We will show a general picture of the solutions, but mainly considering Tarski and Kripke's theories.

**Key Words: Liar Paradox, Semantically Closed Languages, Principle of Bivalence, Principle of Excluded Middle.**

## Símbolos

$\sim$	Negação
$\neg$	Negação Forte
$\rightarrow$	Condicional
$\leftrightarrow$	Bicondicional
$\forall(x)$	Quantificador Universal
$\exists(x)$	Quantificador Existencial
$F(x), G(x), H(x,y)$	Predicados e Relações
$f(x), g(x)$	Funções Sentenciais
$p, q, r$	Variáveis Sentenciais
$A, B, C$	Fórmulas
$x, y, z$	Variáveis individuais
$Tr(x)$	$x$ é verdadeira
$Fs(x)$	$x$ é falsa
$\langle \dots \rangle$	Nomes Estruturais Descritivos
$S_1$	Extensão de um predicado
$S_2$	Antiextensão de um predicado
$\models$	Consequência Semântica
$\{ \dots \}$	Conjunto
$\subseteq$	Subconjunto
$\subset$	Subconjunto próprio
$\in$	Relação de Pertencimento
$\langle \dots \rangle$	n-uplas ordenadas
$\cup$	Operação União
$\cap$	Operação Interseção
$\cup$	Conjunto União
$\emptyset$	Conjunto Vazio
<b>D</b>	Domínio
$>$	Relação “maior que”
$\geq$	Relação “maior ou igual a”
<b>PB</b>	Princípio de Bivalência
<b>PTE</b>	Princípio do Terceiro Excluído
<b>PBR</b>	Princípio de Bivalência Reforçado
<b>(T)</b>	Esquema T
<b>PM</b>	“esta própria sentença é falsa”
<b>PMR</b>	“esta própria sentença não é verdadeira”
<b>CV</b>	“esta própria sentença é verdadeira”

## Sumário

<b>Introdução.....</b>	<b>1</b>
<b>Capítulo 1</b>	
<b>O Paradoxo do Mentiroso.....</b>	<b>5</b>
<b>1.1</b> Paradoxos.....	5
<b>1.2</b> Círculos Viciosos.....	11
<b>1.3</b> Paradoxo do Mentiroso.....	14
<b>1.4</b> O que esperar de uma solução?.....	20
<b>Capítulo 2</b>	
<b>A Solução Ortodoxa.....</b>	<b>24</b>
<b>2.1</b> Tarski.....	24
2.1.1 A definição semântica do conceito de verdade.....	24
2.1.2 O Diagnóstico de Tarski.....	33
<b>2.2</b> A Solução Ortodoxa.....	40
<b>2.3</b> Vantagens e Desvantagens.....	44
<b>Capítulo 3</b>	
<b>A Solução Alternativa.....</b>	<b>51</b>
<b>3.1</b> Interlúdio sobre <i>truth value gaps</i> .....	51
<b>3.2</b> A Solução de Kripke para o Paradoxo do Mentiroso.....	62
3.2.1 Explicação informal da solução de Kripke.....	62
3.2.2 Explicação formal da solução de Kripke.....	72
3.2.3 Resultados da solução de Kripke.....	79
<b>3.3</b> Vantagens e Desvantagens.....	82
<b>Conclusão.....</b>	<b>90</b>
<b>Bibliografia.....</b>	<b>98</b>

## Introdução

A linguagem natural, ou linguagem ordinária, cotidiana, apresenta características muito interessantes e curiosas. Além do fato de que podemos expressar praticamente qualquer coisa nesta linguagem, pelo menos qualquer coisa que pudesse ser expressa em alguma linguagem artificial, a riqueza desta linguagem também é responsável pela derivação de inquietantes e complicados paradoxos. O comportamento dos conceitos semânticos e a presença indisputável de círculos viciosos (envolvendo estes conceitos) levam a embaraçosas situações quanto ao seu uso ou às suas definições. De tal maneira que, se construirmos uma linguagem e pudermos reproduzir formalmente nela as características da linguagem natural, sua consistência é imediatamente ameaçada pela derivação dos paradoxos.

Estes paradoxos são de tal modo ameaçadores, que o máximo de esforço realizado no sentido de eliminá-los tem se revelado sempre insatisfatório, sendo que algumas tentativas de solução chegam mesmo a aumentar o grau de complexidade e fatalidade destes paradoxos. Nossa inquietação, entretanto, não se limita apenas à aparente conclusão de que não podemos construir uma linguagem que espelhe tais traços característicos da linguagem natural. Claro que esta conclusão já seria suficientemente limitadora. Mas, além disso, a situação em que estamos parece nos induzir a pensar que não temos sequer uma boa compreensão destes conceitos tão básicos e importantes para vários empreendimentos filosóficos, lógicos e científicos.

Tomemos o conceito de verdade como exemplo. Este é um conceito cercado por intuições e por regras (indisputáveis e mesmo triviais) que regulam seus usos na linguagem natural e determinam seu significado. Mas, embora o conceito de verdade seja bastante comum na linguagem utilizada para discussões filosóficas, científicas e lógicas, é muito mais comum encontrar teorias que tomaram este conceito como primitivo do que aquelas que tentaram defini-lo. O conceito de verdade, contudo, em uma linguagem que funcione minimamente como a linguagem natural, deriva um paradoxo bastante ameaçador, o Paradoxo do Mentiroso. As tentativas de solução deste paradoxo não apenas mostram como é complicado construir uma linguagem razoavelmente expressiva que não permita sua derivação, mas mostram também que tentar resolvê-lo sempre envolve abrir mão de alguma intuição que julgávamos indisputável. Este paradoxo acaba produzindo uma sombra de dúvidas a respeito do

significado mesmo do conceito de verdade e das intuições que supúnhamos acompanhá-lo, de forma que não sabemos mais se podemos afirmar as coisas que antes julgávamos tão óbvias e triviais. O Paradoxo do Mentiroso e as dúvidas que ele levanta a respeito do comportamento e do significado do conceito de verdade na linguagem natural são, de maneira resumida, o que pretendemos discutir neste trabalho.

Não se pode dizer, entretanto, que haja um corpo de textos exclusivamente e precisamente direcionados apenas ao Paradoxo do Mentiroso. O que acontece é que as principais referências não compartilham sempre o mesmo problema. Assim, enquanto Tarski (que é talvez o grande ponto de partida para uma discussão a respeito do Mentiroso) está preocupado exclusivamente com a construção de linguagens formais, os autores que saem diretamente em defesa de uma solução para as ocorrências do paradoxo na linguagem natural tendem a ignorar alguns resultados formais importantes a respeito da formulação do paradoxo, e acabam respondendo a versões muito simplificadas do problema. Alguns autores costumam classificar a solução de Kripke como voltada exclusivamente para a linguagem natural. Tal classificação não está totalmente correta, já que Kripke discute diretamente os resultados obtidos pela solução de Tarski, tentando mostrar de que maneira poderíamos construir linguagens formais semanticamente fechadas. A vantagem da teoria de Kripke, tanto do ponto de vista metodológico quanto do ponto de vista filosófico, contudo, é que esperamos que a discussão possa ser ampliada abrangendo a linguagem natural e as linguagens formais, e isto é possível a partir da sua teoria.

Com o intuito de tratar o problema na sua maior abrangência e levar em conta as considerações de Tarski sem que isto signifique tratar problemas da linguagem natural de uma maneira ingenuamente formal, tentamos preservar os escopos das teorias bem explicitados. Para tanto, fizemos um esforço em mostrar adaptações do escopo de uma teoria para outra nos momentos em que se fazia necessário compará-las.

É comum os autores criticarem a teoria de Tarski como uma solução incompatível para a linguagem natural. Mas ele sequer propõe uma solução para os paradoxos da linguagem natural. Afim de não cometer este mesmo equívoco, tentamos deixar claro que Tarski não propõe tal coisa, mas estendemos os seus resultados para a linguagem natural tal como parece pressuposto nas insinuações fornecidas pela literatura. A esta solução, pautada na extensão dos resultados de Tarski, chamamos (seguindo Kripke) de Solução Ortodoxa. Também generalizamos as Soluções Alternativas, embora tenhamos dado ênfase à solução de Kripke, que, por sua vez, compartilha poucos pontos com

outras soluções que chamamos de alternativas. A razão para isto é que de fato o texto de Kripke é a primeira defesa genuína de uma linguagem semanticamente fechada, já que ele fornece modelos que efetivamente estabelecem o fechamento semântico como resultado. Por outro lado, é interessante mostrar que há outras formas de interpretar as restrições à lógica clássica, e também as consequências destas outras interpretações para o Paradoxo do Mentiroso. Por isso, não tratamos apenas da solução de Kripke.

Com o objetivo de apresentar o Paradoxo do Mentiroso de uma maneira abrangente, e daí, formular uma organização do quadro de soluções a partir da Solução Ortodoxa e da Solução Alternativa, estruturamos o texto da seguinte maneira. O primeiro capítulo apresenta o Paradoxo do Mentiroso. Começamos discutindo o que são paradoxos e qual a sua importância na Filosofia, propomos algumas definições, e mencionamos os paradoxos que envolvem círculos viciosos de uma maneira geral. O Paradoxo do Mentiroso é então formulado em algumas de suas várias versões. Por fim, tentamos discutir alguns pontos metodológicos a respeito da solução de paradoxos.

No segundo capítulo, apresentamos a Solução Ortodoxa. A Solução Ortodoxa é formulada a partir do diagnóstico de Tarski. Portanto, em primeiro lugar, falamos um pouco do projeto de Tarski e dos resultados que ele estabeleceu. Prosseguimos retirando o seu diagnóstico e as consequências para o paradoxo. A seguir, estendemos sua teoria para o que Kripke chama de Solução Ortodoxa. Depois de mostrar como funciona a solução, discutimos as vantagens e desvantagens desta solução.

No terceiro capítulo apresentamos a Solução Alternativa. Nele discutimos alguns pontos de diferentes propostas alternativas, pontos importantes para melhor apreciar a solução de Kripke. A seguir, apresentamos a teoria de Kripke, e então, discutimos as vantagens e desvantagens da Solução Alternativa.

Na conclusão, retomamos as vantagens e desvantagens que argumentamos a respeito das soluções Ortodoxa e Alternativa. Partindo disto, comparamos as duas vias de solução, e tentamos deixar bem sublinhado o dilema a que nos submetemos para decidir sobre uma destas vias de solução.

Algumas ressalvas devem ser feitas tanto pela natureza do trabalho quanto pelas condições incipientes do autor. Em primeiro lugar, é importante avisar que não tratamos a parte formal das discussões com o rigor que elas exigem. Em particular, não apresentamos as provas dos teoremas que enunciamos, não discutimos vários pontos que decorrem destas teorias para as linguagens formais, e também não alegamos que demos aos resultados semânticos o rigor adequado. É necessário ter em conta que este

não é um trabalho de lógica, e que, de todo modo, um trabalho que preservasse todo o cuidado com as considerações formais, apresentando todas as provas necessárias, certamente não caberia nos limites estipulados para este trabalho. Por outro lado, tentamos cercar as discussões filosóficas e teóricas que decorrem destes pontos, embora obviamente não pretendemos argumentar que fizemos um trabalho exaustivo no que diz respeito a elas. Muitos autores podem ter sido relegados. Há uma imensa bibliografia sobre o Mentiroso. Entretanto, julgamos que o recorte que fizemos aqui ocupa um lugar central na discussão, e tentamos articulá-lo da melhor maneira.

# Capítulo 1

## O Paradoxo do Mentiroso

### 1.1 Paradoxos

*A paradox? You mean one of those things that could destroy the universe?*

(Marty McFly em *Back to the Future*, part II)

O sentido usual do termo “paradoxo” é bastante difuso. Em alguns contextos, por exemplo, dizemos que algo é paradoxal quando envolve contradição, porém “contradição” também não é um termo que usamos de maneira precisa cotidianamente. “*a ama e odeia b*” é uma sentença que, na linguagem cotidiana, pode ser dita como envolvendo uma contradição. De forma um pouco mais precisa, entretanto, uma contradição é uma relação que se dá entre sentenças: quando a verdade de uma sentença implica a falsidade de outra e vice-versa, não podem ser ambas verdadeiras ou ambas falsas. Se *a* tem sentimentos complexos e mistos por *b* sob vários aspectos, “*a ama b*” e “*a odeia b*” podem ser ambas verdadeiras. Mas neste caso ainda é comum dizer que *a* tem sentimentos paradoxais. Por outro lado, nem todo uso preciso de contradição é algo que associamos a um paradoxo. Se uma contradição se segue por um raciocínio claramente incorreto, não dizemos que ocorre um paradoxo, mas antes um erro, um engano.

Em outros contextos, dizemos que algo é paradoxal no sentido de que é um caminho sem saída, um impasse, um enigma sem solução. O *stalemate* do jogo de Xadrez é uma situação que ilustra bem esse sentido de paradoxo. O *rei* de um dos jogadores deve fazer um movimento, e não está sob *xequê* (quando uma peça de um dos jogadores ameaça eliminar o *rei* de seu oponente), mas qualquer um de seus possíveis movimentos resultaria em *suicídio* (quando a peça de um jogador é colocada em uma posição que permite a alguma peça do adversário eliminá-la do jogo), pela disposição das peças do oponente. Como, pela regra, o *rei* nunca pode *suicidar-se*, não lhe é permitido fazer qualquer movimento, e como, assim, um dos jogadores não pode mais

jogar, o jogo encerra com empate (*stalemate*). O *rei* neste caso encontra-se em uma situação paradoxal: ele deve fazer um movimento, mas não lhe é permitido mover-se para nenhuma das posições disponíveis.

Costumamos dizer ainda que alguns objetos e eventos são paradoxais, no sentido de que são absurdos. Imagine um edifício com um conjunto complexo de escadas tal que um lance ascendente de escadas leva do primeiro ao segundo andar e outro lance igualmente ascendente leva do segundo ao primeiro novamente. Este objeto é paradoxal, porque é um absurdo<sup>1</sup>. Um sujeito que viaja no tempo, voltando ao passado e matando um de seus ancestrais, gera um paradoxo, pois seria um absurdo que ele existisse no futuro para empreender tal jornada. De forma geral, paradoxos são absurdos insuperáveis, geram resultados inconcebíveis, qualquer suposição da sua ocorrência acompanha um constrangimento de sentido, um constrangimento que nos leva a temer pelo nosso próprio discernimento.

Originalmente, o termo “paradoxo” vem do grego e significa algo como: além (contrário) da crença (opinião). Mas como podemos ver, seu sentido não é restrito a esta tradução. Não queremos esgotar os diferentes usos do termo paradoxo na linguagem cotidiana. Vamos nos ater a um uso bastante específico, que é aquele do contexto da filosofia da lógica. Neste caso podemos apresentar uma boa definição de paradoxo: uma conclusão aparentemente inaceitável derivada por raciocínios aparentemente aceitáveis a partir de premissas aparentemente aceitáveis<sup>2</sup>. Ou seja, um paradoxo ocorre quando somos forçados a aceitar uma conclusão que julgamos falsa, pela aceitação de regras de inferências e premissas que julgamos respectivamente válidas e verdadeiras. Neste caso, somos levados a desconfiar da verdade de nossas premissas, da validade de nossas inferências, ou mesmo da falsidade da conclusão.

O Paradoxo do Movimento de Zenão é um bom exemplo: suponha que Aquiles pretende se movimentar do ponto  $m$  ao ponto  $n$ , percorrendo uma linha reta (distância  $l$ ) de forma a alcançar a Tartaruga. Para ele percorrer a distância  $l$  deve antes percorrer  $l/2$ . Para percorrer  $l/2$  deve percorrer  $l/4$ , mas antes;  $l/8$ , e assim por diante, infinitamente (já que as distâncias são sempre infinitamente divisíveis). Assim, Aquiles não pode chegar até  $n$ , pois para isso precisa percorrer uma distância infinita (o que é impossível, porque ele nunca pode terminar), por menor que seja a distância entre  $m$  e  $n$ , ou seja, não pode

---

<sup>1</sup> Um apreciador dos quadros de M.C. Escher está acostumado com esse tipo de figura. Normalmente, alguém que conhece o tipo de obra que Escher produz, ao ser apresentado a uma de suas obras, procura, em primeiro lugar, o que há de absurdo na figura.

<sup>2</sup> Cf. Sainsbury (1995, p. 1).

sequer sair do ponto  $m$ . Logo, o movimento é impossível. Ora, sabemos que o movimento ocorre, portanto, não estamos dispostos a assumir que é impossível. Mas então, o que pode estar errado?<sup>3</sup> Todos tendem a concordar que a distância pode ser dividida um número infinito de vezes, resultando em um número infinito de distâncias entre  $m$  e  $n$ . Mas também parece aceitável que para percorrer uma distância  $l$  qualquer é necessário percorrer as suas partes. Mas se uma distância é composta de infinitas partes, um indivíduo que inicia o percurso jamais termina o percurso, ou seja, somos levados pelo argumento à inaceitável conclusão de que o movimento é impossível.

Vale mencionar ainda outra definição semelhante de paradoxo: um paradoxo ocorre quando temos um conjunto de enunciados individualmente aceitáveis, porém coletivamente inconsistentes<sup>4</sup>. Podemos colocar o seguinte exemplo:

- (1) Um indivíduo com um bilhão de fios de cabelos não é careca.
- (2) Um indivíduo com zero fios de cabelos é careca.
- (3) Se um indivíduo com  $n$  fios de cabelos não é careca, então um indivíduo com  $n - 1$  fios de cabelos também não é careca.

O conjunto  $\{(1), (2), (3)\}$  é inconsistente, já que (1) e (3) derivam:

- (4) Um indivíduo com zero fios de cabelos não é careca.

Mas (4) e (2) são contraditórias. Os enunciados (1), (2) e (3) individualmente não implicam contradição, e poderíamos mesmo sustentar consistentemente  $\{(1), (2)\}$ ,  $\{(1), (3)\}$  ou  $\{(2), (3)\}$ , mas nunca  $\{(1), (2), (3)\}$ .

Esta última definição apresenta algo importante: a caracterização de paradoxo a partir da noção de consistência. Em lógica, normalmente, um conjunto de sentenças é consistente se para toda sentença ou ela ou a sua negação não pertence ao conjunto das sentenças que podem ser derivadas daquele conjunto, ou seja, se não podemos derivar dele uma sentença e a sua negação (o que seria uma contradição). Um conjunto de sentenças é inconsistente se permite a derivação de contradições. Se assumirmos uma postura clássica em relação à lógica, então, uma contradição é o que melhor exemplifica uma conclusão inaceitável. Assim, quando um paradoxo tem como conclusão uma contradição, ele satisfaz as duas definições de paradoxo que apresentamos aqui. Satisfaz a primeira definição, porque é um argumento que leva por premissas e inferências

---

<sup>3</sup> As soluções tradicionais a este paradoxo têm questionado um passo inferencial bastante sutil, que pressupõe uma concepção atualista do infinito. Da premissa de que qualquer distância é infinitamente divisível (que **podemos** dividir infinitamente) não se segue que há (**atualmente**) infinitas distâncias entre quaisquer duas distâncias. Neste caso vemos que a forma de desmontar o paradoxo é invalidar parte do raciocínio utilizado no argumento.

<sup>4</sup> Cf. Rescher (2001, p. 6).

aceitáveis a uma contradição (uma conclusão inaceitável), e satisfaz a segunda definição, porque suas premissas, juntamente com as sentenças que expressam as regras inferenciais utilizadas e a conclusão, é um conjunto de sentenças individualmente aceitáveis, mas que coletivamente derivam contradição, portanto, são coletivamente inconsistentes.

Uma discussão importante que há a respeito de paradoxos em filosofia da lógica é a relação entre paradoxos, falácias e enigmas. Na medida em que paradoxos se apresentam como problemas para os quais alguma articulação das premissas e das inferências deve ser considerada, em princípio, eles se assemelhariam aos enigmas e falácias. Nas falácias e nos enigmas de lógica também somos desafiados a encontrar qual parte do argumento foi equivocadamente aceita ou escondida, devemos reconsiderar as partes do argumento de modo a desvendar a resposta.

Mas então, o que distingue um paradoxo de um enigma? A distinção é meramente quantitativa, ou qualitativa? Alguns autores, como Raymond Smullyan, por exemplo, parecem apostar em uma espécie de tese gradualista, assumindo no máximo uma distinção de graus de complexidade. Smullyan produziu diversos trabalhos<sup>5</sup>, nos quais sempre começa por explorar enigmas simples e jogos de raciocínio que fazem uso de alguns conceitos, em princípio inofensivos, mas então, em capítulos mais avançados, estes mesmos conceitos revelam profundos problemas, enigmas complexos e diversos paradoxos conhecidos. Esta interpretação sugere algo bastante otimista: que há solução para os paradoxos complexos e que elas devem seguir os passos das soluções de enigmas simples, pois alguns padrões podem ser reconhecidos entre aquelas simples charadas e os paradoxos mais profundos. Este otimismo, entretanto, é obtido ao custo de um deflacionamento filosófico dos paradoxos. Se a solução de um paradoxo é algo semelhante à solução de um enigma ou de um quebra cabeças, então elas não devem envolver questões filosóficas profundas e significativas, mas apenas artifícios astuciosos, divertidos, e no máximo, de alta complexidade algorítmica.

Quine sugere, em *The Ways of Paradox*, uma classificação dos paradoxos que pode ser bastante instrutiva a este respeito. Há os paradoxos falsídicos, cuja conclusão é sempre falsa, e o erro se encontra em alguma premissa ou alguma inferência. A suposta prova de que  $2 = 1$  é um bom exemplo: suponha  $x = 1$ . Então  $x^2 = x$ . E  $x^2 - 1 = x - 1$ . Dividindo ambos os lados por  $x - 1$ , concluímos que  $x + 1 = 1$ , ou seja, quando  $x = 1$ ,  $2$

---

<sup>5</sup> Dentre eles *The Riddle of Scherazade and other amazing puzzles, ancient and modern* (1997) e *The Lady or the Tiger? And other logic puzzles* (1982).

= 1. Por mais surpreendente que pareça este argumento, a princípio, ninguém disputaria que a conclusão é incorreta e que o passo inferencial incorreto consiste em dividir por  $x - 1$ , que é o mesmo que dividir por 0 quando  $x = 1$ . Paradoxos falsídicos são superficiais e de fácil solução (possuem soluções amplamente aceitas), assemelhando-se às falácias. Há ainda os paradoxos verídicos, que nos levam a conclusões que verdadeiramente se seguem das hipóteses, e deste modo desautorizam as hipóteses, já que a conclusão seria inaceitável. O exemplo é o paradoxo do barbeiro de Russell: suponha uma vila que possui um barbeiro que barbeia todos aqueles que não barbeiam a si mesmos. Este barbeiro barbeia a si mesmo? Se sim, então ele é barbeado pelo barbeiro que só barbeia aqueles que não barbeiam a si mesmos, logo, não barbeia a si mesmo. Mas se não, então ele deve ser barbeado pelo barbeiro que barbeia quem não barbeia a si mesmo, ele mesmo, logo, sim, ele barbeia a si mesmo. Ele barbeia a si mesmo se, e somente se, ele não barbeia a si mesmo. O argumento é uma prova *ad absurdum* de que não deve haver um barbeiro ou uma vila que satisfaçam tal condição. Se supusermos sua existência somos levados a absurdos, ou seja, paradoxos verídicos também não são paradoxos genuínos, mas reduções ao absurdo de hipóteses falsas. Por fim, e mais importantes, temos as antinomias. Antinomias são paradoxos genuínos, onde a conclusão é surpreendente e efetivamente ameaçadora, porém obtida a partir de raciocínios aparentemente impecáveis e premissas aparentemente indisputáveis. Antinomias produzem contradição por modos aceitáveis de raciocínio. É o caso do Paradoxo do Mentiroso, uma sentença que diz de si mesma que é falsa, de modo que se é verdadeira, é falsa, e se falsa; verdadeira. É verdadeira se, e somente se, é falsa. Os argumentos que levam a antinomias parecem corretos, e não há, como nos outros casos, uma solução consensual. Qualquer saída<sup>6</sup> carrega um alto preço. Não há uma saída para estes paradoxos que preserve as mesmas intuições e crenças que foram suficientes para derivá-los.

Um paradoxo verídico traz uma surpresa, mas a surpresa rapidamente se auto-dissipa na medida em que ponderamos a prova. Um paradoxo falsídico traz uma surpresa, mas que é vista como um alarme falso quando nós resolvemos a falácia sublinhada. Uma antinomia, no entanto, traz uma surpresa que pode ser acomodada por nada menos do que um repúdio de parte do nosso patrimônio conceitual. (Quine, 1976, p.9).<sup>7</sup>

<sup>6</sup> Segundo Quine (1976, p. 9) a saída é a revisão do nosso esquema conceitual.

<sup>7</sup> A veridical paradox packs a surprise, but the surprise quickly dissipates itself as we ponder the proof. A falsidical paradox packs a surprise, but it is seen as false alarm when we solve the underlying fallacy. An

Se assumirmos a classificação quiniana<sup>8</sup>, vamos defender que paradoxos não são apenas jogos divertidos, mas que também envolvem problemas sérios e de relevância filosófica. Há mais do que uma diferença de grau de complexidade entre enigmas simples e paradoxos genuínos. A solução de um paradoxo genuíno envolve muito mais do que a mera remoção de enganos ou o abandono de uma hipótese. Um paradoxo genuíno deflagra um impasse frente a soluções equivalentemente desconfortáveis, ameaça princípios e inferências que julgamos aceitáveis, nos convidando à reflexão filosófica, ou nos desafiando a assumir a inconsistência.

Mas os paradoxos genuínos não são relevantes apenas porque são atravessados por reflexões filosóficas. Tal como caracterizamos, os paradoxos genuínos são um modelo paradigmático para como se formulam os próprios problemas filosóficos genuínos. Seguindo a nossa intuição, assumimos princípios e raciocínios que nos levam por eles mesmos a assumir coisas evidentemente contra intuitivas e surpreendentes. Nossa intuição deriva resultados que contrariam outras intuições. A partir daí, somos lançados à difícil tarefa de compreender e criticar nossas intuições ou reformular nossos princípios e o nosso raciocínio para melhor espelharem estas intuições. A possibilidade de contestar uma intuição com considerações conceituais e vice-versa, ou formular e reformular teorias que expliquem tais discrepâncias é característico do trabalho filosófico que um problema genuíno exige.

Além disso, os paradoxos são peças fundamentais no desenvolvimento de teorias. Em alguns casos eles geram a demanda por elas, quando deflagram problemas para os quais devemos erguer alguma solução, e em outros casos eles determinam as suas correções necessárias, quando impedem que elas funcionem, criando obstáculos. Podemos dizer que a elaboração de paradoxos em uma área tem como resultado o aprimoramento e a sofisticação do seu aparato teórico. E como já antecipado pelos medievais, paradoxos são também excelentes testes para a eficiência de teorias. Certamente uma forma bastante convincente de defender uma teoria é mostrando que ela resolve paradoxos pendentes para as teorias anteriormente aceitas.

Obviamente, nem sempre os princípios e as intuições colocadas sob suspeição por um paradoxo são de natureza filosófica, ou mesmo de natureza lógica. Assim, embora a

---

antinomy, however, packs a surprise that can be accommodated by nothing less than a repudiation of part of our conceptual heritage.

<sup>8</sup> Neste trabalho seguimos as grafias “quiniano”, “fregiano”, e “kripkiano”. Estas grafias são estabelecidas pela regra de terminação “iano” do *Dicionário Houaiss da Língua Portuguesa*. Edição virtual de 2001.

definição que seguimos ocorra no domínio da filosofia da lógica, vamos assumir que uma classificação dos paradoxos a partir do domínio a que seus problemas se referem é possível. Podem, então, haver paradoxos físicos, astronômicos ou matemáticos, para os quais considerações filosóficas ou lógicas não trazem qualquer elucidação, paradoxos que envolvem impasses na aplicação de conceitos por razões específicas da ciência em questão. Isto não impede, todavia, que alguns paradoxos pertencentes a diferentes domínios compartilhem semelhanças muito significativas, mesmo do ponto de vista filosófico.

Há um debate em torno deste ponto. Alguns autores seguem a posição inicialmente defendida por Ramsey, entendendo que há pelo menos dois tipos de paradoxos de naturezas distintas, de um lado os lógicos e matemáticos, de outro lado os epistêmicos e semânticos, e que as soluções para estes paradoxos não devem ser guiadas pelo critério de homogeneidade, independentemente de haver ou não semelhança entre suas formulações, mas que elas devem ser motivadas por razões específicas, determinadas pelas teorias próprias destes domínios. Outros autores seguem a posição de Russell em buscar soluções que se direcionem às características estruturais dos paradoxos, e não de seus domínios específicos ou dos conceitos em uso.

Um tipo bastante especial de paradoxos semelhantes que chamou a atenção de Russell e de muitos lógicos entre os séculos XIX e XX é aquele dos Círculos Viciosos. A importância destes paradoxos é o que pretendemos apresentar na seção a seguir.

## 1.2 Círculos Viciosos

Um paradoxo que promoveu grande repercussão na história da lógica é o Paradoxo de Russell. Este paradoxo gerou um problema para o logicismo de Frege, projeto que ambicionava construir uma linguagem (a partir da lógica e da teoria de conjuntos) capaz de reduzir a aritmética e estabelecer bases rigorosas para as suas demonstrações, além de fundamentar seus conceitos. O Paradoxo de Russell estava inicialmente vinculado às expectativas da lógica, ela mesma, enquanto uma nova ciência. O Paradoxo de Russell pode ser basicamente formulado da seguinte maneira:

**Paradoxo de Russell:** há conjuntos que pertencem a si mesmos, como o conjunto de todos os objetos abstratos, que sendo ele mesmo um objeto abstrato, deve pertencer a si mesmo. Mas há conjuntos que não pertencem a si mesmos, como o conjunto de todos os homens, que, sendo um conjunto, não pertence ao conjunto de todos os homens.

Supondo que para toda propriedade, temos um conjunto daqueles objetos que a satisfazem, poderíamos formar, então, o conjunto de todos os conjuntos que não pertencem a si próprios. Chamemos este conjunto  $R = \{x: x \in R \leftrightarrow x \notin x\}$ . O problema que surge é:  $R$  pertence a  $R$ ? Se  $R \in R$ , então  $R$  é um conjunto que satisfaz a condição de não pertencer a si próprio, logo,  $R \notin R$ . Mas se  $R \notin R$ , então  $R$  é um conjunto que satisfaz a condição para pertencer a  $R$ , logo,  $R \in R$ . Assim,  $R \in R$ , se e somente se,  $R \notin R$ .

Mas as repercussões deste paradoxo não se detiveram ao domínio do projeto logicista. A partir do diagnóstico lançado por Russell, o problema dos círculos viciosos ganhou autonomia, abrangendo não apenas a teoria de conjuntos, mas mesmo a semântica e a filosofia. Russell, como uma forma de diagnosticar e justificar filosoficamente a restrição a estes fenômenos, formulou<sup>9</sup> o Princípio do Círculo Vicioso (PCV), segundo o qual:

“o que quer que envolva *todos* de uma coleção não pode ser um desta coleção;” ou, conversamente: “se, suposto que uma coleção tivesse um total, ela tivesse membros definíveis apenas em termos deste total, então a coleção dita não tem total.” (Russell, 1908, p.225).<sup>10</sup>

O PCV justifica uma proibição formal a partir da defesa de uma intuição, a idéia de que conjuntos, por exemplo, não podem pertencer a si mesmos. Uma coleção que tem elementos definíveis apenas em termos da coleção ela mesma não tem total, no sentido de que qualquer enunciado acerca de todos desta coleção é um *nonsense*. Ela não pode ser completada, já que seus elementos supõem que ela seja completa para serem definidos, e ela supõe a definição de seus elementos para ser completa<sup>11</sup>. Um reajuste na teoria de conjuntos deve ser imposto de modo a evitar casos mal comportados sob a relação de pertença. A teoria axiomática dos conjuntos inicialmente

---

<sup>9</sup> O precursor deste princípio como um diagnóstico dos paradoxos da teoria de conjuntos foi Poincaré. Mas Russell o utiliza de uma maneira específica, justificando filosoficamente a sua teoria dos tipos.

<sup>10</sup> “*Whatever involves all of a collection must not be one of the collection;*” or, conversely: “*If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total.*”

<sup>11</sup> Que círculos sejam fenômenos que devemos impedir, e que sejam inconsistentes ou absurdos não é uma unanimidade. Há teorias (Teoria de Hiperconjuntos, por exemplo) que pretendem, ao contrário, explicar muitos fenômenos na linguagem e no mundo a partir deles, e mesmo provar que não derivam formalmente inconsistências. Como bem enfatiza Boolos, alguns consideram intuitivo que haja conjuntos que pertencem a si próprios, outros defendem exatamente o inverso, mas nenhum deles possui um argumento qualquer para defender suas intuições. É uma questão de considerar algo intuitivo ou não, e nisto as opiniões divergem (cf. Boolos, 1971, p.220).

formulada por Zermelo, hoje estendida e denominada **ZF**<sup>12</sup>, segue a idéia básica de que a relação de pertença deve ser bem fundada. Conjuntos podem pertencer a conjuntos que pertencem a conjuntos, etc. Mas todo conjunto  $x$  (com exceção do conjunto vazio) deve ter como elemento ao menos um conjunto  $y$  cujos elementos não pertencem a  $x$ . Ficam proibidos casos como:  $x = \{x\}$ . Deste modo, proíbe que um conjunto pertença a si mesmo, ou que qualquer relação de pertença acabe por resultar em círculos viciosos.

Mas, além de fundar parte das considerações de Russell a respeito do paradoxo por ele mesmo formulado, o PCV caracteriza de maneira precisa a estrutura de uma série de outros paradoxos semelhantes que ocorrem mesmo em outros domínios.

Assim, podemos dizer que paradoxos de círculos viciosos são argumentos que derivam uma contradição a partir de premissas e inferências que têm como característica estrutural o envolvimento de círculos viciosos, seja através de conceitos que estão no escopo deles mesmos, sejam conjuntos que devem pertencer a si mesmos, seja a referência de sentenças que refiram a si mesmas.

Entre exemplos de paradoxos matemáticos que envolvem a mesma estrutura, podemos mencionar:

**Paradoxo Burali Forti:** Toda série bem ordenada tem um número ordinal. Um número ordinal é um número que expressa uma determinada posição em uma determinada sequência. Os números naturais estão ordenados pela relação “menor que ( $<$ )”. A cada número natural há um número cardinal correspondente, que expressa a sua quantidade, e um número ordinal correspondente, que expressa a sua posição na sequência formada pelo conjunto dos números naturais ordenados pela relação  $<$ <sup>13</sup>. Podemos formar uma série que inclui todos os números ordinais, e ela mesma é bem ordenada, já que, dados quaisquer dois números ordinais  $\alpha$  e  $\beta$ , ou  $\alpha < \beta$ , ou  $\beta < \alpha$ , ou  $\alpha = \beta$ . Segue-se que a série de todos os números ordinais tem um ordinal, digamos  $\omega$ . Mas a série de todos os ordinais deve incluir  $\omega$ , e esta série bem ordenada de todos os ordinais incluindo  $\omega$  deve ter um ordinal  $\omega + 1$ , que é maior do que  $\omega$ . Ou seja, a série

---

<sup>12</sup> **ZF** remete aos nomes dos autores que elaboraram a teoria: Zermelo (que iniciou a concepção iterativa de conjuntos com os primeiros axiomas da teoria axiomática de conjuntos) e Fraenkel que introduziu importantes axiomas posteriormente. Também é comum encontrar **ZFC**: teoria axiomática de conjuntos mais o Axioma da Escolha.

<sup>13</sup> Se quisermos expressar estes tipos de números a partir de noções de conjuntos temos: um conjunto é uma coleção de elementos abstraídos de suas particularidades, a cardinalidade de um conjunto é a quantidade de elementos que ele tem (dois conjuntos têm a mesma cardinalidade se os seus elementos podem ser colocados em correspondência biunívoca, se para cada elemento de um conjunto corresponde um, e somente um, elemento do outro conjunto). A ordinalidade de um conjunto é a sequência formada por todos os seus elementos levando em conta a ordem em que eles são dispostos.

de todos os ordinais tem um ordinal maior que todos os ordinais, logo; não é a série de todos os ordinais<sup>14</sup>.

Entre exemplos de paradoxos semânticos que envolvem a mesma estrutura, podemos mencionar:

**Paradoxo de Grelling:** algumas expressões são verdadeiras de si mesmas (autológicas) no sentido de se aplicarem a si mesmas, como “curto” e “*english*”. Outras não são verdadeiras de si mesmas (heterológicas), como “longo” e “*portuguese*”. O problema é: a expressão “heterológica” ela mesma é heterológica ou não? Se ela é heterológica, então não se aplica a si mesma, portanto, não é heterológica. Se ela não é heterológica, então ela se aplica a si mesma, logo, ela é heterológica. “heterológica” é heterológica se, e somente se, não é heterológica.

Em todos estes paradoxos ocorrem círculos viciosos. Seja porque permitimos que um conjunto pertença a si mesmo, ou que uma série de ordinais tenha um ordinal (ao falarmos do conjunto de todos os conjuntos, ou do ordinal de todos os ordinais), seja porque permitimos que um predicado instancie a si mesmo. Todos estes casos caracterizam exemplos de relações viciosas<sup>15</sup>.

No âmbito da semântica, estes paradoxos envolvendo círculos viciosos são também conhecidos como paradoxos de autorreferência, principalmente por culpa de um paradoxo que se tornou muito conhecido em contextos filosóficos desde a antiguidade, o Paradoxo do Mentiroso.

### 1.3 Paradoxo do Mentiroso

A versão mais antiga do Paradoxo do Mentiroso vem da antiguidade e é também conhecida como Paradoxo de Epimênides: suponha que Epimênides é um cretense que diz que todos os cretenses são sempre mentirosos. Se entendermos que mentir é fazer enunciados falsos, então o único modo de tornar o enunciado de Epimênides verdadeiro inclui tomar o enunciado ele mesmo como falso. O enunciado de Epimênides se refere a todos os enunciados de todos os cretenses, incluindo os seus próprios enunciados (referindo-se a si mesmo), e diz de todos eles que eles são falsos. Portanto, se ele é

---

<sup>14</sup> Cf. Russell (1908, p.224).

<sup>15</sup> É importante notar que alguns círculos não derivam paradoxos ou contradições, e são aparentemente inócuos. “esta frase está escrita em português” envolve circularidade, mas é verdadeira e não parece suspeita. Isto parece indicar que a presença de círculos é no máximo condição necessária para a ocorrência de paradoxos desse tipo, mas não, condição suficiente.

verdadeiro, então ele é falso. Por outro lado, se todos os enunciados de todos os outros cretenses e mesmo todos os enunciados de Epimênides, incluindo este enunciado atual, são falsos, então o que ele diz é verdadeiro. Sob estas condições<sup>16</sup>, o enunciado de Epimênides é verdadeiro se, e somente se, é falso.

Podemos também imaginar a seguinte situação que evita mencionar qualquer coisa sobre mentiras<sup>17</sup>: um professor *a* de matemática de uma universidade renomada não aprecia de modo algum o trabalho dos professores de filosofia da mesma instituição, e os considera incompetentes e farsantes. Sabendo que ocorre uma apresentação de comunicações destes professores na sala ao lado, *a* decide escrever em seu quadro negro: “todas as sentenças escritas no quadro negro da sala 9 são falsas”. Acontece que ao escrever tal sentença, *a* acreditava estar na sala 10, mas contrário à sua suposição ele mesmo estava na sala de número 9, e a única sentença escrita no quadro negro da sala 9 era a sua própria sentença. O problema é: a sentença escrita por *a* é verdadeira ou falsa?

Temos uma versão ainda mais simples:

“esta própria sentença é falsa”

Nesta versão (na sentença que denominaremos como **PM**) temos uma sentença que diz de si mesma que é falsa. A referência do que é dito ser falso pela sentença é dada pelo demonstrativo “esta”, que, no contexto, resulta em apontar para a própria sentença ela mesma. O problema novamente resulta quando perguntamos se a sentença é verdadeira ou falsa. Se ela é verdadeira, então o que ela diz deve ser o caso, então, ela é falsa. Se ela é falsa, então o que ela diz não é o caso, e ela não é falsa, logo, ela é verdadeira. A sentença é verdadeira se, e somente se, ela é falsa.

---

<sup>16</sup> O enunciado de Epimênides não constitui necessariamente um paradoxo, pois há uma circunstância em que ele seria simplesmente falso, qual seja, quando há enunciados de cretenses que são verdadeiros e outros que são falsos. Havendo, portanto, ao menos um enunciado cretense que não é falso, o enunciado de Epimênides é falso, e daí não derivamos que ele mesmo deve ser verdadeiro. A presença do quantificador nesta versão determina que a ocorrência de um paradoxo dependa de circunstâncias empíricas.

<sup>17</sup> Mentiras podem envolver sofisticados aspectos intencionais, o que complicaria em muito uma discussão já bastante dura. Em princípio, não podemos nem mesmo dizer que se um enunciado é mentiroso, então ele falso. Se entendermos a mentira como um recurso para enganar o interlocutor, por exemplo, podemos dizer que um sujeito pode mentir dizendo a verdade, situação bastante comum nos jogos de cartas. Nestes contextos um sujeito pode blefar duplamente, prevendo que seu adversário adivinharia o blefe, e assim, induzindo-o a pensar que ele não tem determinada carta, ao dizer que tem tal carta, quando de fato ele tem a carta.

Há outros recursos disponíveis na linguagem natural, e mesmo nas linguagens formais<sup>18</sup>, para construir a autorreferência. Poderíamos atribuir nomes às sentenças. Assim, suponha que o nome da seguinte sentença é Jack.

“Jack é falsa”

Ou poderíamos nos referir às sentenças de maneira descritiva:

“a sentença impressa na linha 6, pág. 16 deste texto é falsa”

O Paradoxo do Mentiroso pode mesmo ser expresso de maneira bastante fiel com várias sentenças que reproduzem o círculo vicioso:

- i) ii é verdadeira.
- ii) iii é verdadeira.
- iii) i é falsa.

Mas o que há de surpreendente e aparentemente absurdo em todas estas versões do paradoxo?

Em um primeiro encontro, é difícil não considerar asserções deste gênero como piadas, dificilmente assuntos de séria investigação intelectual. Mas quando a questão subjacente de alguém envolve a noção de verdade de uma maneira central, por exemplo, ao estudarmos as propriedades semânticas de uma linguagem, as piadas tomam um novo ar de seriedade: elas se tornam paradoxos genuínos. E uma das importantes lições da ciência do século vinte, em campos tão diversos como teoria de conjuntos, física, e semântica, é que paradoxos são importantes. O significado de um paradoxo nunca é o paradoxo ele mesmo, mas aquilo de que ele é um sintoma. Pois um paradoxo demonstra que a nossa compreensão de algum conceito básico ou grupo de conceitos é crucialmente defeituosa, que os conceitos não funcionam em casos limites. E embora os casos limites possam nos parecer estranhos ou improváveis, ou mesmo divertidos, o defeito ele mesmo é uma característica dos conceitos, e não dos casos limites que o tornam evidente. Se os

---

<sup>18</sup> Deixaremos este ponto a ser esclarecido no capítulo 2, onde mencionaremos os teoremas relevantes.

conceitos são importantes, isto não é assunto para rir. (Barwise, J. 1987, p. 4).<sup>19</sup>

O problema é que estas versões derivam contradições a partir de suposições e princípios que julgamos profundamente intuitivos e indisputáveis. Por um lado sabemos que há um tipo de entidades para as quais usualmente predicamos “verdadeiro” ou “falso”. Sejam tais entidades sentenças declarativas<sup>20</sup> bem formadas, ou aquilo expresso por elas (enunciados ou proposições)<sup>21</sup>, seguimos com uma forte convicção o princípio de que para tais entidades, ou elas são verdadeiras ou elas são falsas<sup>22</sup>. Não estamos dispostos a admitir que haja sentenças declarativas (proposições) simultaneamente verdadeiras e falsas, tampouco sentenças declarativas (proposições) nem verdadeiras nem falsas. Normalmente uma sentença gramaticalmente correta, que expressa algo que somos capazes de compreender (possui significado), e que não é uma pergunta ou uma exclamação, é algo que estamos dispostos a tomar como verdadeiro ou falso, ou como expressando uma proposição verdadeira ou falsa.

Por outro lado, um outro princípio bem razoável que seguimos diz que uma sentença é verdadeira, se e somente se, o que ela diz é o caso, e que uma sentença é falsa, se e somente se, o que ela diz não é o caso<sup>23</sup>. Se  $x$  é uma sentença verdadeira, então o que ela diz é o caso. Se  $x$  diz que a neve é branca, então é o caso que a neve é branca. Se  $x$  diz que a neve é branca e se isto é o caso, então  $x$  é verdadeira. Mas  $x$  é

---

<sup>19</sup> *On first encounter, it's hard not consider assertions of this sort as jokes, hardly matters of serious intellectual inquiry. But when one's subject matter involves the notion of truth in a central way, for example when studying the semantic properties of a language, the jokes take on a new air of seriousness: They become genuine paradoxes. And one of the important lessons of twentieth century science, in fields as diverse as set theory, physics, and semantics, is that paradoxes matter. The significance of a paradox is never the paradox itself, but what it is a symptom of. For a paradox demonstrates that our understanding of some basic concept or cluster of concepts is crucially flawed, that the concepts break down in limiting cases. And although the limiting cases may strike us as odd or unlikely, or even amusing, the flaw itself is a feature of the concepts, not the limiting cases that bring it to the fore. If the concepts are important ones, this is no laughing matter.*

<sup>20</sup> Obviamente estamos excluindo sentenças exclamativas ou interrogativas, pois não são verdadeiras nem falsas.

<sup>21</sup> Não é necessário ser preciso a respeito destas distinções aqui, ou manter um uso discriminado destas expressões (sentenças e enunciados ou proposições). Os paradoxos aqui mencionados possuem versões extremamente semelhantes para as teorias em que os portadores de verdade são sentenças ou proposições. Uma teoria que tente compreender a linguagem natural, obviamente deve falar de proposições. Teorias acerca de considerações puramente formais (como Tarski) geralmente vão preferir driblar as polêmicas filosóficas acerca de entidades como proposições, tratando apenas de sentenças. Teremos maior cuidado em precisar e manter estas distinções ao apresentar propostas de solução que as pressupõem.

<sup>22</sup> Este princípio é como o Princípio de Bivalência que apresentaremos mais adiante, porém não especifica o portador de verdade, e não é apresentado como um princípio formal. O que quer que seja que usualmente dizemos ser verdadeiro ou falso é sempre verdadeiro ou falso.

<sup>23</sup> Veremos mais adiante que este princípio será apresentado, em um contexto mais formal, como esquema (T).

falsa se, e somente se, x diz que a neve é branca e não é o caso que a neve é branca. Este princípio, por mais redundante e pouco informativo que pareça, elucida o uso dos predicados de verdade. Negar que utilizamos os predicados de verdade desta maneira seria extremamente contra-intuitivo.

De início, estamos mesmo dispostos a tomar estes princípios como evidentes, e quem quer que os negue, nos pareceria sustentar crenças incorretas e mesmo contraditórias. Acontece que o Paradoxo do Mentiroso e as suas versões aqui apresentadas levam destes princípios evidentes a contradições trágicas, pois supomos que **PM** é do tipo de coisas que julgamos serem sempre verdadeiras ou falsas. **PM** é uma sentença declarativa gramaticalmente correta que possui significado<sup>24</sup>. Mas se **PM** é verdadeira, então o que diz é o caso, ou seja, **PM** é falsa, sendo assim verdadeira e falsa. Mas se **PM** é falsa, o que diz não é o caso, ou seja, **PM** é verdadeira, sendo assim verdadeira e falsa.

Podemos colocar este paradoxo como um conjunto inconsistente de sentenças da seguinte maneira:

- (1) A sentença **PM** diz de si mesma que ela é falsa.
- (2) Se uma sentença é verdadeira, então o que ela diz é o caso.
- (3) Se o que uma sentença diz é o caso, então ela é verdadeira.
- (4) Toda sentença é verdadeira ou falsa (Princípio da Bivalência).<sup>25</sup>

O conjunto {(1), (2), (3), (4)} é inconsistente, e, no entanto {(1), (2), (3)}, {(1), (4)} ou {(2), (3), (4)} são consistentes. Supondo que devemos evitar a inconsistência, temos o seguinte trilema: devemos abandonar (1), abandonar juntamente (2) e (3), ou devemos abandonar (4). Abandonar (1) é recusar o problema antes de ele se manifestar. Talvez esta estratégia faça sentido ao tratarmos de linguagens puramente formais, já que podemos criar recursos que simplesmente proíbem a ocorrência de sentenças do tipo **PM**. Mas para a linguagem natural esta estratégia não é possível, não podemos simplesmente controlar as circunstâncias tal que elas sejam sempre favoráveis à restrição de autorreferências danosas. Abandonar (2) e (3) também seria formalmente satisfatório, mas além de consistir em uma restrição completamente carente de

---

<sup>24</sup> Alguns autores (como Martin, 1967) tentam argumentar que **PM** não possui significado, que é um *nonsense*. Mas é difícil defender esta posição, pois **PM** não contém termos que sejam semanticamente incompatíveis (como ocorre nos erros categoriais), mas também não é algo que não possamos compreender de maneira geral, e além de ter uma forma muito semelhante à de várias sentenças que possuem significado, ela mesma pode conter *tokens* (proferimentos ou inscrições particulares) totalmente inócuos. Alguém pode, por exemplo, enunciar **PM** no contexto em que aponta para uma determinada frase verdadeira escrita em um quadro. Neste caso **PM** é simplesmente falsa.

<sup>25</sup> Cf. Kirkham (2003, p. 376-377).

justificação filosófica e bastante especulativa<sup>26</sup>, produziria um uso de “verdadeiro” na linguagem sem qualquer semelhança com o significado real de “verdadeiro” tal como concebemos. Abandonar (4), por sua vez, é uma estratégia que apresenta pontos polêmicos e disputáveis do ponto de vista formal e também filosófico. Em princípio, se abandonar significa negar, então, esta estratégia formalmente implicaria contradições<sup>27</sup>. Mesmo se nós entendermos “abandonar” em um sentido mais fraco, muito esforço ainda é necessário para manter a consistência, a expressividade e a adequação da lógica às intuições que pretendíamos manter, como veremos mais tarde neste trabalho. E do ponto de vista filosófico, esta estratégia também não é ausente de objeções. O que justificaria que sentenças como **PM** não sejam verdadeiras nem falsas? **PM** é uma sentença gramaticalmente correta e que compreendemos bem, além de não compartilhar qualquer traço substancial com aquelas sentenças que usualmente são ditas nem verdadeiras nem falsas, ou seja, **PM** não é uma pergunta ou uma exclamação, não é um *nonsense*, e também não possui falhas de referência<sup>28</sup>. Se aceitarmos que **PM** seja nem verdadeira nem falsa, portanto, temos dificuldade em reformular o Princípio de Bivalência, determinando que tipo de entidades devemos dizer que são sempre verdadeiras ou falsas.

Em suma, o Paradoxo do Mentiroso é um paradoxo genuíno que vai muito além da complexidade dos enigmas de raciocínio. Pela simplicidade como pode ser apresentado e pelo pouco que ele pressupõe (por pressupor coisas extremamente aceitáveis e mesmo triviais), seu grau de profundidade e de dificuldade se torna ainda mais ameaçador. Coloca sob ameaça e suspeição importantes intuições, e desafia teorias filosóficas que pretendam elucidar os conceitos de verdade e significado. A solução para este paradoxo é um importante e legítimo problema.

---

<sup>26</sup> Já que também é difícil pensar qual outro princípio regimentando o uso de “verdadeiro” seria um substituto efetivo. Mais adiante apresentaremos teorias que reformulam este princípio, mas seguindo bem de perto sua intuição básica, o que não pode ser caracterizado como uma restrição no sentido forte.

<sup>27</sup> Apresentaremos a prova deste fato em momento mais oportuno. Mas a idéia é que no sentido clássico da negação, da conjunção, da disjunção e do predicado “falsa”, “nem verdadeira nem falsa” também significa “verdadeira e falsa”.

<sup>28</sup> Há certo consenso entre os autores de filosofia da linguagem e da lógica em assumirem que sentenças com falhas de referência e sentenças *nonsense* são sentenças nem verdadeiras nem falsas. Um exemplo de uma sentença *nonsense* é “sábados comem álgebra”. Nestes casos não entendemos o que deve ser o caso para a sentença ser verdadeira ou falsa, e há incompatibilidade semântica entre os termos. Um exemplo de sentenças sem referência é “o atual rei da França é careca”. Como não há tal coisa como um rei da França atualmente, não podemos dizer que ele é careca, nem que ele não é careca, já que em ambos os casos pressupomos a sua existência, logo, nem a sentença nem a sua negação são verdadeiras, ou seja, ela não é verdadeira nem falsa.

#### 1.4 O que esperar de uma solução?

Diante do impasse criado pelos paradoxos é de se esperar que alguma solução astuciosa se apresente. Pelo que já vimos a respeito da diferença qualitativa entre paradoxos genuínos e enigmas de raciocínio, sabemos que estas soluções não podem ser da mesma natureza que as soluções puramente algorítmicas. Não é o caso de encontrarmos a peça que faltava no quebra-cabeça ou desvendar o enigma eliminando completamente os problemas. Não há um *clique* que indica quando encaixamos a peça correta. Se houvesse uma solução efetiva para um paradoxo neste sentido, tenderíamos a dizer que não era um paradoxo<sup>29</sup>. A solução de um paradoxo deixa rastros. Não dissolvemos o problema, apenas apresentamos bons argumentos para escolher um caminho que julgamos menos danoso ou mais fecundo. Obviamente, na medida em que escolhemos reconsiderar algumas de nossas intuições, estes reajustes podem alterar nossas concepções e promover avanços conceituais significativos, tal que sejamos capazes de olhar para os antigos paradoxos e não conseguir mais ver os problemas envolvidos. Mas disto não se segue que resolvemos os paradoxos da mesma maneira que resolvemos um enigma ou desmentimos uma falácia.

Uma solução para um paradoxo genuíno deve partir, antes de tudo, de um bom diagnóstico. Devemos avaliar a derivação do paradoxo nas suas premissas essenciais, e determinando exatamente quais princípios ou quais inferências ele ameaça. E é discutindo as evidências destes princípios, as intuições que os guiam, as inferências, as teorias e os seus pressupostos, que estaremos aptos para propor reajustes capazes de driblar os impasses, seja evitando que ocorram, seja ajustando as consequências de maneira justificável.

Neste intuito, devemos esperar que as soluções sejam compostas por considerações filosóficas e formais. A solução deve apontar os princípios ou regras de inferências a serem abandonados. Além disto, ela deve mostrar os resultados que seguem de abandonar tais coisas. Se uma solução não aponta o que deve ser abandonado e não mostra que abandonar tais coisas elimina o problema, ela não é efetiva, mas apenas especulativa. Mas a solução também deve apresentar razões independentes para tais reajustes. Deve haver elucidação e justificativa destes reajustes,

---

<sup>29</sup> Curiosamente, a maioria dos paradoxos para os quais uma solução consensual e genuína é encontrada passam a ser considerados como algo que nunca fora realmente um paradoxo. Os autores que se propõem a resolver um paradoxo tendem a querer mostrar que na verdade tal problema não é um paradoxo.

ou a solução não passa de uma função algorítmica. E estes reajustes devem ser aconselháveis por razões que não se limitem apenas a evitar que paradoxos ocorram. Seria completamente infrutífero se não justificássemos os desvios que introduzimos de modo a evitar paradoxos, mesmo no caso de paradoxos puramente matemáticos, já que estes reajustes devem de alguma maneira interferir no significado que atribuímos às operações e aos conceitos utilizados.

No caso do Paradoxo do Mentiroso, devemos nos atentar ainda para uma peculiar situação em relação às linguagens puramente formais e à linguagem natural. O Paradoxo do Mentiroso não é um fenômeno exclusivo das linguagens formais ou da linguagem natural. O mesmo fenômeno, senão uma versão essencialmente semelhante dele, atravessa toda linguagem com recursos que permitam expressar a autorreferência e para a qual pretendemos ter um predicado de verdade. Há estratégias que permitem construir uma linguagem formal na qual o Paradoxo do Mentiroso não ocorra. Obviamente, estas estratégias não se aplicam à linguagem natural. Também não é possível extrair consequências que produzimos em contextos formais para aplicar à linguagem natural sem qualquer ressalva. A linguagem natural não é estável, não possui apenas sentido literal, não é determinada apenas por relações lógicas e é muitas vezes ambígua. Portanto, não podemos simplesmente supor uma formalização da linguagem natural. Contudo, na medida em que o uso de conceitos nas linguagens formais é determinado de modo a melhor espelhar seu significado real (tal como concebemos na linguagem usual e em qualquer linguagem), devemos esperar que alguma aproximação seja possível. Ao menos, é bastante razoável esperar que as soluções propostas para o Paradoxo do Mentiroso nas linguagens formais e na linguagem natural não sejam completamente independentes e incompatíveis umas com as outras.

Alguns requisitos adicionais<sup>30</sup> podem ser úteis para avaliarmos as soluções. Em primeiro lugar, o requisito da *especificidade*: uma solução deve apontar com precisão um passo inferencial ou uma premissa no argumento como sujeitos a revisão, e não de maneira vaga, a solução deve efetivamente desmontar o argumento que leva ao paradoxo.

O requisito da *proibição das Postulações ad hoc*: uma solução deve justificar a revisão sugerida independentemente do argumento de ela desfazer o paradoxo. “Existe a velha história do homem que vai ao médico e, levantando os braços acima da cabeça,

---

<sup>30</sup> Cf Kirkham (2003, pp. 378-379).

diz ‘doutor, me dói sempre que faço isso’. Ao que o médico responde: ‘então não faça’<sup>31</sup>. Não é este o tipo de solução que procuramos. Normalmente, esperamos poder levantar os braços acima da cabeça. Queremos um diagnóstico bem justificado, e possivelmente um tratamento que nos permita eliminar a dor.

Devemos ainda *evitar estragos excessivos*: ao desabilitar uma premissa ou uma regra de inferência antes aceitáveis, corremos o risco de deslegitimar juntamente argumentos inócuos sobre os quais não deveria haver qualquer dúvida. A solução não pode ser excessivamente ampla, a ponto de mutilar raciocínios que deveríamos preservar. Este é o princípio que Haack chama de “não corte seu nariz fora porque você odeia sua cara”<sup>32</sup>. A proibição da autorreferência é uma solução que ilustra excesso de estragos. Muitas sentenças inócuas, e mesmo algumas necessárias para as linguagens formais são autorreferentes.

Mas a solução também não pode ser pouco ampla. É o requisito da *completude*: todas as versões do mesmo paradoxo devem ser cobertas pela solução. Haack denomina este princípio como “não salte da frigideira para o fogo”. Não adianta resolver uma versão do paradoxo se isto nos leva imediatamente para uma versão tão ameaçadora quanto a primeira. Este requisito levanta uma discussão delicada. Uma solução completa (neste sentido específico de completa) do Paradoxo do Mentiroso é uma solução para todos os paradoxos que envolvem círculos viciosos ou apenas para paradoxos semânticos que envolvem círculos viciosos, ou mesmo apenas aquelas que envolvem predicados de verdade?<sup>33</sup>. Dado que paradoxos envolvendo círculos viciosos têm uma semelhança estrutural, devemos formular um diagnóstico geral, ou é bastante que sejamos capazes de esgotar as versões restritas de cada um destes paradoxos?

Por fim, temos o requisito de *conservar intuições*: não devemos permitir que os reajustes promovidos por uma solução nos distanciem das intuições a respeito dos conceitos em questão. É fácil produzir uma reformulação do conceito de verdade tal que nenhuma versão do Paradoxo do Mentiroso seja possível, mas resta saber se o conceito resultante ainda pode ser minimamente utilizado para traduzir o significado intuitivo de verdade.

---

<sup>31</sup> Idem, p. 378.

<sup>32</sup> Cf. Haack (1998, p.190). Talvez uma forma melhor de expressar este requisito utilizando a analogia de Haack seja: “não corte a sua cabeça porque não gosta do seu nariz”, já que neste caso, sim, você elimina um problema causando um estrago enorme e desproporcional.

<sup>33</sup> Neste trabalho vamos nos restringir às versões do Paradoxo do Mentiroso e às teorias semânticas. De todo modo, vamos tentar sugerir em que medida as soluções propostas satisfazem ou não este requisito no sentido mais amplo.

Obviamente, não podemos esperar que uma solução de um paradoxo genuíno cumpra perfeitamente bem todos os requisitos que sugerimos aqui, ou não poderíamos de fato classificar o problema como um paradoxo genuíno. Alguma insatisfação deve remanescer. Entretanto, podemos procurar nos aproximar o máximo possível de uma solução ideal, ou ao menos estes critérios nos serão úteis para refutar uma a uma as tentativas de solução insatisfatórias.

## Capítulo 2

### A Solução Ortodoxa

#### 2.1 Tarski

O que os autores costumam considerar como a Solução Ortodoxa<sup>34</sup> para o Paradoxo do Mentiroso é uma espécie de extensão da teoria que Tarski formulou em “O Conceito de Verdade nas Linguagens Formalizadas” (1933)<sup>35</sup>. Conforme iremos esclarecer na seção 2.2, Tarski não apresenta em sentido estrito uma solução para o Paradoxo do Mentiroso, mas é com base no diagnóstico fornecido por ele que tal solução se torna possível. Para compreendermos bem o diagnóstico de Tarski, devemos antes ter uma idéia do que ele pretendia com a sua teoria.

##### 2.1.1 A Definição Semântica do Conceito de Verdade

O conceito de verdade está presente no cotidiano e seu significado parece bastante intuitivo. A tarefa de definir o que é a verdade, contudo, tem levantado disputas filosóficas intermináveis. Muitos resultados já foram alcançados em Filosofia e mesmo nas ciências pressupondo a significação deste conceito ou tomando-o como um conceito primitivo (isento de definição). Em lógica, por exemplo, é comum definir argumento válido como aquele que não permite que a conclusão seja falsa quando as premissas são verdadeiras. Um físico pode querer saber se alguma descrição sugerida por uma teoria a respeito do movimento de partículas é verdadeira, e mesmo um historiador pode querer decidir se o relato de algum evento do passado é verdadeiro ou falso, se o evento realmente ocorreu de tal maneira ou não. Mas o que significa dizer que algo é verdadeiro?

---

<sup>34</sup> O termo foi cunhado por Kripke (1975, p.58). Kripke utiliza a expressão “*Orthodox Approach*”. Estamos traduzindo como “Solução Ortodoxa”, embora a tradução mais adequada fosse “Abordagem Ortodoxa”, pois esta diferença não implica em uma má compreensão do texto de Kripke, e ajuda nos nossos propósitos.

<sup>35</sup> Denominaremos esta obra de fundamental importância pela abreviação **CTFL** (em referência ao título em inglês: *The Concept of Truth in Formalized Languages*. Tradução de 1956). O artigo original, em polonês, foi publicado em 1933. A tradução que utilizamos, em português, foi realizada por Cezar A. Mortari em “A Concepção Semântica da Verdade. Textos Clássicos de Tarski” (2006).

Desde Aristóteles, muitas teorias tentaram responder a esta pergunta fornecendo uma definição do conceito de verdade. Assim, Aristóteles afirma na *Metafísica*:

Dizer do que é que não é, ou do que não é que é, é falso, enquanto que dizer do que é que é, ou do que não é que não é, é verdadeiro<sup>36</sup>.

Esta é a concepção clássica da verdade. Uma teoria mais recente é a Teoria da Correspondência. Segundo esta teoria “a verdade de uma sentença consiste em sua concordância (ou correspondência) com a realidade”<sup>37</sup>. Tarski entende que a Teoria da Correspondência é uma versão da concepção clássica. Existem, obviamente, concepções alternativas da verdade, como por exemplo, as teorias que defendem que verdadeiro é o que satisfaz critérios de utilidade, ou teorias que defendem que verdadeiro é o que é coerente, mas estas teorias definem a verdade como significando algo totalmente distinto do que é significado pela concepção clássica. Tarski parece interessado em desenvolver a concepção clássica. Mas é necessário marcar que Tarski não está se posicionando no debate filosófico acerca do que é a verdade, ou rejeitando outras concepções como incorretas. Ele pretende apenas elucidar o que é significado por verdade na concepção clássica, sobretudo em Aristóteles, na medida em que a concepção clássica parece capturar o significado intuitivo do conceito de verdade<sup>38</sup>. Ele considera, entretanto, que a concepção clássica ainda não foi elaborada de maneira clara e precisa, sendo principalmente obscuro o que “correspondência com a realidade” significa. A concepção semântica da verdade vai tornar claro e preciso este significado, fornecendo uma definição do conceito. É este o objetivo central de Tarski.

A definição do conceito de verdade também faz parte de um projeto mais amplo de Tarski, que é o da reabilitação dos conceitos semânticos nas linguagens formais. Os positivistas lógicos do início do século XX já haviam lançado dúvidas acerca dos conceitos semânticos, na expectativa de *limpar* a lógica dos resquícios de metafísica. Para os positivistas somente enunciados traduzíveis em termos puramente lógicos e matemáticos deveriam ser aceitos na linguagem formal. Com a sua teoria, Tarski

---

<sup>36</sup> Citação de Tarski (em “A Concepção Semântica da Verdade e os Fundamentos da Semântica” 1944, na tradução em português de 2006, p.160).

<sup>37</sup> Cf. Tarski (1944). Tradução (2006), p.160.

<sup>38</sup> Tarski chega a mencionar os resultados de uma pesquisa que indicaria que 90% das pessoas concordariam que uma sentença tal como “está nevando” é verdadeira, se, e somente se, está nevando. Ele contrapõe a aceitação às instâncias do seu esquema (**T**) com a versão menos precisa de correspondência (cf. Tarski, 1944. Trad. (2006), p.188).

pretende estabelecer as bases para uma semântica aceitável, científica, dedutiva. E como ele faz isto? Reduzindo os conceitos semânticos a conceitos puramente lógicos e matemáticos, formulando a metateoria da linguagem em questão (uma teoria da morfologia desta linguagem) capaz de expressar as relações semânticas da linguagem em termos das suas relações sintáticas. Esta teoria é expressa na metalinguagem.

O conceito de verdade, entretanto, não é tão claramente um conceito semântico<sup>39</sup>. Mas como Tarski vai definir verdade a partir do conceito semântico de satisfação, é dito que Tarski fornece uma definição semântica do conceito de verdade, e uma definição que autoriza a utilização deste conceito na linguagem científica formal, já que os conceitos semânticos eles mesmos serão reduzidos à morfologia da própria linguagem<sup>40</sup>.

Com estas preocupações, Tarski exige que a sua definição satisfaça dois requisitos fundamentais: Adequação Material e Correção Formal. Por uma definição materialmente adequada ele entende uma definição que apreenda o significado expresso pela concepção clássica da verdade. A forma que Tarski elabora para este requisito pode ser expressa no esquema **(T)**:

**(T)**  $x$  é verdadeira se e somente se  $p$

Onde “ $p$ ” deve ser substituída por qualquer sentença da linguagem, e “ $x$ ” deve ser substituída por um nome desta sentença. Se tomarmos como exemplo a sentença “a neve é branca”, então temos:

a sentença “a neve é branca” é verdadeira se, e somente se, a neve é branca.

A sentença anterior é uma instância materialmente adequada do conceito de verdade. Como consequência da definição de Tarski, toda sentença da linguagem deve resultar

---

<sup>39</sup> Um conceito semântico é um conceito que relaciona uma expressão linguística com objetos não-linguísticos. O conceito de satisfação é um conceito semântico: Tarski satisfaz a expressão “é um filósofo que discutiu o Paradoxo do Mentiroso”. O conceito de denotação também é um conceito semântico: “o autor da *Metafísica*” denota Aristóteles. O conceito de verdade não tem um uso exatamente semelhante aos destes conceitos. O conceito de verdade funciona como uma propriedade de sentenças. Mas se esperamos elucidar o significado que é defendido pela concepção clássica, vemos que algo mais ou menos semelhante ocorre. Uma sentença é verdadeira quando o que ela diz é o caso, verdade relaciona uma sentença com um estado de coisas.

<sup>40</sup> Este ponto é delicado. Tarski coloca que vai fornecer uma definição semântica da verdade. Mas em última instância o que ele faz é definir verdade a partir do conceito semântico de satisfação. E ele define o conceito de satisfação a partir de conceitos sintáticos. Talvez Tarski devesse dizer que ele vai produzir uma definição sintática, e não apenas uma definição semântica.

uma instância verdadeira deste esquema. E cada instância deste esquema é uma definição parcial do conceito de verdade, uma definição para o que significa dizer que tal sentença é verdadeira.

Quanto à correção formal da definição, podemos dizer que Tarski espera que a definição satisfaça tanto as leis usuais da lógica clássica – como os princípios de não-contradição e terceiro-excluído – quanto que ela siga os critérios de correção formal para as definições – como os critérios de eliminabilidade e não-criatividade – <sup>41</sup>.

O Princípio de Não-Contradição (**PNC**): o **PNC** é um princípio da lógica que afirma que a conjunção de uma sentença qualquer  $p$  com sua negação  $\sim p$  é invariavelmente falsa, ou melhor, que a negação desta conjunção é verdadeira sob todas as interpretações:  $\models \sim(p \wedge \sim p)$ . Mas a definição de verdade fornecida por Tarski vai acarretar uma versão semântica<sup>42</sup> do mesmo princípio: para todas as sentenças  $x$ ,  $x \notin \text{Tr}$  ou  $\ast \notin \text{Tr}$ <sup>43</sup>, onde  $\text{Tr}$  é a classe das sentenças verdadeiras. O princípio afirma que não é o caso que todas as sentenças são verdadeiras, e que, em particular, de cada sentença, pelo menos ela ou a sua negação não é verdadeira, ou seja, se uma sentença é verdadeira, sua negação não é verdadeira. O princípio não exclui, contudo, que haja uma sentença tal que nem ela nem a sua negação sejam verdadeiras<sup>44</sup>, mas exclui que uma sentença e sua negação sejam ambas verdadeiras. Como Tarski entende que afirmar a verdade da negação de uma sentença é o mesmo que afirmar a falsidade  $F_s$  de uma sentença, então podemos dizer que este princípio afirma que  $\text{Tr} \cap F_s = \emptyset$ . Quantificando sobre variáveis sentenciais, temos:  $\forall p \sim(\text{Tr}(\langle p \rangle) \wedge \text{Tr}(\langle \sim p \rangle))$

O Princípio do Terceiro-Excluído (**PTE**): o **PTE** é um princípio da lógica que afirma que a disjunção de uma sentença qualquer com a sua negação é uma verdade lógica:  $\models p \vee \sim p$ . A definição de verdade de Tarski também acarreta uma versão

<sup>41</sup> Cf. Abílio Rodrigues (2005, pp.32-34).

<sup>42</sup> É uma versão semântica porque introduz o conceito de verdade. A rigor os dois princípios são semânticos, já que são expressos pelo martelo semântico “ $\models$ ”. Vamos convencionar aqui que há um princípio da lógica e um princípio semântico, embora ambos sejam semânticos (no sentido usual), apenas para distinguir uma tautologia que pode ser obtida no cálculo proposicional e de predicados ( $\models \sim(p \wedge \sim p)$ ) da afirmação consequente da definição de verdade que nenhuma sentença é verdadeira e falsa:  $\models \forall p \sim(\text{Tr}(\langle p \rangle) \wedge \text{Tr}(\langle \sim p \rangle))$ . Segundo Tarski, devemos distinguir as leis semânticas das leis lógicas (cf. Tarski. 1944, trad. 2006, p. 177).

<sup>43</sup> Cf. Tarski, **CTFL** (Trad. 2006, p. 66). Note que “ $\ast$ ” é o nome da negação da sentença  $x$ .

<sup>44</sup> A expressão “ou” que ocorre na formulação do princípio não é traduzida como uma disjunção exclusiva, mas como uma disjunção inclusiva. Ou seja, pode haver uma sentença  $x$  tal que nem  $x$  nem  $\ast$  pertencem a  $\text{Tr}$ .

semântica de **PTE**: para todas as sentenças  $x$ ,  $x \in \text{Tr}$  ou  $x \notin \text{Tr}$ <sup>45</sup>. Para qualquer sentença da linguagem, ou ela ou a sua negação é verdadeira. A classe  $\text{Tr} \cup \text{Fs}$  é exaustiva em relação ao domínio das sentenças da linguagem, todas as sentenças da linguagem ( $S$ ) pertencem a ela.  $\text{Tr} \cup \text{Fs} = S$ . Também podemos expressar este princípio como:  $\forall p(\text{Tr}(\langle p \rangle) \vee \text{Tr}(\langle \sim p \rangle))$ . A versão semântica de **PTE** é também conhecida como Princípio de Bivalência (**PB**): toda sentença é sempre verdadeira ou falsa, ou; para toda sentença, ou ela ou a sua negação é verdadeira<sup>46</sup>. Note que se todas as instâncias do esquema (**T**) devem ser verdadeiras, então **PTE** acarreta **PB**:

1.  $\models p \vee \sim p$ . Para toda sentença  $p$ .
2.  $\models \text{Tr}(\langle p \rangle) \leftrightarrow p$ . Esquema (**T**), onde  $p$  é substituído por qualquer sentença e  $\langle p \rangle$  é um nome desta sentença<sup>47</sup>.
3.  $\models \text{Tr}(\langle p \rangle) \vee \text{Tr}(\langle \sim p \rangle)$ . Substituição em 1.

Tarski não faz distinção entre **PTE** semântico e **PB**. Isto ocorre devido ao fato de Tarski identificar “não-verdadeira” com “falsa”, identificar  $\langle \sim \text{Tr}(\langle p \rangle) \rangle$  com  $\langle \text{Tr}(\langle \sim p \rangle) \rangle$ . Se esta identificação não procede, então uma instância de **PTE** envolvendo o predicado  $\text{Tr}$  deve ser:  $\langle \text{Tr}(\langle p \rangle) \vee \sim \text{Tr}(\langle p \rangle) \rangle$ . Uma instância de **PB**, por sua vez, deve ser:  $\langle \text{Tr}(\langle p \rangle) \vee \text{Tr}(\langle \sim p \rangle) \rangle$ . Mas esta distinção não está presente em Tarski, e só faz algum sentido no

<sup>45</sup> Cf. Tarski, **CTFL** (Trad. 2006, p. 66). Novamente a disjunção é inclusiva. Um ponto que talvez justifique porque nos dois princípios a disjunção é inclusiva, é que estes princípios assemelham-se respectivamente a partes das formulações das seguintes noções fundamentais da metateoria: Consistência e Completude. Uma classe  $X$  é consistente sse para toda sentença  $x$ ,  $x \notin \text{Cn}(X)$  (consequências de  $X$ ) ou  $x \in \text{Cn}(X)$ . Uma classe  $X$  é completa sse para toda sentença  $x$ ,  $x \in \text{Cn}(X)$  ou  $x \notin \text{Cn}(X)$ . A classe  $\text{Tr}$  está em acordo com **PNC** se, de uma sentença e sua negação, ao menos uma sentença não pertence a  $\text{Tr}$ , e está em acordo com **PTE** se, de uma sentença e sua negação, ao menos uma sentença pertence a  $\text{Tr}$ . A classe  $\text{Tr}$ , contudo, é consistente se, de uma sentença e sua negação, pelo menos uma não pertence a  $\text{Cn}(\text{Tr})$ . E a classe  $\text{Tr}$  é completa se, de uma sentença e sua negação, pelo menos uma pertence a  $\text{Cn}(\text{Tr})$ .

<sup>46</sup> Outra vez, estamos tomando o **PTE** como um princípio da lógica e **PB** como um princípio semântico, mas ambos são semânticos, são expressos pelo martelo semântico “ $\models$ ”. Queremos apenas distinguir a tautologia  $\langle p \vee \sim p \rangle$  da afirmação “toda sentença é verdadeira ou falsa”.

<sup>47</sup> Vamos utilizar esta notação para expressar nomes estruturais descritivos de expressões. Estes nomes são descrições dos termos que ocorrem na expressão. Suponha a sentença “a neve é branca”. O nome estrutural descritivo desta sentença é “a sentença cuja primeira palavra é ‘a’, a segunda é ‘neve’, a terceira é ‘é’, e a quarta é ‘branca’”. Tarski utiliza este tipo de nomes na sua metalinguagem para evitar problemas com nomes por funções de citação (já que estas funções seriam intensionais). Na metalinguagem temos nomes para cada expressão da linguagem, e a concatenação entre os nomes dos termos de uma sentença resulta um nome estrutural descritivo da sentença. Suponha que “ng” é o nome de “ $\sim$ ”, “un” é o nome de “ $\vee$ ”, “Vk” é nome da variável “ $x$ ”, “Vl” é nome da variável “ $y$ ”, “gr” é o nome da relação “ $>$ ”, e “ $\frown$ ” é o símbolo de concatenação. Então,  $(((((ng \frown un) \frown Vk) \frown un) \frown Vl) \frown Vk) \frown gr) \frown Vl$  é o nome estrutural descritivo da sentença  $\langle \sim(\forall x \forall y(x > y)) \rangle$ .

contexto em que sentenças podem ser não verdadeiras sem que isto signifique que elas são falsas. Para Tarski, toda sentença é sempre verdadeira ou falsa.

O critério de Eliminabilidade: uma definição obedece ao critério de eliminabilidade se ela define um conceito *e* de tal forma que para toda expressão *E* em que *e* ocorre podemos produzir uma expressão *E'* equivalente a *E* na qual *e* não ocorre. A definição deve relacionar o *definiendum* (expressão que queremos definir) e o *definiens* (expressão por meio da qual definimos outra expressão) por um “se e somente se”, de modo que o *definiens* seja condição suficiente e necessária para o *definiendum*, mas o conceito a ser definido deve ocorrer apenas no *definiendum*, e estar completamente eliminado no *definiens*<sup>48</sup>.

O critério de Não-Criatividade: este critério determina que uma definição formalmente correta não deve permitir a derivação de teoremas que não eram permitidos antes da definição. A definição, portanto, não pode autorizar uma contradição, já que, então, autorizaria qualquer teorema, incluindo aqueles que não eram permitidos antes da definição.

A definição que Tarski procura, retomando, deve ser materialmente adequada, no sentido de apreender o que é expresso pelo esquema **(T)** e acarretar como verdadeiras todas as instâncias deste esquema. A definição também deve ser formalmente correta, no sentido de obedecer às leis lógicas usuais, além de acarretar o **PNC** e o **PB**. E a definição também deve garantir a eliminabilidade do conceito de verdade e não permitir a derivação de teoremas que não eram deriváveis antes da definição.

Se a linguagem para a qual Tarski quer definir verdade é uma linguagem finita, então o problema tem uma solução imediata: se cada instância de **(T)** é uma definição parcial, e se **(T)** obedece aos demais requisitos (não é incompatível com as leis lógicas e está de acordo com os critérios de eliminabilidade e não-criatividade), então se fizermos uma conjunção de todas as instâncias de **(T)**, temos uma definição:

$$\text{Tr}(\langle p_1 \rangle) \leftrightarrow p_1 \wedge \text{Tr}(\langle p_2 \rangle) \leftrightarrow p_2 \wedge \dots \wedge \text{Tr}(\langle p_n \rangle) \leftrightarrow p_n.$$

Obviamente se a linguagem possui um número infinito de sentenças, este método não irá funcionar, já que teremos uma definição expressa por uma sentença infinita, o que não estaria de acordo com o que é pressuposto pela correção formal de uma

---

<sup>48</sup> A exigência de eliminabilidade determina que apenas definições explícitas sejam consideradas como formalmente corretas (cf. Abílio Rodrigues, 2005, p.34). Conforme veremos, Tarski terá de recorrer a definições recursivas, que eliminam o conceito de verdade por um número finito de passos, mas que não obedece no sentido estrito o critério de eliminabilidade. Para tornar sua definição formalmente correta, portanto, Tarski precisa lançar mão de um recurso que transforma definições recursivas em definições explícitas (cf. Tarski, **CTFL**. Trad. 2006, p. 44, nota 24).

definição. Mas então, surge uma questão: se a linguagem possui um número infinito de sentenças, há alguma maneira de generalizar (**T**), de modo a produzir uma definição igualmente satisfatória? Tarski responde a isto negativamente. Se introduzirmos um quantificador e uma variável em (**T**) de forma a produzir uma generalização, não podemos mais significar o que queremos. A sentença “Para todo  $p$ , ‘ $p$ ’ é uma sentença verdadeira se e somente se  $p$ ” não significa o que pretendemos, já que não podemos fazer substituições dentro dos sinais de aspas em “‘ $p$ ’”. A ocorrência da letra “ $p$ ” nesta expressão é parte de um nome, como “Só” em “Sócrates”, e não como uma parte sintática. Não podemos substituir ambas as expressões na sentença por sentenças da linguagem e ao mesmo tempo manter que em um caso nos referimos ao nome da sentença e noutro à sentença ela mesma.

Não podemos fazer esta generalização, então o recurso é procurar uma definição recursiva. Em lógica é comum formularmos definições recursivas para conjuntos infinitos. A linguagem formal que possui um número infinito de sentenças é um conjunto gerado indutivamente, e qualquer predicado de sentenças desta linguagem deve ser definido recursivamente. Geramos um conjunto infinito de sentenças da seguinte maneira: tomamos inicialmente um conjunto de sentenças *base*, que serão as nossas sentenças primitivas, e um conjunto finito de operações sobre sentenças, determinando sob quais condições estas operações resultam em novas sentenças. Reiterando a aplicação destas operações, temos sempre mais sentenças, e assim podemos gerar um conjunto infinito de sentenças. Para um conjunto de sentenças gerado indutivamente, uma definição recursiva de algum predicado sentencial vai determinar sob quais condições este predicado se aplica da seguinte maneira: há os casos de *base*, sentenças atômicas que expressam relações da linguagem específica, para as quais supomos se o predicado deve ou não se aplicar. A partir dos casos de base mais as regras para o funcionamento dos operadores intersentenciais (que ocorrem em sentenças não-atômicas), definimos as condições para a aplicabilidade do predicado às demais sentenças da linguagem, de modo que temos um procedimento para verificar a aplicabilidade do predicado para um número infinito de sentenças.

Mas Tarski também não pode definir verdade diretamente por este método. Na linguagem que possui estrutura quantificacional, sentenças moleculares não são compostas em todos os casos de sentenças atômicas, em alguns casos elas são compostas de fórmulas abertas (que possuem variáveis livres – são compostas de funções sentenciais). E funções sentenciais não são ditas verdadeiras ou falsas, mas

satisfeitas ou não por objetos e sequências. Uma sentença, por exemplo, como  $\langle \forall x (F(x) \vee G(x)) \rangle$  é composta não por sentenças, mas pela função sentencial  $\langle F(x) \vee G(x) \rangle$ , que por sua vez é composta pelas funções  $\langle F(x) \rangle$  e  $\langle G(x) \rangle$ . Sentenças, no entanto, também são funções sentenciais, são funções sentenciais que não possuem variáveis livres. O conceito de satisfação, entretanto, que se aplica tanto a sentenças quanto a funções sentenciais, pode ser utilizado na definição recursiva que Tarski pretende formular. O conceito de satisfação será definido recursivamente, e o conceito de verdade será definido a partir do conceito de satisfação.

Mas a definição não é uma definição geral para qualquer linguagem. Tarski apresenta um método para construir a definição que pode ser estendido a outras linguagens, embora cada linguagem obviamente terá condições próprias de satisfação para as funções sentenciais de *base*. A definição instrutiva que Tarski apresenta é aquela que se aplica ao cálculo de classes. Vamos apresentar aqui uma definição para o cálculo de predicados de primeira ordem.

Definição de “satisfaz”: sejam  $M$  e  $N$  variando sobre sequências de objetos do domínio,  $A$  e  $B$  variando sobre fórmulas (funções sentenciais) da linguagem e os índices subscritos variando sobre números naturais e indicando a posição dos objetos na sequência,  $M_i$  indicando a  $i$ -ésima coisa na sequência  $M$ . Temos:

$\alpha$ ) Para predicados monádicos: para todo  $i$ ,  $M$ :  $M$  satisfaz  $\langle F(x_i) \rangle$  sse  $F(M_i)$ <sup>49</sup>.

Para predicados diádicos: para todo  $i, j$ ,  $M$ :  $M$  satisfaz  $\langle G(x_i, x_j) \rangle$  sse  $G(M_i, M_j)$ . E assim por diante para os outros predicados de diferentes aridades<sup>50</sup>.

$\beta$ ) Para todo  $M$ ,  $A$ :  $M$  satisfaz  $\langle \sim A \rangle$  sse  $M$  não satisfaz  $\langle A \rangle$ .

$\gamma$ ) Para todo  $M$ ,  $A$ ,  $B$ :  $M$  satisfaz  $\langle A \wedge B \rangle$  sse  $M$  satisfaz  $\langle A \rangle$  e  $M$  satisfaz  $\langle B \rangle$ .

$\delta$ ) Para todo  $M$ ,  $A$ ,  $i$ :  $M$  satisfaz  $\langle \forall x_i(A) \rangle$  sse toda sequência  $N$  que difere de  $M$  no máximo na posição  $i$  satisfaz  $\langle A \rangle$ .<sup>51</sup>

<sup>49</sup> Para todo número  $i$  (que indica a posição de um objeto na sequência) e para toda sequência  $M$ , a sequência  $M$  satisfaz a função  $\langle F(x_i) \rangle$  sse a coisa denotada pelo  $i$ -ésimo termo da sequência  $M$  é  $F$ . Se  $F$  denota a classe dos filósofos gregos, e se  $M_i$  denota Aristóteles, então  $M$  satisfaz  $\langle F(x_i) \rangle$ .

<sup>50</sup> Como a definição vai exigir uma cláusula para cada predicado da linguagem, vemos que uma linguagem com um número infinito de predicados (como a linguagem natural) vai demandar uma definição infinita. Mas este ponto apenas reforça a posição defendida por Tarski em vários momentos de que não há uma definição materialmente adequada e formalmente correta para a linguagem natural. Todavia, este problema poderia ser de algum modo evitado. Basta introduzir uma letra para esquema de predicados, com a sinalização de que devemos poder substituir no lugar desta letra qualquer predicado da linguagem. O problema é que neste caso não temos uma sentença e sim uma fórmula aberta, e poderia ser objetado que a definição possui uma parte incompleta e sem sentido (cf. Kirkham, 1992. pp. 226-31).

Tendo obtido a definição de “satisfaz”, podemos definir a expressão “é verdadeira”: uma sentença é verdadeira se, e somente se, ela é satisfeita por todas as sequências de objetos do domínio. Sentenças não possuem variáveis livres. Ou elas possuem constantes individuais e letras de predicados ou relações, ou todas as suas variáveis estão ligadas por quantificadores. No primeiro caso, a sentença é verdadeira, simplesmente, se as constantes instanciam os predicados ou relações: se  $\langle F(x,y) \rangle$  é a relação “ $x$  é maior que  $y$ ”, e  $a$  e  $b$  são nomes<sup>52</sup> de indivíduos do domínio (constantes, portanto), então  $\langle F(a,b) \rangle$  é verdadeira se, e somente se,  $a$  é maior que  $b$ . No segundo caso, temos uma função sentencial sem variáveis livres, onde todas as variáveis estão ligadas por quantificadores, logo, ela é satisfeita por todas as sequências ou por nenhuma. Suponha a sentença  $\langle \forall x \exists y (H(x,y)) \rangle$  de uma linguagem cujo domínio é  $D = \{1,2\}$ , e onde  $\iota(H) = \{\langle 1,2 \rangle, \langle 2,1 \rangle\}$ <sup>53</sup>. Para este domínio existem quatro sequências relevantes:  $M_1 = \langle 1,1 \rangle$ ,  $M_2 = \langle 1,2 \rangle$ ,  $M_3 = \langle 2,1 \rangle$  e  $M_4 = \langle 2,2 \rangle$ . Para avaliarmos se a nossa sentença é verdadeira, devemos verificar se ela é satisfeita por todas as sequências. Pela regra  $\delta$  da definição de “satisfaz”, ela é satisfeita por uma sequência  $M$  se toda sequência  $N$  que difere de  $M$  no máximo em  $x$  satisfaz  $\langle \exists y (H(x,y)) \rangle$ . Mas uma sequência  $M$  satisfaz  $\langle \exists y (H(x,y)) \rangle$ , por sua vez, se existe uma sequência  $N$  que difere de  $M$  no máximo em  $y$  e satisfaz  $\langle H(x,y) \rangle$ , e como vimos, esta última função sentencial é satisfeita por  $M_2$  e  $M_3$ . Assim,  $\langle \exists y (H(x,y)) \rangle$  será satisfeita por todas as sequências e  $\langle \forall x \exists y (H(x,y)) \rangle$  também será satisfeita por todas. Uma sentença é verdadeira sse ela é satisfeita por todas as sequências, e falsa sse ela não é satisfeita por nenhuma sequência. Nesta linguagem, por exemplo,  $\langle \exists y \forall x (H(x,y)) \rangle$  é falsa, já que não é satisfeita por nenhuma sequência.

Enfim, Tarski define verdade pelo conceito semântico de satisfação, e elimina este conceito semântico recursivamente. Este procedimento obedece a todos os requisitos de correção formal, além de preservar o critério de adequação material, determinando como consequências semânticas da teoria todas as instâncias de **(T)**, e determinando

---

<sup>51</sup> Adaptamos a definição de Haack, (1978, p.155). A definição para sentenças que envolvem outros conectivos ou o quantificador existencial pode, obviamente, ser obtida, a partir das equivalências conhecidas.

<sup>52</sup> Tarski não utiliza uma linguagem com constantes individuais e nomes. A razão para isto é que se introduzir, então ele vai ter que utilizar uma função denotação que leve dos nomes aos indivíduos, e então uma definição de denotação deveria complementar a sua definição de satisfação (cf. Kirkham, 1992, p. 229).

<sup>53</sup>  $\iota(\dots)$  é uma função interpretação. Neste caso ela leva de um predicado ou relação para o conjunto das  $n$ -uplas ordenadas que satisfazem o predicado ou relação. Note que estamos utilizando o símbolo “ $\langle \dots \rangle$ ” para as  $n$ -uplas ordenadas, e o símbolo “ $\{ \dots \}$ ” para nomes estruturais descritivos.

ainda como consequências semânticas os princípios: **PB** e **PNC**. A definição de satisfação de Tarski envolve nomes estruturais descritivos, que apenas descrevem a morfologia de expressões da linguagem, envolve termos lógicos, e sentenças (ou traduções de sentenças) da própria linguagem. Daí a razão para Tarski acreditar que reduziu o conceito de verdade à sintaxe da linguagem, e que, assim, reabilitou o conceito para as ciências formais e dedutivas.

### 2.1.2 O Diagnóstico de Tarski

Conforme vimos no capítulo 1, o Paradoxo do Mentiroso é um paradoxo que resulta de princípios fortemente indisputáveis, princípios que se assemelham a versões informais do **PB** e do esquema (**T**) introduzidos neste capítulo como consequências da definição formalmente correta e materialmente adequada de Tarski. O problema que surge, então, é: a definição semântica de Tarski permite a derivação do Mentiroso? Claro que uma resposta positiva a esta questão torna a teoria de Tarski profundamente insatisfatória. Uma definição que permite a derivação do Mentiroso leva a contradições, viola o **PNC**, e, portanto, não é formalmente correta.

Mas se assumirmos **PB** e (**T**), e permitirmos a autorreferência, então sentenças como **PM** parecem estar autorizadas, e estas sentenças derivam contradições:

1. **PM** = “**PM** é falsa” (**PM** é um nome para a sentença “**PM** é falsa”).
2. “**PM** é falsa” é verdadeira sse **PM** é falsa. (esquema (**T**)).
3. **PM** é verdadeira sse **PM** é falsa. (de 1 e 2, por substituição).
4. **PM** é verdadeira ou falsa. (**PB**).
5. **PM** é verdadeira. (hipótese 1).
6. **PM** é falsa. (3, 5, por *Modus Ponens*).
7. **PM** é verdadeira e falsa. (5, 6, por conjunção).
8. **PM** não é verdadeira. (eliminação da hipótese 1).
9. **PM** é falsa. (hipótese 2).
10. **PM** é verdadeira. (3, 9, por *Modus Ponens*).
11. **PM** é verdadeira e falsa. (9, 10, conjunção).
12. **PM** não é falsa. (eliminação da hipótese 2).
13. **PM** não é verdadeira e **PM** não é falsa. (8, 12, conjunção).
14. **PM** é verdadeira ou falsa e **PM** não é verdadeira nem falsa. (4, 13, conjunção).

Como vemos, 14 é uma contradição. Se esperarmos que o **PB** seja instanciado pela sentença do mentiroso, então somos levados a negar o **PB** para a sentença do mentiroso. Sob as duas hipóteses, se **PM** é verdadeira ou se **PM** é falsa, chegamos a: **PM**  $\in$  Tr e  $\sim$ **PM**  $\in$  Tr. Ou seja, se **PB** vale, a classe Tr é inconsistente. Mas então, afirmamos que **PM**  $\notin$  Tr e  $\sim$ **PM**  $\notin$  Tr, negando o **PB**. E daí temos uma nova contradição, já que afirmamos que **PM** é verdadeira ou falsa e **PM** não é verdadeira nem falsa.

Em relação às ocorrências deste paradoxo na linguagem natural, o diagnóstico de Tarski é mais ou menos definitivo:

(I) Fizemos implicitamente a pressuposição de que a linguagem na qual a antinomia é construída contém, além de suas expressões, os nomes para essas expressões, assim como termos semânticos tais como ‘verdadeiro’, referindo-se a sentenças dessa linguagem. Também pressupomos que todas as sentenças que determinam o uso adequado desse termo podem ser afirmadas na linguagem. Uma linguagem com essas propriedades será chamada ‘*semanticamente fechada*’.

(II) Fizemos a pressuposição de que, nessa linguagem, valem as leis ordinárias da lógica.

(III) Fizemos a pressuposição de que podemos formular e afirmar em nossa linguagem uma premissa empírica, tal como o enunciado (2), que apareceu em nosso argumento<sup>54</sup>.

Tarski argumenta que (III) é desnecessário para o Paradoxo do Mentiroso. De fato, na linguagem natural a autorreferência é um fenômeno indisputável, não somos capazes de controlar as circunstâncias empíricas ou o uso de nomes. Sempre podemos nomear as sentenças de uma forma arbitrária, de modo que em algumas circunstâncias uma sentença pode envolver o nome que utilizamos para denotar a sentença ela mesma. Mesmo nas linguagens formais, como veremos a seguir, há um teorema que garante a autorreferência. Mas, independentemente disto, poderíamos formular versões do Paradoxo do Mentiroso que não dependessem da autorreferência empírica. O paradoxo dos predicados heterológicos de Grelling é um exemplo, já que “heterológico” significa exatamente “não verdadeiro de si mesmo”, e não precisamos de uma circunstância que justificasse a dúvida pela autoaplicabilidade ou não da expressão. Sendo um predicado que se aplica a predicados, é natural que pudéssemos responder se ele se aplica ou não a

---

<sup>54</sup> Cf. Tarski 1944 (trad. 2006, p. 168). Note que na premissa (III) Tarski menciona um enunciado (2) anterior do seu texto. Aqui, na nossa derivação, a premissa que permite a autoreferência seria o enunciado 1 da página 33.

si mesmo. O diagnóstico de Tarski para o Mentiroso, portanto, é principalmente fundamentado em (I) e (II). O fechamento semântico mais as leis da lógica ordinária levam ao Paradoxo do Mentiroso. Se a linguagem permite falar de suas próprias expressões, se ela permite dizer quais de suas próprias sentenças são verdadeiras e quais não são verdadeiras, e se todas as instâncias do esquema **(T)** devem ser sentenças verdadeiras da própria linguagem, ela é semanticamente fechada. Ou seja, ela fornece sua própria semântica, é suficientemente expressiva para não depender de uma outra linguagem que forneça a sua semântica. Mas neste caso, se as leis da lógica ordinária (mais especificamente, se os princípios semânticos **PB** e **PNC**) valem, então se ocorrem sentenças do tipo **PM**, estas sentenças mostram que a classe de sentenças verdadeiras desta linguagem é inconsistente. Esta linguagem não permite um uso consistente do conceito de verdade.

Bom, é um fato aparentemente inegável que a linguagem natural é semanticamente fechada. Não há uma linguagem que contenha ou que seja capaz de expressar a linguagem natural<sup>55</sup> e que não possa ser expressa na linguagem natural ela mesma. Em geral, utilizamos a linguagem natural para explicar o significado de fragmentos de linguagens formais, utilizamos fragmentos de linguagens formais mais expressivos para explicar outros fragmentos menos expressivos, e utilizamos a linguagem natural para explicar a própria linguagem natural. A linguagem natural é universal neste sentido, não possui limitações de expressividade, não é parte própria de uma linguagem mais rica.

Mas, que a linguagem natural obedeça às leis da lógica ordinária, já não é algo indisputável<sup>56</sup>. O ponto que Tarski salienta<sup>57</sup> é: a linguagem natural não possui uma estrutura exatamente especificada. Uma linguagem com uma estrutura exatamente especificada é uma linguagem para a qual podemos determinar exatamente e sem ambiguidades quais são as suas expressões com significado, destas quais constituem as suas sentenças, e para as quais então, podemos formular condições de assertividade, podemos determinar quais sentenças devemos afirmar. Linguagens formais permitem uma especificação de estrutura: possuem termos primitivos e regras de definição. A

---

<sup>55</sup> Estou sendo impreciso ao usar a expressão “a linguagem natural”. Isto é uma generalização das linguagens naturais (coloquiais, cotidianas, ordinárias), como o inglês, o francês, o português, etc. Mas esta generalização é inofensiva, já que nos casos em que a estamos utilizando supomos que o que é dito vale para as linguagens coloquiais em geral.

<sup>56</sup> Apesar de que, como vimos no cap. 1, princípios muito semelhantes aos da lógica clássica parecem ocupar algum lugar na semântica da linguagem natural.

<sup>57</sup> Cf. Tarski, 1944 (trad. 2006, p. 170).

partir dos termos primitivos e das regras de definição, podemos definir novos termos. Nestas linguagens, temos uma caracterização precisa de sentenças, e introduzimos axiomas e regras de derivação. Portanto, determinamos também quais sentenças são teoremas, já que os teoremas serão aquelas sentenças derivadas ou de outros teoremas ou dos axiomas que introduzimos. Nas linguagens formais, os axiomas e os teoremas são as sentenças que nós estamos autorizados a afirmar. Mas a linguagem natural está repleta de ambiguidades, de modo que não sabemos sempre quais expressões são sentenças com significado, ou quais sentenças devemos afirmar. A linguagem natural, portanto, não pode ser dita como obedecendo às leis da lógica. E, portanto, também não podemos afirmar que a linguagem natural é inconsistente.

O que pode ser dito, contudo, é que se uma linguagem semelhante à linguagem natural é semanticamente fechada e obedece às leis lógicas usuais, então esta linguagem permite a derivação de paradoxos e contradições, e, portanto, ela é inconsistente. Assim, deve haver certo cuidado na construção das linguagens formais.

As linguagens formais não precisam, contudo, preservar o fechamento semântico. Podemos construir as linguagens formais de modo que os termos semânticos aplicáveis às suas expressões não ocorram nela mesma, e que ela não seja capaz de referir-se às suas próprias expressões. Formulamos a semântica desta linguagem (L), então, na sua metalinguagem (ML), e esta metalinguagem é exatamente uma linguagem que contém<sup>58</sup> a linguagem L, possui os nomes estruturais descritivos das expressões de L e também os predicados semânticos que se aplicam às expressões de L, mais o vocabulário lógico suficiente para expressar estas relações. A metalinguagem, portanto, deve ser essencialmente mais rica do que a linguagem. Quando formulamos a semântica de uma linguagem devemos distinguir a linguagem na qual formulamos a semântica (ML) da linguagem para a qual estamos formulando a semântica (L). O predicado verdade a ser definido em ML não se aplica às sentenças de ML, mas às sentenças de L. O predicado verdade que se aplica às sentenças de ML deve ser definido e expresso em MML, e assim por diante, produzindo uma Hierarquia de Linguagens. O esquema (T), portanto, deve ser entendido como:

$$(T) x \in Tr \text{ sse } p$$

---

<sup>58</sup> Se  $ML \supset L$  (L é um subconjunto próprio de ML), então tudo que pertence a L também pertence a ML. Mas em outros momentos Tarski diz que basta que ML contenha expressões que significam o mesmo que todas as expressões de L, que são traduções de expressões de L. Note que neste último caso a definição dos predicados semânticos depende de já termos definido uma função tradução ou o conceito semântico de sinonímia.

Onde **(T)** é uma sentença da metalinguagem. Então, “*p*” deve ser substituído por sentenças da metalinguagem que são as mesmas sentenças da linguagem L, ou que são traduções destas sentenças. E “*x*” deve ser substituído por nomes estruturais descritivos destas sentenças. **(T)** é uma sentença da metalinguagem ML que expressa as condições de um uso materialmente adequado de Tr para as sentenças da linguagem L. Mas **(T)** ela mesma, assim como outras sentenças da metalinguagem, não são ditas pertencerem ou não à mesma classe Tr especificada em **(T)**. Para as sentenças da metalinguagem ML, uma metametalinguagem MML deve ser formulada, onde possamos expressar a semântica de ML.

Pelo diagnóstico apontado por Tarski, se eliminarmos (I), a antinomia do mentiroso não pode mais ser derivada. Mas nas linguagens formais a distinção L-ML em alguns casos também não é suficiente para evitar o fechamento semântico, pois nem sempre podemos garantir que ML é essencialmente mais rica do que L. Quando L é uma linguagem com ordem infinita, não podemos formular uma ML na qual definiríamos os conceitos semânticos para L. Nestas definições pressuporíamos variáveis de ordem infinitamente altas, e como a ML deve ser formulada abrangendo variáveis de ordem mais alta que a maior ordem das variáveis de L, então a definição dos conceitos semânticos não é possível para estas linguagens<sup>59</sup>.

A impossibilidade de formular uma definição para linguagens com ordem infinita é reforçada pelo Teorema I da seção 5 (**CTFL**) de Tarski. Este teorema espelha o fato, já apontado pelo diagnóstico do mentiroso, que linguagens semanticamente fechadas, capazes de expressar sua própria metalinguagem, derivam contradições e são inconsistentes. O ponto é que, em uma linguagem que é capaz de expressar a aritmética, a metalinguagem é replicada na própria linguagem<sup>60</sup>. E em particular, temos:

---

<sup>59</sup> Em **CTFL**, Tarski considera quatro tipos de linguagens: 1) linguagens em que todas as variáveis pertencem a uma e mesma categoria semântica; 2) linguagens em que o número de categorias nas quais as variáveis são incluídas é maior do que 1 mas finito; 3) linguagens em que as variáveis pertencem a infinitamente muitas categorias mas a ordem dessas variáveis não excede um número natural *n* determinado; 4) linguagens que contêm variáveis de ordem arbitrariamente alta, de ordem infinita (cf. **CTFL**, trad. 2006, p.89). Para as três primeiras, o problema de formular uma definição de Tr para L em ML tem uma solução positiva a partir do método de definição apresentado por Tarski. Linguagens do tipo 4, entretanto, não permitem que se defina Tr para elas em ML do mesmo modo, já que uma tal definição deveria ser expressa abrangendo toda ordem arbitrariamente alta da linguagem, que pode ser infinita.

<sup>60</sup> Em geral, a ML é mais rica do que a L,  $ML \supset L$ . Neste caso, para cada expressão de L temos um nome dessa expressão em ML, e para cada nome em ML; uma expressão correspondente em L, e ML tem ainda as mesmas expressões de L, como vimos. Agora suponha que ao invés de formular a metalinguagem pelo método indicado por Tarski, utilizássemos o seguinte método análogo (os números de Gödel): criamos um sistema de códigos, tal que selecionamos números naturais para denotar expressões da linguagem e construímos nomes de expressões com números naturais. Agora para cada sentença da linguagem (como L é a própria aritmética), que consiste exatamente em uma sentença da aritmética, temos um nome desta

**Lema da Diagonalização de Gödel:** para qualquer fórmula  $\langle H(x) \rangle$  na linguagem da aritmética com  $x$  como sua única variável livre, há uma sentença  $p$  nesta linguagem tal que  $\langle p \leftrightarrow H(\langle p \rangle) \rangle$  é uma tautologia. Uma sentença da linguagem  $L$  é equivalente a uma sentença de  $ML$  que afirma que a sentença de  $L$  é  $H$ .

Este lema, cuja prova iremos omitir<sup>61</sup>, coloca a autorreferência fora de qualquer dúvida. Para qualquer predicado da linguagem, vai haver uma tautologia que estabelece a equivalência entre uma sentença e a sentença da metalinguagem que atribui este predicado à sentença ela mesma, ou seja, a sentença diz de si mesma que ela é  $H$ . Em uma linguagem que seja capaz de expressar a aritmética, qualquer predicado que seja atribuído a expressões vai permitir círculos viciosos, já que vai haver uma sentença da aritmética que diz de si mesma que ela instancia tal predicado.

Como consequência deste lema, para o predicado  $Tr$ , temos: **Teorema I (Teorema da Indefinibilidade, ou Teorema Tarski-Epimênides):** se  $\langle Tr(x) \rangle$  é qualquer fórmula na linguagem da aritmética com  $x$  como única variável livre, então há uma sentença  $p$  nesta linguagem tal que  $\langle \sim(p \leftrightarrow Tr(\langle p \rangle)) \rangle$  é uma tautologia<sup>62</sup>. Se para todo predicado temos uma sentença que diz de si mesma que satisfaz este predicado, então para o predicado “não é verdadeira” temos uma sentença que diz de si mesma que ela não é verdadeira<sup>63</sup>.

Mas havíamos estabelecido que todas as instâncias de **(T)** deveriam ser acarretadas de uma definição materialmente adequada. Logo, temos  $\models p \leftrightarrow Tr(\langle p \rangle)$ , mas temos também  $\models \sim(p \leftrightarrow Tr(\langle p \rangle))$ , ou seja, temos uma sentença, tal que, tanto ela quanto a sua negação pertencem ao conjunto das consequências de  $Tr$ . A classe  $Tr$  é inconsistente.

O que o Teorema I mostra? Ele mostra que não é possível definir ou expressar  $Tr$  em uma linguagem tal que  $Tr$  se aplique a sentenças desta mesma linguagem, ou então, uma versão do Mentiroso pode ser derivada. Mais especificamente, o teorema mostra

sentença em  $ML$  que é um número determinado. Mas como a sentença de  $ML$  é uma sentença que expressa também alguma relação entre números, ela também pode ser expressa na própria aritmética. Neste caso,  $L$  é capaz de expressar a sua própria  $ML$ .

<sup>61</sup> Omitimos esta prova em razão da natureza deste trabalho. Nosso objetivo não é acompanhar de modo preciso os desenvolvimentos formais das teorias. De qualquer maneira, uma apresentação bem detalhada deste lema pode ser encontrada em Heck (*Formal Backgrounds for Theories of Truth*), ou McGee (1991, p.24), ou Hartry Field (2008, pp.26-27).

<sup>62</sup> Esta tautologia segue do Lema de Gödel: aplique o Lema de Gödel à negação de  $Tr$ , obtendo  $\models p \leftrightarrow \sim Tr(\langle p \rangle)$ , o que é equivalente a  $\models \sim(p \leftrightarrow Tr(\langle p \rangle))$ .

<sup>63</sup> Cf. Field (2008, p. 28).

que o conceito de verdade para a aritmética não pode ser definido ou mesmo expresso na própria aritmética. Portanto, devemos introduzir  $Tr$  em uma  $ML$  essencialmente mais rica do que  $L$ , e as sentenças de  $ML$  são ditas verdadeiras ou não por um predicado  $Tr'$  que deve ser introduzido em  $MML$ . Mas quando  $L$  é uma linguagem de ordem infinita (como o cálculo geral de classes), não podemos formular uma  $ML$  que seja essencialmente mais rica e que não possa ser expressa na própria linguagem, então, a definição não é possível<sup>64</sup>. Para estas linguagens será possível apenas adicionar  $Tr$  a  $ML$  como um termo primitivo.

Em suma, o projeto de Tarski era o de fornecer uma definição do conceito de verdade, estendível a outros conceitos semânticos, e que tornasse preciso o que era significado pela concepção clássica. Tal definição não apenas permitiria eliminar as suspeitas a respeito dos conceitos semânticos nas ciências dedutivas, retirando sua carga metafísica, como garantiria um uso consistente do conceito, um uso que não permitisse a derivação de paradoxos. Os critérios exigidos para uma possível definição eram: que ela fosse materialmente adequada, tendo para todas as sentenças da linguagem uma instância correspondente no esquema (**T**), e que ela fosse formalmente correta, não restringindo os princípios da lógica clássica, mais especificamente, acarretando como tautologias os princípios semânticos **PNC** e **PB**, e obedecendo aos critérios de definições formais. Ao investigar o comportamento dos predicados semânticos em linguagens semanticamente fechadas, Tarski conclui que uma definição do predicado verdade para estas linguagens, ou mesmo um uso consistente destes predicados, não é possível. Em linguagens semanticamente fechadas a preservação das leis lógicas ordinárias (principalmente o **PB**) leva a paradoxos. Daí Tarski suplantar a sua definição com uma hierarquia de linguagens justificada na distinção  $L$ - $ML$ . Uma definição de verdade na hierarquia de linguagens funciona do seguinte modo: para uma determinada linguagem, o predicado verdade que se aplica às suas sentenças é definido na sua metalinguagem, e este predicado é aplicável apenas às sentenças desta linguagem. A distinção  $L$ - $ML$ , entretanto, não é o bastante para evitar o fechamento semântico. Linguagens com recursos disponíveis para expressar a aritmética são capazes de expressar a sua própria  $ML$ , de modo que  $ML$  não é essencialmente mais rica do que  $L$ . Este fato reforça a impossibilidade da definição para certas linguagens. Para linguagens

---

<sup>64</sup> A solução de Tarski é adicionar  $Tr$  a  $ML$  sem definição, através de um método axiomático que regimenta o uso de  $Tr$  em  $L$ . Neste caso, todavia,  $Tr$  e os conceitos semânticos que se aplicam a  $L$  não são reduzidos à morfologia de  $L$ , mas apenas adicionados à  $ML$ .

que possuem variáveis de ordem arbitrariamente alta uma definição de verdade não é possível nem mesmo na metalinguagem, já que tal metalinguagem deveria estar equipada de recursos expressivos que abrangessem toda ordem expressa na linguagem, mas como podem ocorrer ordens infinitamente altas na linguagem, este requisito da metalinguagem não pode ser satisfeito. Para este tipo de linguagens a redução dos conceitos semânticos a partir de definições explícitas não é de fato possível.

## 2.2 A Solução Ortodoxa

Como podemos ver, Tarski não propõe, em sentido estrito, uma solução para o Paradoxo do Mentiroso, muito menos uma que seja relevante para o Mentiroso enquanto um fenômeno da linguagem natural, pois a estrutura da linguagem natural, como salienta Tarski, não é exatamente especificada, e nenhuma prova de inconsistência pode lhe ser atribuída senão por aproximação. O que Tarski efetivamente fornece é um método de produzir definições formalmente corretas e materialmente adequadas dos conceitos semânticos para linguagens bem construídas e que obedecem a certas restrições. Para estas linguagens, é possível assegurar um uso consistente do predicado verdade (e dos demais conceitos semânticos<sup>65</sup>) tal que paradoxos não ocorram.

Mas um tipo de solução parece surgir diretamente ou como extensão do diagnóstico de Tarski e da sua Hierarquia de Linguagens. Pelo diagnóstico de Tarski, o Mentiroso é o resultado de assumirmos o fechamento semântico mais as leis lógicas usuais. A Solução Ortodoxa<sup>66</sup> consiste basicamente em fornecer uma interpretação do predicado verdade que não assumo o fechamento semântico. O predicado verdade é sempre relativo a uma linguagem, e as linguagens estão hierarquicamente ordenadas de acordo com a sua capacidade expressiva<sup>67</sup>. Acima de tudo, na solução ortodoxa as leis

---

<sup>65</sup> Os resultados que Tarski obtém para o conceito de satisfação (que define verdade) não são específicos. Na seção 6 de **CTFL** ele afirma que todos os resultados que ele alcançou se aplicam aos demais conceitos semânticos.

<sup>66</sup> A Solução Ortodoxa é formulada a partir de uma adaptação dos resultados de Tarski. Não há um autor que defenda a Solução Ortodoxa de maneira estrita, embora muitos pontos sejam compartilhados por autores como Burge (1979), e também Charles Parsons (1974).

<sup>67</sup> Na teoria de Tarski os conceitos semânticos são sempre relativos a uma linguagem específica por duas razões independentes. Na medida em que os portadores dos predicados semânticos são as sentenças e não as proposições, então a semântica de cada linguagem específica é fornecida separadamente. E além disto, uma mesma linguagem é estratificada em várias linguagens já que a semântica de uma linguagem é expressa na sua metalinguagem, que é uma linguagem distinta. Assim, não há um predicado comum de

lógicas usuais são mantidas, incluindo o **PB**, já que toda sentença é sempre verdadeira ou falsa, e o **PNC**, já que nenhuma sentença é verdadeira e falsa.

Suponha uma linguagem  $L_n$  que contenha as sentenças “a neve é branca” e “a grama é branca”. Esta linguagem possui uma sentença verdadeira e uma sentença falsa: “a neve é branca” é verdadeira-em- $L_n$ . E “a grama é branca” é falsa-em- $L_n$ . Mas os predicados de verdade que se aplicam às sentenças de  $L_n$  pertencem a  $L_{n+1}$ , e sentenças que envolvem estes predicados não são verdadeiras nem falsas em  $L_n$ , mas verdadeiras ou falsas em  $L_{n+1}$ . Como vimos, a sentença que expressa a verdade de uma sentença da linguagem é uma sentença da metalinguagem. Neste caso, “‘a neve é branca’ é verdadeira-em- $L_n$ ” e “‘a grama é branca’ é falsa-em- $L_n$ ” são sentenças verdadeiras em  $L_{n+1}$ <sup>68</sup>. Assim, as sentenças de uma linguagem  $L_n$  são ditas verdadeiras ou falsas em  $L_n$ , e estes predicados pertencem a uma linguagem  $L_{n+1}$ . As sentenças de  $L_{n+1}$  são verdadeiras ou falsas em  $L_{n+1}$ , e estes predicados pertencem a  $L_{n+2}$ . E assim por diante. Não há um predicado de verdade comum a toda linguagem. Mas o **PB** é mantido para toda linguagem. Para qualquer linguagem  $L_n$  há um predicado  $Tr$ , que pertence a sua metalinguagem  $L_{n+1}$ , tal que para toda sentença  $x$  desta linguagem,  $x \in Tr$  ou  $\neg x \in Tr$ , ou seja,  $x$  é verdadeira-em- $L_n$  ou  $x$  é falsa-em- $L_n$ . E não há uma sentença  $x$  desta linguagem tal que  $x$  é verdadeira-em- $L_n$  e  $x$  é falsa-em- $L_n$ , portanto o **PNC** também é mantido.

Impostas estas restrições, é fácil ver que fenômenos como o Paradoxo do Mentiroso não podem mais ocorrer. A sentença “esta própria sentença é falsa” envolve um uso irrestrito do predicado “é falsa” proibido pela hierarquia de linguagens. Por outro lado, sentenças como “esta própria sentença é falsa-em- $L_n$ ” são inofensivas. Sendo uma sentença que envolve predicados semânticos para  $L_n$ , ela deve ser uma sentença de  $L_{n+1}$ . Portanto, ela não é verdadeira nem falsa em  $L_n$ , mas verdadeira ou falsa em  $L_{n+1}$ . Como ela diz de si mesma que é uma sentença falsa de  $L_n$ , e ela não é uma sentença de  $L_n$ , ela é simplesmente falsa-em- $L_{n+1}$ . Mas disto não se segue que ela é verdadeira como antes. Quando estávamos supondo o fechamento semântico e não subscrevíamos o nível de linguagem da sentença, não podíamos afirmar que **PM** é falsa, já que isto era exatamente o que era dito por **PM**, se **PM** era falsa, então **PM** era verdadeira. Agora podemos afirmar que **PM** é falsa-em- $L_{n+1}$ , pois isto não é o que é dito por **PM**. E se

---

verdade, mas coisas como “verdadeira no português” ou “verdadeira no inglês”, mas também coisas como “verdadeira-em-L” ou “verdadeira-em-ML”.

<sup>68</sup> Portanto, vemos que “a grama é branca” é falsa-em- $L_n$ , pois a grama não é branca. Mas a sentença “‘a grama é branca’ é falsa-em- $L_n$ ”, por sua vez, é verdadeira-em- $L_{n+1}$ , pois dizer de uma sentença falsa que ela é falsa é verdadeiro.

formulamos **PM** como “esta própria sentença é falsa-em- $L_{n+1}$ ”, diremos que ela agora é falsa-em- $L_{n+2}$ , já que não é uma sentença de  $L_{n+1}$ . Deste modo, **PM** não constitui uma sentença paradoxal e não deriva uma contradição. Uma linguagem semanticamente aberta, como as linguagens hierarquicamente ordenadas a partir de sua capacidade expressiva, é consistente.

No capítulo 1, formulamos o Mentiroso como um conjunto de sentenças individualmente plausíveis, mas coletivamente inconsistentes. Nesta formulação, (1) expressava a ocorrência da sentença **PM** na linguagem, (2) e (3) expressavam algo parecido com o esquema (**T**), e (4) expressava algo como o **PB**. Podemos dizer que a solução ortodoxa elimina (1). Preservadas as devidas distinções entre L e ML para toda linguagem, a hierarquia de linguagens e a relativização do predicado de verdade à linguagem para a qual ele se aplica, (**T**) e **PB** são claramente e fundamentalmente mantidos na solução ortodoxa. Mas que haja uma sentença na linguagem que diz de si mesma que ela é falsa é algo que a solução definitivamente recusa. Não podemos formular tal sentença. Para que uma sentença diga de si mesma que ela é falsa, ela deve incluir um predicado que pertence à sua metalinguagem, sem que isto a torne uma sentença da metalinguagem, permanecendo uma sentença da linguagem. Teria de ser algo como “esta sentença é falsa-em- $L_{n+1}$ ”, e esta sentença ela mesma ser uma sentença verdadeira ou falsa em  $L_{n+1}$ . Mas isto não é possível, já que a ocorrência do predicado “falsa-em- $L_{n+1}$ ” automaticamente torna a sentença uma sentença de  $L_{n+2}$ .

A distinção L-ML desautoriza a autorreferência. Como predicados semânticos de uma linguagem são expressos na sua ML, e a estes predicados eles mesmos apenas aplicamos predicados semânticos de MML, não há mais autorreferência ou círculo vicioso. Tome uma sentença  $p$  e o predicado semântico  $H$ . Se  $p$  é uma sentença de L,  $\langle H(\langle p \rangle) \rangle$  é uma sentença de ML, e  $\langle H(\langle H(\langle p \rangle) \rangle) \rangle$  é uma sentença de MML. Supondo que  $L \subset ML$ , todo predicado que pertence a L, também pertence a ML, mas como ML deve ser essencialmente mais rica do que L (L é parte própria de ML), há predicados que pertencem a ML e que não pertencem a L, quais sejam, os predicados semânticos que se aplicam às sentenças de L. Deste modo não há uma expressão que afirme de si mesma que ela é H.

Como já vimos no capítulo 1, o conjunto {(2), (3), (4)} é consistente, e portanto, eliminar (1) é uma solução efetiva. Mas como também já argumentamos, eliminar (1) parece ser uma solução que está disponível apenas para linguagens formais. Se a linguagem natural é semanticamente fechada, como normalmente supomos, como

podemos eliminar (1)? Não podemos simplesmente introduzir uma hierarquia de linguagens a partir da distinção L-ML na linguagem natural, porque a linguagem natural é a sua própria metalinguagem.

Uma saída possível é substituir a hierarquia de linguagens por uma hierarquia de predicados semânticos. Podemos, por exemplo, argumentar que na linguagem natural os predicados de verdade são aparentemente ambíguos e possuem hierarquias implícitas de níveis. Quando entendemos sentenças como “Tudo que John falou é verdadeiro”, estamos implicitamente atribuindo índices hierarquicamente ordenados aos predicados de verdade, de modo que se John disse alguma sentença que envolva predicados de verdade, então a estes predicados são atribuídos índices menores que os da sentença a respeito do que foi dito por John. Mas sem uma atribuição de índices que torne explícita a hierarquia de predicados da linguagem, a linguagem torna-se ambígua e sujeita a paradoxos como aqueles que temos visto.

Esta saída, em princípio, bloqueia o Paradoxo do Mentiroso para a linguagem natural tão bem quanto a solução que apresentamos para as linguagens formais. Se tivermos uma sentença como “Esta própria sentença é falsa”, devemos ler tal sentença como “Esta própria sentença é  $F_{Sn}$ ”, e esta sentença que acabamos de mencionar é ela mesma  $Tr_{n+1}$  ou  $F_{Sn+1}$ . Se ela diz de si mesma que ela é  $F_{Sn}$ , então ela é  $F_{Sn+1}$ . Mas o que isto quer dizer? Isto quer dizer que o uso de predicados semânticos deve obedecer a algum tipo de hierarquia de abrangência, normalmente implícita, e que ao proferir uma sentença como **PM**, nós utilizamos um predicado semântico de nível implícito **n**. Portanto, a sentença proferida deve ser avaliada como satisfazendo ou não o mesmo predicado semântico de nível **n+1**. Retomando a formulação do Mentiroso como um conjunto inconsistente de sentenças, esta solução também consiste em eliminar (1). **PM** não é uma sentença que diz de si mesma que ela é falsa. O que **PM** diz é que ela mesma é  $F_{Sn}$ , já que níveis implícitos são necessários para eliminar a ambiguidade de predicados semânticos<sup>69</sup>.

Enfim, a Solução Ortodoxa baseia-se no diagnóstico de Tarski. Ela assume que o fechamento semântico mais as leis da lógica clássica levam à inconsistência. Recusando-se a aceitar qualquer revisão da lógica clássica, promove a abertura

---

<sup>69</sup> Tyler Burge (1979) defende uma solução bastante parecida com esta. Para ele a justificação para uma introdução de índices nos predicados de verdade é que estes predicados se comportam quase como indexicais, não porque seu sentido é determinado por uma regra sensível ao contexto, mas porque a sua extensão varia de acordo com o contexto. A teoria que Burge apresenta, contudo, não é apenas uma adaptação da solução ortodoxa, mas também uma extensão dos modelos de Kripke (que apresentaremos no capítulo 3). Portanto, ela não se classificaria como um exemplo da Solução Ortodoxa.

semântica. Como para a linguagem natural o fechamento semântico é uma característica indisputável, e não eliminável, uma hierarquia nas atribuições de predicados semânticos é introduzida de modo a tornar explícitos os índices que implicitamente são atribuídos a estes predicados ordinariamente. O essencial da solução ortodoxa, portanto, parece ser a recusa em assumir que os predicados semânticos (como os predicados de verdade) sejam predicados unívocos. Não há um único predicado de verdade comum a toda sentença ou toda linguagem, mas vários. No registro das linguagens formais, há um predicado de verdade para cada linguagem. A linguagem natural, que por sua vez não está sujeita a uma estratificação, possui níveis implícitos de atribuição do predicado verdade.

Aquilo que estamos chamando aqui de Solução Ortodoxa é fundamentalmente dependente da teoria de Tarski. A conclusão desta teoria, entretanto, a respeito da linguagem natural é que o problema de consistência para ela – e assim uma possível solução para Paradoxo do Mentiroso ou não – tem um significado mais ou menos vago, já que ela não tem uma estrutura exatamente especificada. Mas uma linguagem semelhante com estrutura especificada seria inconsistente. A conclusão mais próxima da teoria de Tarski, então, seria a de que a linguagem natural é inconsistente. Todavia, a Solução Ortodoxa é apenas uma adaptação da teoria de Tarski, e esta solução sustenta a consistência da linguagem natural tentando introduzir nela uma hierarquia de predicados de verdade.

### 2.3 Vantagens e Desvantagens

A Solução Ortodoxa tem sido considerada o ponto de partida para as discussões a respeito do Paradoxo do Mentiroso. Mesmo os autores que discordam da solução compartilham em parte o quadro teórico no qual ela se desenvolve. Isto se deve em grande parte à precisão do diagnóstico de Tarski. Este diagnóstico é pressuposto por todas as tentativas realmente sérias de solução. E como argumentamos no capítulo 1, uma condição mínima para uma boa solução, tratando-se de paradoxos, é partir de um diagnóstico preciso. É de fundamental importância levar em consideração premissas que sejam essenciais, e determinar com precisão de que exatamente as conclusões paradoxais resultam. O diagnóstico de Tarski satisfaz esta condição. Segundo este diagnóstico, o Paradoxo do Mentiroso é resultado de assumirmos o fechamento semântico mais a prevalência das leis da lógica clássica (mais precisamente, **PB**).

Parece difícil pensar um diagnóstico alternativo que levasse em consideração aspectos igualmente relevantes e abrangentes, ou que fosse tão preciso.

Além disso, a Solução Ortodoxa tem a vantagem de fazer parte de uma teoria genuína. Não é apenas um conjunto de considerações esparsas ou meras sugestões. Apesar de que, como vimos, Tarski ele mesmo não tenha estendido sua teoria de modo a aplicá-la como uma solução do Mentiroso, principalmente enquanto um paradoxo da linguagem natural, esta solução pressupõe em grande parte os resultados obtidos pela sua teoria. É fácil apontar sugestões que resolveriam o paradoxo sem efetivamente formular uma teoria. Na medida em que tentamos construir uma teoria, entretanto, restringindo aqueles pontos que julgamos responsáveis pela derivação dos paradoxos, então podemos constatar se de fato a solução é realizável e se ela efetivamente resolve nossos problemas. O Mentiroso é um problema conhecido em filosofia desde a Antiguidade, e talvez por isso, é também um problema para o qual muito já foi dito, muitas sugestões já foram apresentadas. A abordagem de Tarski, todavia, inaugura uma nova fase na discussão, justamente por apresentar uma teoria.

Podemos dizer também que a Solução Ortodoxa é uma solução efetiva, ou que cumpre o requisito da *especificidade* que colocamos no capítulo 1. De fato, mostramos que eliminar (1) bloqueia o paradoxo. O conjunto {(2), (3), (4)} é consistente, ou seja, esquema (T) mais **PB** não levam por eles mesmos a paradoxos. As novas versões autorizadas das sentenças que antes expressavam **PM** são inofensivas.

A Solução Ortodoxa também é suficientemente abrangente, cumprindo o requisito da *completude*. Todas as versões do Paradoxo do Mentiroso e mesmo aqueles outros paradoxos semânticos envolvendo círculos viciosos são resolvidos por ela (paradoxos envolvendo “satisfação”, “definibilidade”, “denotação”, “tradução”, etc.). Como dissemos, a hierarquia de linguagens relativiza o uso de todos os predicados semânticos, um predicado semântico é sempre expresso na metalinguagem e toma como argumentos apenas expressões da linguagem. A teoria de Tarski, mais especificamente, não resolve, porém, as versões do Paradoxo do Mentiroso para a linguagem natural, o que constitui uma grande desvantagem para sua teoria enquanto uma solução para o paradoxo. Mas supondo-se que pudéssemos estendê-la como temos feito, ampliamos sua abrangência. O problema, como veremos mais adiante, é que muitas outras objeções podem ser independentemente levantadas contra estender esta solução para a linguagem natural. Em última instância, poderemos no máximo afirmar que a Solução Ortodoxa é parcialmente completa.

Por último, temos de dizer que a Solução Ortodoxa tem a vantagem bastante significativa de preservar o esquema (T), o PNC, o PB, e a lógica clássica em geral. A vantagem disso é que a solução é formalmente correta, ela não ameaça derivar contradições ou abandonar princípios importantes para a lógica clássica. Por isso mesmo, ela também é uma solução econômica, já que como veremos mais adiante uma solução que abandone algum destes princípios vai demandar uma grande reforma de modo a evitar resultados indesejáveis. Além disso, na medida em que preserva o esquema (T), o PNC e o PB, a Solução Ortodoxa conserva algumas intuições importantes. Como dissemos no Capítulo 1, versões semelhantes destes princípios parecem guiar nossa intuição a respeito do significado do conceito de verdade e orientar o seu uso ordinário. Enfim, é bom que uma solução não elimine (T) ou PB e PNC. Mas veremos a seguir que isto também é razão apenas para afirmar que a Solução Ortodoxa satisfaz parcialmente o requisito de *conservar intuições*.

Apesar de todas estas vantagens, alguns autores têm levantado ainda algumas possíveis desvantagens da Solução Ortodoxa. Grosso modo, elas resultam do desconforto em abandonar o fechamento semântico, sobretudo na estratégia utilizada para reproduzir este abandono na linguagem natural, introduzindo uma hierarquia de níveis de predicados de verdade. Vale ressaltar mais uma vez, entretanto, que Tarski não tentou formular uma solução para o paradoxo da linguagem natural. Apenas, na medida em que ele estabelece um método para construir definições do predicado de verdade que não derivem versões deste paradoxo em linguagens formalmente construídas, podemos formular uma possível adaptação.

A primeira objeção que quero considerar toca justamente neste último ponto. A Solução Ortodoxa pretende resolver o paradoxo da linguagem natural juntamente com os paradoxos das linguagens formais. A objeção que se tem levantado a este respeito é a artificialidade da solução. A tentativa de estender a solução para a linguagem natural é bastante forçada. Introduzir índices que expressem uma hierarquia é um recurso meramente técnico. Esse tipo de saída parece resolver muito bem o problema *no papel*, mas diz pouco sobre o problema enquanto um fenômeno real da linguagem. Como dissemos no Capítulo 1, eliminar (1) é recusar o problema antes de ele se manifestar, o que não é uma saída disponível para a linguagem natural. O fechamento semântico parece garantir a ocorrência de sentenças como PM, e não podemos simplesmente eliminar este traço característico da linguagem natural, ou escondê-lo através da introdução de um recurso puramente técnico – hierarquia de predicados de verdade –

que faça com que a linguagem natural se comporte como uma linguagem semanticamente aberta. Mas se esta objeção procede, duas consequências importantes devem ser aqui apontadas. Em primeiro lugar, não existe tal coisa como uma Solução Ortodoxa do Mentiroso no sentido forte, já que esta solução não pode ser estendida à linguagem natural, e o que resta são exatamente as mesmas considerações feitas por Tarski a respeito das linguagens formais. Em segundo lugar, por esta razão a solução deve ser considerada no máximo como parcialmente completa.

Aliás, mesmo para as linguagens puramente formais pode-se considerar a Solução Ortodoxa excessivamente artificial. Embora potencialmente justificada na distinção L-ML, a hierarquia de linguagens é uma manobra puramente técnica, já que a mera compreensão de que há uma linguagem na qual falamos e uma linguagem a respeito da qual falamos não implica uma separação radical, uma relativização dos predicados ou mesmo uma hierarquia de níveis. Mesmo os recursos técnicos das linguagens formais devem ser apoiados por boas razões em alguma medida, já que eles devem espelhar o significado que atribuímos aos conceitos envolvidos. E de fato, a hierarquia de linguagens é uma solução carente de justificação filosófica. Faltam boas razões para justificá-la. A razão suficiente para adotá-la tem sido: se não o fizermos, paradoxos serão permitidos. Estamos mais ou menos na mesma situação do paciente que é aconselhado pelo médico a simplesmente não fazer aquilo que lhe causa a dor. De que adianta um diagnóstico preciso se a solução é apenas um paliativo? Por esta razão também tem sido comum dizer que a Solução Ortodoxa não satisfaz o requisito da *proibição das postulações ad hoc*.

Uma outra objeção é a de que a Solução Ortodoxa produz *estragos excessivos*. Tratando de paradoxos genuínos, devemos estar até certo ponto dispostos a conviver com estragos, já que nem sempre há uma solução capaz de nos reconciliar com todas as nossas intuições e ao mesmo tempo eliminar as consequências indesejáveis do paradoxo. Algumas soluções, no entanto, são excessivamente radicais, e levam junto com o problema partes saudáveis das nossas teorias, elas como que receitam que amputemos nossos braços para eliminarmos a dor que sentimos ao levantá-los. A Solução Ortodoxa joga fora junto com as sentenças do tipo **PM** todo tipo de sentenças que envolvam autorreferência. Mas a autorreferência não é sempre um problema. Há casos totalmente inócuos como “esta própria sentença está escrita em português”, que é uma sentença verdadeira que não deriva qualquer paradoxo. A Solução Ortodoxa, porém, reformularia a sentença. Na medida em que ela envolve a expressão “está escrita

em português”, que é uma expressão que é atribuída a sentenças da linguagem, então ela mesma deve ser expressa na metalinguagem. Já a sentença que diz que a sentença que estamos avaliando é verdadeira porque realmente ela está escrita em português é uma sentença da metametalinguagem. Se distinguirmos linguagem e metalinguagem radicalmente, separando as expressões de uma das expressões de outra, não há mais autorreferência.

A Solução Ortodoxa tem também como consequência que não há mais um único predicado de verdade comum na linguagem, mas um predicado para cada linguagem. E uma linguagem é algo bastante restrito aqui, já que não apenas duas linguagens com domínios absolutamente independentes são distintas, mas mesmo a linguagem e a sua própria metalinguagem são duas linguagens diferentes com seus próprios predicados de verdade. Mesmo uma linguagem bastante específica e limitada vai exigir uma estratificação de níveis e um predicado para cada nível. Esta proliferação de predicados não está em harmonia com o significado intuitivo do conceito de verdade, principalmente não está de acordo com o uso do predicado de verdade na linguagem natural. Na linguagem natural há um único predicado de verdade, o mesmo predicado que atribuímos às sentenças proferidas no uso ordinário, ou às sentenças proferidas a respeito da linguagem ela mesma, ou às sentenças que utilizamos construindo provas formais. A objeção é: na medida em que a solução rejeita um predicado comum de verdade ela não *conserva intuições*. Dissemos que a Solução Ortodoxa preserva parcialmente nossas intuições, mantendo o esquema **(T)** e o **PB**, que espelham nossa concepção intuitiva do conceito de verdade. Mas, como os conceitos semânticos estão submetidos a uma hierarquia, as versões resultantes de **(T)** e **PB** que a solução preserva são de fato apenas parcialmente intuitivas.

Uma objeção ainda mais perturbadora foi levantada por Kripke (1975, pp.54-55). Por esta objeção, aos predicados de verdade não podem ser atribuídos nem mesmo níveis implícitos. Pelo menos não da maneira que é imposta pela hierarquia pré-fixada de níveis. Nesta hierarquia, as sentenças não possuem uma liberdade para buscar seu próprio nível. Se uma sentença atribui verdade a um conjunto de outras sentenças, o seu nível já está previamente fixado na metalinguagem que possa abranger a sentença de maior nível do conjunto. Isto implicaria, no caso de uma adaptação da solução para a linguagem natural, que algumas situações poderiam levar a uma confusão de níveis. Suponha que Jones diga:

“Todas as declarações de Nixon sobre *Watergate* são falsas”

Então, o nível da sentença de Jones deve ser maior do que o nível de todas as sentenças ditas por Nixon. Mas podemos incluir entre as declarações de Nixon:

“Todas as declarações de Jones sobre *Watergate* são verdadeiras”

Neste caso, a sentença de Jones deve ser de um nível maior que a sentença de Nixon e a sentença de Nixon deve ser de um nível maior que a sentença de Jones. Jones pode querer dizer que tudo que Nixon diz é falso, mesmo sem conhecer todas as sentenças ditas por Nixon. Mas se entre estas sentenças temos a sentença acima, não há mais uma atribuição possível de níveis. A situação reproduz a autorreferência de um modo que ela não pode mais ser simplesmente eliminada. Se tudo que Jones e Nixon disseram sobre *Watergate* foram exatamente as sentenças acima, então se o que Jones diz é verdadeiro, o que Jones diz não é verdadeiro, já que ele mesmo diz, portanto, que suas próprias declarações não são verdadeiras. Este exemplo evidencia como a Solução Ortodoxa é insatisfatória para a linguagem natural, mostrando um caso em que a introdução de níveis não é suficiente para impedir a derivação de paradoxos. E este exemplo mostra ainda que a hierarquia pré-fixada de níveis não permite que algumas sentenças possam ser paradoxais em algumas circunstâncias e inócuas em outras. O exemplo acima seria inócuo sob certas circunstâncias: se, por exemplo, Nixon tivesse enunciado alguma outra sentença sobre o *Watergate*, e se esta sentença fosse verdadeira, então as duas sentenças acima seriam falsas e não haveria nenhum paradoxo. A introdução de níveis pré-fixados sempre vai caracterizar as sentenças acima como patológicas, independentemente das circunstâncias empíricas.

A introdução de índices de níveis para os predicados da linguagem natural, portanto, não é apenas contra intuitiva, mas também é uma restrição incompatível com os usos destes predicados. E esta solução também é insuficiente, já que podemos formular versões do paradoxo que ela não resolve, como o próprio exemplo de Kripke bem ilustra. Por um lado, a Solução Ortodoxa, enquanto uma solução exclusiva para os paradoxos das linguagens formais, é restrita, não possui considerações relevantes para os paradoxos essencialmente semelhantes da linguagem natural, e, além disso, também é uma solução exclusivamente técnica, não apresentando razões independentes para suas restrições, e apenas impondo-as com a justificativa de evitar paradoxos. Por outro

lado, a Solução Ortodoxa enquanto uma solução ampla que se aplica aos paradoxos da linguagem natural é artificial, produz estragos excessivos, é contra intuitiva, e mesmo insatisfatória.

Os autores que discutem o Paradoxo do Mentiroso costumam apresentar a Solução Ortodoxa como parte da teoria de Tarski, o que, como temos visto, é um erro. Tarski não discute a linguagem natural. Mas como os autores que discutem este problema estão interessados tanto na linguagem natural quanto nas linguagens formais, eles não preservam com muito cuidado esta distinção. Apresentei aqui uma adaptação da teoria de Tarski para a linguagem natural que cumpre razoavelmente bem o papel de Solução Ortodoxa. As objeções levantadas são direcionadas principalmente a esta solução, e não ao trabalho de Tarski que está inserido em outras questões.

## Capítulo 3

### A Solução Alternativa

#### 3.1 Interlúdio sobre *Truth Value Gaps*<sup>70</sup>

A alternativa principal que se tem levantado em relação à Solução Ortodoxa é a restrição da lógica clássica. Se a abertura semântica da linguagem não produz uma solução satisfatória para o Mentiroso, e se partimos do diagnóstico de Tarski que afirma que os paradoxos resultam de assumirmos o fechamento semântico mais a prevalência da lógica clássica, então, a única saída direta para aqueles que julgam o fechamento semântico uma característica indisputável da linguagem é reajustar a lógica clássica para que ela não derive mais paradoxos.

O Paradoxo do Mentiroso pode ser expresso como um conjunto inconsistente de sentenças individualmente aceitáveis, onde a primeira sentença é uma sentença do tipo **PM**, a segunda e a terceira expressam as duas direções do esquema (**T**), e a quarta expressa o **PB**. No capítulo anterior, vimos que a Solução Ortodoxa basicamente rejeita a primeira destas sentenças, rejeita a ocorrência na linguagem de sentenças que envolvem autorreferência. Resta como opção à Solução Alternativa, rejeitar o esquema (**T**) ou o **PB**. Rejeitar o esquema (**T**) ou uma de suas direções não parece uma boa opção. Primeiramente, porque como já dissemos o que é dito por tal esquema é bastante trivial e indisputável, e depois porque na medida em que (**T**) é muito intuitivo, seria difícil encontrar razões que justificassem independentemente a sua rejeição<sup>71</sup>. A opção é restringir o **PB**, que é um princípio essencial na derivação do paradoxo. Se a sentença do mentiroso não deve ser verdadeira ou falsa, então não há mais paradoxo. O paradoxo decorria da seguinte maneira: se o **PB** é aceito, toda sentença deve ser verdadeira ou falsa, então, **PM** deve ser verdadeira ou falsa. Mas por (**T**), **PM** é verdadeira se, e

<sup>70</sup> Falhas de valores de verdade, ou falhas nas atribuições dos valores de verdade.

<sup>71</sup> Como veremos mais tarde, as restrições ao **PB** irão acarretar restrições ao esquema (**T**), mas neste caso haverá razões que explicam tais restrições. De qualquer maneira, este tipo de restrição deve ser levado em consideração quando avaliarmos as vantagens e desvantagens de uma solução.

somente se, **PM** é falsa (contradição)<sup>72</sup>. Como **PM** não é mais necessariamente verdadeira ou falsa, não temos mais um paradoxo.

O ponto desta solução, portanto, é afirmar que uma sentença como a sentença do Mentiroso não é verdadeira nem falsa, negamos o **PB** para tal sentença, ela produz uma falha na atribuição de valores de verdade (Tr ou Fs), ou seja, ela incorre em *truth value gaps*. O **PB** afirma: para toda sentença  $p$ ,  $\text{Tr}(\langle p \rangle)$  ou  $\text{Fs}(\langle p \rangle)$ . Então, **GAP** afirma: não é o caso que para toda sentença  $p$ :  $\text{Tr}(\langle p \rangle)$  ou  $\text{Fs}(\langle p \rangle)$ , ou seja, para pelo menos uma sentença  $p$ :  $\sim\text{Tr}(\langle p \rangle)$  e  $\sim\text{Fs}(\langle p \rangle)$ .

Esta posição de modo algum é sem precedentes, sobretudo em teorias a respeito da linguagem natural. Há mais de um fenômeno que já motivou a defesa de que há sentenças na linguagem natural que não são sempre verdadeiras ou falsas. Mesmo sentenças declarativas gramaticalmente bem formadas em alguns casos, como veremos, podem envolver algum tipo de falha que nos impede de avaliá-las como verdadeiras ou falsas.

As sentenças que envolvem falhas de referência, por exemplo, são candidatas a satisfazerem esta condição. Suponha a sentença “o atual rei da França é careca”. Se ela é verdadeira, então o atual rei da França é careca, mas se ela é falsa, então o atual rei da França não é careca. Em ambos os casos, pressupõe-se que o atual rei da França existe. Se não existe o atual rei da França (como é o caso), então “o atual rei da França é careca” não é verdadeira, mas também não é falsa, já que neste caso “o atual rei da França não é careca” também não é verdadeira. Quando uma sentença possui falha de referência, é comum dizer que algo que é pressuposto não é verdadeiro, e, portanto, nem a sentença nem a sua negação são verdadeiras, a sentença não é verdadeira nem falsa. A falha de referência exemplifica um tipo de falha mais abrangente, que são as falhas pressuposicionais<sup>73</sup>. Em geral, uma pressuposição é uma relação entre sentenças (ou enunciados) tal que se  $p$  pressupõe  $q$ , então quando  $q$  não é verdadeira,  $p$  não pode ser

<sup>72</sup> Por (T),  $(\text{Tr}(\langle \text{PM} \rangle) \leftrightarrow \text{Tr}(\langle \sim \text{PM} \rangle)) \leftrightarrow (\text{PM} \leftrightarrow \sim \text{PM})$ . Mas para qualquer sentença  $p$ ,  $(p \leftrightarrow \sim p) \leftrightarrow ((p \wedge \sim p) \vee (\sim p \wedge \sim \sim p))$ . E por dupla negação, mais a tautologia  $(\langle p \vee p \rangle \leftrightarrow p)$ ,  $(p \leftrightarrow \sim p) \leftrightarrow (p \wedge \sim p)$ . Ou seja, se **PM** é verdadeira ou falsa, então **PM** é verdadeira sse **PM** é falsa, o que é equivalente a uma contradição.

<sup>73</sup> Há uma discussão entre as posições de Russell (1905) e Strawson (1950) a respeito deste problema. Se considerarmos que “o atual rei da França existe” é parte do que é dito por “o atual rei da França é careca”, então esta última não é uma sentença cujo valor é nem verdadeira nem falsa, mas ela é simplesmente falsa. E pela mesma razão, a sua negação é verdadeira. A negação deveria ser entendida como “não é o caso que existe um rei da França atualmente e ele é único e ele é careca”. Esta é a posição de Russell. Strawson retoma este problema, recusando que a pressuposição seja parte do que é dito pela sentença, mas algo que é implicado por ela, em um sentido específico de implicar. Ao enfatizar a relação de pressuposição aqui para elucidar sentenças com falhas referenciais, estamos obviamente seguindo a interpretação de Strawson.

verdadeira e também não pode ser falsa. A falha de referência é um caso de falha pressuposicional. A sentença “o atual rei da França é careca” pressupõe a sentença “o atual rei da França existe”<sup>74</sup>. Mas há outros exemplos de relações pressuposicionais que não envolvem referência: a sentença “Ivan parou de bater em sua mulher” pressupõe “Ivan costumava bater em sua mulher”, e a sentença “foi o John quem quebrou a sua máquina” pressupõe “alguém quebrou a sua máquina”. Em todos estes casos a falha pressuposicional justifica a afirmação de que alguma sentença não é verdadeira nem falsa. Se Ivan nunca bateu em sua mulher, então “Ivan parou de bater em sua mulher” não é verdadeira, mas não porque ele continua a bater nela, e sim porque “Ivan não parou de bater em sua mulher” também não é verdadeira.

Outro fenômeno que parece garantir exemplos de sentenças sem valor de verdade é o das sentenças *nonsense*. Quando há incompatibilidade semântica entre os termos de uma sentença, quando, por exemplo, o predicado da sentença é atribuído a um tipo de coisas para o qual não é significativamente atribuível, ainda que a gramática da sentença esteja preservada, costumamos dizer que tal sentença é um *nonsense* ou um erro categorial. Tome a sentença “Sócrates é um número ímpar”. Tal sentença certamente não é verdadeira, mas também não podemos dizer que ela é falsa, pois isto pareceria implicar que “Sócrates não é um número ímpar” é verdadeira, e que, desta feita, Sócrates é um número par, pois de todo número para o qual dizemos ser ou não ímpar, se não é ímpar, então é um número par. De fato, há uma diferença entre dizer que uma sentença não é verdadeira porque o objeto denotado pelo termo de sujeito não instancia o conjunto do predicado que lhe é atribuído, mas instancia um outro predicado da mesma família, e por outro lado, dizer que uma sentença não é verdadeira porque o termo do sujeito denota um objeto que não instancia o predicado nem qualquer outro predicado da mesma família, de maneira que o predicado não é significativamente atribuível a tal objeto. Certamente a sentença “a neve é azul” não é verdadeira, mas “a paciência é triangular” também não é verdadeira, em um sentido completamente

---

<sup>74</sup> Para uma discussão sobre *truth value gaps* em linguagens com falhas de referência (falhas pressuposicionais) ver van Fraassen (1966, ou 68). A fonte original de uma possível discussão a respeito de relações pressuposicionais é Frege (*Sobre o Sentido e a Referência*. cf. p. 75, trad. Paulo Alcoforado). Van Fraassen (1968) tenta formular uma solução para o Paradoxo do Mentiroso a partir da semântica pressuposicional. Grosso modo, a idéia é que, dadas as condições da relação pressuposicional, se uma sentença pressupõe a si mesma (a expressão que ocupa o lugar de sujeito na sentença refere à sentença ela mesma), ela não pode ser falsa, mas verdadeira ou nem verdadeira nem falsa. Como **PM** é uma sentença que diz de si mesma que ela é falsa, então ela pressupõe a si mesma e não pode ser falsa, logo, ela é nem verdadeira nem falsa. Esta solução, porém, está sujeita às objeções que apontaremos neste capítulo sobre os modelos de *gaps* tradicionais, e também à objeção independente de que não é claro que haja uma relação pressuposicional presente em **PM**, como há nos outros casos que exemplificamos.

diferente do primeiro. Nestes casos também dizemos que tais sentenças não são verdadeiras nem falsas<sup>75</sup>. Um *nonsense* não é uma sentença gramaticalmente incorreta. “Sábados comem álgebra” é nem verdadeira nem falsa por razões que são também independentes daquelas que tornam “golfe joga dormem” nem verdadeira nem falsa. A razão que torna a última expressão nem verdadeira nem falsa é que ela não é uma sentença, é gramaticalmente incorreta. “Sábados comem álgebra”, por sua vez, é gramaticalmente correta, e não possui condições de verdade.

Mas o que significa exatamente restringir a bivalência? Em geral os autores que seguem essa via de solução entendem que restringir o **PB** é negar o **PB**. Esta é a tese que chamamos **GAP**. Como dissemos acima, se o **PB** afirma que toda sentença é sempre verdadeira ou falsa, então a negação do **PB** deve afirmar que não é o caso que toda sentença é sempre verdadeira ou falsa. Acontece que a negação do **PB** na presença do esquema (**T**) resulta na negação do **PTE**, que, por sua vez, resulta em uma contradição:

Suponha que para alguma sentença a negação do **PB** é verdadeira:  $\sim(\text{Tr}(\langle p \rangle) \vee \text{Fs}(\langle p \rangle))$ , que é o mesmo que:  $\sim(\text{Tr}(\langle p \rangle) \vee \text{Tr}(\langle \sim p \rangle))$ . Assim:

1.  $\sim(\text{Tr}(\langle p \rangle) \vee \text{Tr}(\langle \sim p \rangle))$ .
2.  $\text{Tr}(\langle p \rangle) \leftrightarrow p$  e  $\text{Tr}(\langle \sim p \rangle) \leftrightarrow \sim p$ . Esquema (**T**).
3.  $\sim(p \vee \sim p)$ . Substituição em 1 das equivalências (**T**).
4.  $\sim\sim(\sim p \wedge \sim\sim p)$ . *De Morgan*.
5.  $p \wedge \sim p$ . Dupla Negação e Comutação.

---

<sup>75</sup> Para uma discussão sobre *nonsense* e *truth value gaps*, ver Arthur Pap (1960) e Martin, R. L. (1967). Martin (1967) também propõe uma solução para o Mentiroso a partir da caracterização de *nonsense*. Para isso ele precisa supor uma diferença no teste de correção categorial entre sentenças comuns e sentenças autorreferenciais. Nas primeiras, o que deve ser avaliado como sendo ou não do tipo ao qual o predicado se aplica com significado é o sentido da expressão de sujeito na frase: a sentença “o atual rei da França é careca” não é um *nonsense*, pois mesmo que a referência de “o atual rei da França” não seja preenchida, a expressão é significativa, e é exatamente do tipo de expressões às quais o predicado “é careca” é aplicado com significado. As sentenças autorreferenciais não correm o risco de conterem expressões cujas referências venham a falhar, portanto, o teste de compatibilidade é entre o predicado e aquilo que é referido pela expressão de sujeito na frase. No caso de **PM**, a referência da expressão de sujeito é a sentença ela mesma, e o predicado é o predicado “é falsa”, que só é aplicado com significado a sentenças semanticamente corretas, que sejam verdadeiras ou falsas. A conclusão é que para verificar se **PM** é semanticamente correta (não é um *nonsense*) precisamos ver se a sentença ela mesma é semanticamente correta. Como o teste entra em círculo vicioso, Martin diz que pode ser afirmado que **PM** não é semanticamente correta, e daí, não é verdadeira nem falsa. As objeções a que esta solução está sujeita são: em primeiro lugar, as mesmas objeções que veremos a seguir a respeito dos modelos tradicionais de *gaps* valem para esta solução. E em segundo lugar, podemos objetar que não há um argumento positivo que diz que **PM** é um *nonsense*, mas apenas uma conclusão parcial, que **PM** não pode ser dita satisfazer o teste extremamente forçado de correção categorial.

Ou seja, se afirmarmos que a negação do **PB** é verdadeira, então somos levados a afirmar que alguma sentença do tipo  $\langle \sim(p \vee \sim p) \rangle$  é verdadeira, o que na lógica clássica é o mesmo que dizer que  $\langle p \vee \sim p \rangle$  é falsa. Mas na lógica clássica, isto também implica em dizer que  $\langle p \wedge \sim p \rangle$  é verdadeira. Mas então, se nós afirmarmos que há sentenças que não instanciam o **PB**, devemos assumir a contradição? Esta não parece uma consequência aceitável. Inicialmente, estamos supondo que a motivação para admitirmos falhas de valores de verdade na linguagem é que para algumas sentenças supor que elas sejam sempre verdadeiras ou falsas levaria a absurdos<sup>76</sup>. No caso do Mentiroso, por exemplo, levaria mesmo a assumir que **PM** é verdadeira e falsa. Mas não precisamos introduzir *gaps* na linguagem para isso (para assumir contradições). Basta não resolvermos o paradoxo. A contradição não é um bom resultado. Não é de maneira alguma apoiada pelas nossas intuições. E assumir a contradição também não resolve os conflitos nas atribuições de valores para sentenças que têm algum tipo de patologia (que têm alguma falha semântica), apenas trivializa as atribuições de valores de verdade. Podemos dizer aqui que uma revisão à lógica clássica deve preservar o máximo possível. A consistência, por exemplo, deve ser preservada. Os autores que defendem restrições à lógica clássica como um caminho viável para solucionar paradoxos estão ao mesmo tempo atentos ao tipo de reajustes que suas soluções exigem, já que uma revisão da lógica clássica não pode ser excessivamente radical, ou os estragos provocados pela solução não irão compensar os seus benefícios.

Precisamos distinguir *truth value gaps* de *truth value gluts*<sup>77</sup>. Enquanto **GAP** afirma que nem toda sentença é verdadeira ou falsa, negando o **PB**, **GLUT** afirma: há sentenças que são verdadeiras e falsas, para alguma sentença  $p$ ,  $\text{Tr}(\langle p \rangle) \wedge \text{Tr}(\langle \sim p \rangle)$ . Assumindo que uma sentença é falsa quando sua negação é verdadeira, **GLUT** determina  $\text{Tr}$  como um conjunto inconsistente de sentenças (a classe das consequências de  $\text{Tr}$  inclui tanto uma sentença quanto a sua negação). E preservando o esquema (**T**),

---

<sup>76</sup> Se assumirmos que “o atual rei da França é careca” é verdadeira ou falsa, por exemplo, inevitavelmente assumimos que “o atual rei da França existe” é verdadeira (segundo a interpretação pressuposicional que estamos seguindo), o que, nas circunstâncias atuais, não é o caso. Se aceitarmos afirmar o que é implicado pela verdade do **PB**, e mantermos o que afirmamos com base no conhecimento que temos dos fatos atuais, vamos acabar afirmando coisas contraditórias.

<sup>77</sup> Excesso na atribuição de valores de verdade. São avaliações de sentenças que permitem que algumas sentenças não sejam sempre exclusivamente verdadeiras ou falsas. *Gluts* ocorrem quando alguma sentença é verdadeira e falsa, ou quando temos mais do que dois valores de verdade para a avaliação de sentenças. O **PB** não é expresso como uma disjunção exclusiva, como enfatizamos no capítulo 2. Logo, a ocorrência de *truth value gluts* não envolve necessariamente uma negação do **PB**, mas envolve, neste caso (quando temos uma sentença verdadeira e falsa), uma negação do princípio semântico de não-contradição.

também temos que **GLUT** leva a assumir sentenças do tipo  $\langle p \wedge \sim p \rangle$ <sup>78</sup>. Ora, vimos que a tese de que há sentenças nem verdadeiras nem falsas, sem as devidas considerações, leva a contradições. E vimos também que a tese de que há sentenças verdadeiras e falsas leva a contradições, além de caracterizar diretamente um conjunto inconsistente de sentenças. Mas são teses distintas. Enquanto **GAP** é uma proposta de revisão da lógica com implicações mínimas, e para as quais os autores pretendem levantar reparos que preservem o essencial, **GLUT** exige uma mudança radical. Tanto **GAP** quanto **GLUT** caracterizam revisões da lógica clássica. Mas os autores que defendem **GAP** estão em uma posição em relação à lógica clássica que é diferente daquela dos autores que defendem **GLUT**<sup>79</sup>. As consequências que um defensor de **GAP** está disposto a aceitar são bem menos drásticas.

Desta feita, as teorias que defendem a ocorrência de *truth value gaps* encaram o seguinte dilema: ou distinguem o **PB** do **PTE** impedindo que a negação do primeiro implique uma negação do segundo, ou assumem a equivalência entre **PB** e **PTE**, porém, tomam a restrição ao **PB** em um sentido distinto, que não implique afirmar a verdade da negação do **PB**. Pois, de outro modo, vimos que negar o **PTE** leva a afirmar uma contradição. Este dilema revela a existência de pelo menos duas abordagens a respeito de *gaps*. A primeira é a chamada tradição dos *gaps* fregianos, a segunda; os *gaps* de Kripke. Vamos avaliar rapidamente alguns pontos da tradição fregiana, e deixaremos Kripke para a seção 3.2 deste capítulo. A referência ao nome de Frege na primeira tradição se deve a basicamente duas razões: a primeira é que é normalmente considerado que Frege foi o precursor da discussão sobre *truth value gaps*<sup>80</sup>. A segunda

<sup>78</sup> Substituindo as equivalências (**T**) na fórmula  $\langle \text{Tr}(\langle p \rangle) \wedge \text{Tr}(\langle \sim p \rangle) \rangle$ , temos a fórmula  $\langle p \wedge \sim p \rangle$ .

<sup>79</sup> Field, H. (2008. pp. 79-82) propõe uma classificação das lógicas que é útil para nós aqui: ele chama **BDM** (lógicas *De Morgan* Básicas) às lógicas que compartilham as regras de introdução e eliminação dos conectivos de conjunção e disjunção, do quantificador universal, e introdução e eliminação dos mesmos conectivos com negação, e introdução e eliminação da dupla negação. A Lógica Clássica é **BDM** mais o **PTE** e o princípio de explosão **EXP**. O princípio **EXP** afirma que de uma contradição obtemos qualquer sentença, de uma contradição, o sistema explode e se torna trivial. A lógica clássica assume tanto **PTE** quanto **EXP**. Entre **BDM** e a Lógica Clássica, contudo, há duas lógicas alternativas distintas. De um lado, temos aquelas que negam **EXP**, mas preservam o **PTE**: são lógicas (como a Lógica Paraconsistente e a Lógica Dialeteísta) que aceitam contradições, e tentam tornar as consequências disto menos absurdas. De outro lado, temos as lógicas que preservam **EXP**, mas rejeitam o **PTE**: são lógicas (como a lógica de Kleene, que Field denomina como Lógica Paracompleta, em referência à noção de completude, por analogia ao nome da Lógica Paraconsistente) que não permitem a afirmação de contradições, mas que vêm razões para não afirmarem o **PTE**. Todavia, nem todas as teorias que defendem a restrição do **PB** restringem o **PTE**, como veremos adiante.

<sup>80</sup> A aceitação de *truth value gaps* é originalmente motivada pelo princípio de composicionalidade da referência de Frege e pela tese (também de Frege) de que a referência de uma sentença é um valor de verdade. Segundo o princípio de composicionalidade, o valor semântico de uma expressão é uma função do valor semântico de suas partes. O valor semântico de uma sentença (sua referência) é uma função do

razão é que a idéia de que há apenas um ato ilocucionário relativo a asserções (o asserir) vem de Frege<sup>81</sup>. Nesta tradição, a negação é sempre parte do conteúdo locucionário, e nunca algo como um tipo de ato ilocucionário alternativo a asserir. Portanto, negar uma sentença é sempre asserir a negação. A tradição fregiana de *gaps* entende que restringir o **PB** é negar o **PB** no sentido fregiano de negar, qual seja; afirmar a verdade da negação<sup>82</sup>.

Já fizemos uma distinção no capítulo 2 entre **PB** e **PTE**. O primeiro é um princípio que diz que de toda sentença, ela ou sua negação pertence à classe das sentenças verdadeiras, o segundo afirma que a disjunção de uma sentença com a sua negação é uma verdade lógica. Portanto, o primeiro envolve a introdução de um predicado semântico nas sentenças, enquanto o segundo não envolve. Desde então, temos reservado o nome “**PB**” para o princípio semântico e “**PTE**” para o princípio da lógica. Mas agora, vamos propor uma distinção entre **PB** e **PTE** semântico. Vamos propor esta distinção com base nas evidências levantadas pela ocorrência de *truth value gaps*, que nos inclinam a não mais tomar como equivalentes as expressão “é falso” e “não é verdadeiro”. As sentenças que envolvem falhas pressupicionais ou erros categoriais foram avaliadas como nem verdadeiras nem falsas. Em princípio, podemos dizer destas sentenças que elas não são verdadeiras, já que “não é verdadeira”, no uso comum do predicado, de maneira ambígua significa tanto “é falsa” quanto “é nem verdadeira nem falsa”. Na concepção clássica e na semântica de Tarski, “não é verdadeira” é equivalente a “é falsa”<sup>83</sup>. Mas a partir de agora vamos distinguir.

A partir de agora convencionaremos que uma sentença é falsa sse a sua negação é verdadeira (como antes), mas vamos também convencionar que afirmar a verdade da negação de uma sentença ( $\text{Tr}(\sim p)$ ) é diferente de negar que uma sentença é verdadeira ( $\sim \text{Tr}(p)$ ). O **PB** afirma que toda sentença é verdadeira ou é falsa, e deve ser expresso

valor de suas partes, ou seja, depende da referência das partes. Portanto, uma sentença onde ocorre um termo sem referência é uma sentença cuja referência também falha. Neste caso, nas linguagens onde ocorrem termos sem referência, há sentenças que não são verdadeiras nem falsas.

<sup>81</sup> Ou melhor, remonta à tradição fregiana tratar a negação sempre como parte do conteúdo locucionário. Para uma discussão mais detalhada disto, ver Terence Parsons (1984).

<sup>82</sup> O autor que utilizaremos aqui para apresentar a tradição dos *gaps* fregianos é van Fraassen (1968). Mais adiante, contudo, veremos que van Fraassen não segue totalmente a semântica fregiana, já que ele acaba tomando certas medidas não-composicionais. De qualquer maneira, é claro no texto de van Fraassen a motivação de *gaps* a partir dos problemas levantados por Frege (*Sobre o Sentido e a Referência*), e van Fraassen desenvolve formalmente vários aspectos, além de tornar explícita e argumentada a distinção entre as expressões “é falsa” e “não é verdadeira”, que é uma distinção muito importante para a tradição fregiana de *gaps*. O autor que propõe uma divisão entre a tradição fregiana e a tradição alternativa a partir dos modelos de Kripke para a interpretação de *truth value gaps* é Terence Parsons (1984).

<sup>83</sup> Cf. Tarski (1969. trad. 2006, pp. 204-205).

da mesma maneira: para toda sentença  $p$ ,  $\text{Tr}(\langle p \rangle) \vee \text{Tr}(\langle \sim p \rangle)$ . Mas o **PTE** semântico (ou melhor, o **PB** reforçado: **PBR**) não é mais expresso do mesmo modo. Como o **PTE** da lógica expressa a disjunção de uma sentença com a sua negação, a forma mais correta de formular uma instância do **PTE** envolvendo o predicado “é verdadeira” seria: para toda sentença, ela é verdadeira ou ela não é verdadeira. Ou seja, **PBR**: para toda sentença  $p$ ,  $\text{Tr}(\langle p \rangle) \vee \sim \text{Tr}(\langle p \rangle)$ .

Assim, sentenças patológicas que envolvem algum tipo de falha semântica (*nonsense*, falha de referência e outras falhas pressuposicionais) são sentenças sobre as quais estamos em condições de afirmar que não são verdadeiras, mesmo que isto não implique afirmar que elas sejam falsas. Sentenças como “o atual rei da França é careca” ou “Sócrates é um número ímpar” são não-verdadeiras em um sentido distinto daquele onde as sentenças são falsas. Elas são não-verdadeiras, mas não são falsas. Ocorre algum tipo de erro que nos impede de afirmar a verdade tanto quanto a falsidade destas sentenças. Não podemos afirmar a verdade destas sentenças nem de suas respectivas negações, mas podemos afirmar a não-verdade delas e das suas negações. A partir desta distinção, avaliar estas sentenças como nem verdadeiras nem falsas não irá nos submeter à acusação de negar **PTE** semântico e afirmar uma contradição. Ao admitir que há sentenças nem verdadeiras nem falsas, estamos negando o **PB**, mas não estamos negando o **PTE** semântico (**PBR**), pois sentenças nem verdadeiras nem falsas ainda são não-verdadeiras, e o **PBR** afirma que toda sentença é verdadeira ou não-verdadeira. Se tornarmos explícita a ambiguidade presente na expressão “não é verdadeira”, então o **PBR** diz: toda sentença é verdadeira, ou falsa, ou nem verdadeira nem falsa<sup>84</sup>. As sentenças patológicas tornam o **PB** falso, mas preservam o **PTE** semântico.

Aliás, embora tenhamos argumentado a importância de preservar o esquema (**T**), uma consequência direta de negar o **PB** é que não podemos mais garantir que todas as instâncias deste esquema sejam sempre verdadeiras. Se uma sentença  $p$  não é verdadeira nem falsa, então, qual valor atribuímos a  $\langle \text{Tr}(\langle p \rangle) \rangle$ ? Seguindo a distinção que fizemos entre “falso” e “não-verdadeiro”, podemos dizer que afirmar de uma sentença nem verdadeira nem falsa que ela é verdadeira é falso, e que, portanto,  $\langle \text{Tr}(\langle p \rangle) \rangle$  é falsa<sup>85</sup>. Assim, a instância de (**T**) para  $p$  ( $\text{Tr}(\langle p \rangle) \leftrightarrow p$ ), quando  $p$  é nem verdadeira nem falsa, tem de um lado do bicondicional uma sentença nem verdadeira nem falsa, e de outro

<sup>84</sup> Cf. van Fraassen (1968. p. 147).

<sup>85</sup> Mais tarde veremos que esta não é a única maneira de avaliar sentenças que atribuem predicados de verdade a sentenças patológicas.

lado uma sentença falsa. Na melhor das hipóteses<sup>86</sup>, vamos avaliar este bicondicional como uma sentença que não é verdadeira. Isto é de fundamental importância, pois nós havíamos estabelecido que a negação do **PB** na presença do esquema (**T**) resultava em uma negação do **PTE**. Mas agora vemos que a negação do **PB** torna algumas instâncias de (**T**) não-verdadeiras, então não se segue da negação do **PB** que  $\langle \sim(p \vee \sim p) \rangle$  é verdadeira para algum  $p$ .

Um problema que permanece é: ainda que negar o **PB** não signifique negar o **PBR**, e ainda que negar **PB** não acarrete afirmar que  $\langle \sim(p \vee \sim p) \rangle$  é verdadeira pra algum  $p$ , o que garante que o **PTE** da lógica seja sempre verdadeiro? Vimos que podemos encontrar boas razões para avaliar algumas sentenças da linguagem como nem verdadeiras nem falsas. E vimos que se não identificarmos os predicados “é falsa” e “não é verdadeira”, podemos negar o **PB** sem negar o **PTE** (o semântico e o lógico), garantindo que não nos submetemos a afirmações contraditórias. Mas o que garante que para toda sentença  $p$  a sentença  $\langle p \vee \sim p \rangle$  é sempre verdadeira? Em princípio não podemos garantir tal coisa. Se  $\langle p \rangle$  é uma sentença patológica (nem verdadeira nem falsa), então  $\langle \sim p \rangle$  também é uma sentença patológica. Se “o maior número primo é ímpar” não é verdadeira nem falsa, então “o maior número primo não é ímpar” também não é verdadeira nem falsa. Mas uma disjunção com duas sentenças que não são verdadeiras nem falsas dificilmente permite uma interpretação que a torne verdadeira. Isto parece ser uma das consequências indesejáveis da aceitação de *gaps*. Mesmo as teorias que contornam a acusação de inconsistência (a acusação de que defender **GAP** significa afirmar coisas contraditórias), distinguindo a expressão “é falsa” da expressão “não é verdadeira” acabam por não avaliar o **PTE** da lógica como uma verdade lógica. Ainda que evitem negar o princípio (o que resultaria em afirmar uma contradição), elas não o preservam no sentido mais forte<sup>87</sup>.

<sup>86</sup> Mais tarde será introduzida uma regra que permite avaliar sentenças bicondicionais que envolvem *truth value gaps*.

<sup>87</sup> A solução de van Fraassen (1966, 68) para este problema é a *Superavaliação*. A *Superavaliação* é uma interpretação para linguagens em que ocorrem *truth value gaps* que permite manter todas as tautologias da lógica clássica. Uma sentença é verdadeira nesta interpretação se ela é verdadeira em todas as avaliações admissíveis. Embora “o atual rei da França é careca” seja uma sentença nem verdadeira nem falsa, assim como “o atual rei da França não é careca”, a sentença “o atual rei da França é careca ou não é careca” é verdadeira na *Superavaliação*, porque ela é verdadeira em todas as avaliações admissíveis. Pela mesma razão, uma sentença que é falsa em todas as avaliações admissíveis é falsa na *Superavaliação*. Esta solução, contudo, não procede dentro de uma semântica puramente fregiana, pois não está de acordo com o princípio de composicionalidade (que foi o motivador da aceitação de *gaps*). Na *Superavaliação* o valor de verdade das sentenças não é sempre uma função apenas dos valores das partes das sentenças.

Seguindo a tradição fregiana de *gaps*, a tese **GAP** afirma: não é o caso que toda sentença é sempre verdadeira ou falsa, há sentenças que não são verdadeiras nem falsas. Esta tese é ilustrada pelos fenômenos de falhas semânticas que mencionamos anteriormente (falhas de referência, *nonsenses*, e falhas pressuposicionais). Aqui os autores que, insatisfeitos com a Solução Ortodoxa do Paradoxo do Mentiroso, procuram por uma solução alternativa, encontram boas razões para incluir **PM** entre os exemplos de sentenças que satisfazem **GAP**. A solução, deste modo, consiste em afirmar que a sentença **PM** não é verdadeira nem falsa. E como já argumentamos, isto elimina o paradoxo, já que o paradoxo depende da suposição de que **PM** é verdadeira ou falsa. Mas:

Antes de nós começarmos a nos sentir convencidos demais a respeito disto, entretanto, nós devemos encarar um segundo paradoxo, o qual eu vou chamar *Mentiroso Reforçado* e que foi feito especialmente para aqueles filósofos esclarecidos que não são iludidos pela bivalência. (van Fraassen, 1968. p.147).<sup>88</sup>

Qual é o problema desta solução? O problema desta solução é que este modelo de *gaps* parece ser suficiente para avaliar os outros tipos de patologia da linguagem, mas não é suficiente para avaliar o Paradoxo do Mentiroso. De fato, ao afirmar que **PM** não é verdadeira, significando que ela não é verdadeira nem falsa, eliminamos o paradoxo. Mas esta avaliação deixa aberto o caminho para uma versão mais forte do paradoxo. Esta versão mais forte é o Paradoxo do Mentiroso Reforçado (**PMR**):

“esta própria sentença é falsa ou nem verdadeira nem falsa”<sup>89</sup>

Pela distinção fundamentada nesta interpretação dos *gaps*, “falso” difere de “não-verdadeiro”, e daí o **PB** não equivale ao **PBR**. Mas, dada esta distinção, esta versão do paradoxo não pressupõe mais o **PB**, e sim o **PBR**, que a interpretação pretende preservar. Mesmo que se admita que haja sentenças na linguagem que não são verdadeiras nem falsas, o paradoxo permanece. Suponha que toda sentença é sempre verdadeira ou falsa ou nem verdadeira nem falsa. Se avaliarmos **PMR** como verdadeira, então ela é falsa ou nem verdadeira nem falsa, ou seja, se ela é verdadeira, então ela não

---

<sup>88</sup> *Before we begin to feel too smug about this, however, we must face a second paradox, which I shall call the **Strengthened Liar** and which was designed especially for those enlightened philosophers who are not taken in by bivalence.*

<sup>89</sup> Ou simplesmente: “esta própria sentença não é verdadeira”.

é verdadeira. Se avaliarmos **PMR** como falsa, então ela é verdadeira, já que ser falsa é parte do que ela diz de si mesma (é um dos disjuntos). Mas se avaliarmos **PMR** como nem verdadeira nem falsa, então ela é verdadeira, já que um dos disjuntos diz exatamente que ela é nem verdadeira nem falsa. Logo, **PMR** é verdadeira se, e somente se, **PMR** não é verdadeira. E este paradoxo não depende de preservarmos o **PB**, mas apenas de preservarmos o **PBR**.

Os *gaps* fregianos, funcionam como uma forma eficaz de avaliar sentenças que possuem falhas de referência, sentenças *nonsense*, e mesmo a versão simples do Mentiroso. Mas o problema com uma possível solução do Paradoxo do Mentiroso com base neste modelo de *gaps* é que ela não é suficiente para eliminar uma versão muito semelhante do mesmo problema. Resolve a avaliação para a sentença “esta própria sentença é falsa”, mas não resolve satisfatoriamente a avaliação para a sentença “esta própria sentença não é verdadeira”<sup>90</sup>. As tentativas de solução que tentam classificar a sentença do mentiroso como nem verdadeira nem falsa (como a tentativa de van Fraassen, que retira consequências das relações pressuposicionais para a sentença do Mentiroso, ou a tentativa de Martin, que tenta classificar a sentença do Mentiroso como um *nonsense*), não são indisputavelmente satisfatórias. Evitar a abertura semântica da linguagem restringindo o **PB**, como estas tentativas de solução sugerem que deve ser entendida esta restrição, é pular da frigideira para o fogo. Eliminamos os problemas da Solução Ortodoxa ao custo de motivar uma versão reforçada do paradoxo, para a qual temos pouco que fazer.

Como dissemos, há duas vias de interpretação da tese de que há *truth value gaps* na linguagem. A primeira, que chamamos de tradição fregiana, entende que restringir o **PB** é negar tal princípio, e evita a acusação de inconsistência distinguindo o **PB** do **PTE** (o lógico e o semântico). Mas em última instância, vimos que tal interpretação não torna viável uma solução para o Paradoxo do Mentiroso, além de não preservar o **PTE** da lógica no sentido forte. A seguir vamos abordar a teoria de Kripke para o Paradoxo do Mentiroso. Teremos a oportunidade de entender a segunda via de interpretação dos *truth*

---

<sup>90</sup> Van Fraassen, na sua tentativa de solução para o Paradoxo do Mentiroso a partir da semântica pressuposicional (1968), apresenta uma solução que considera satisfatória para o Mentiroso Reforçado. O problema desta solução é que vamos ter que avaliar a sentença que diz que **PMR** não é verdadeira ela mesma como não verdadeira. Aliás, a tentativa de explicar que tipo de relação direta a pressuposição tem com a sentença do Mentiroso não é exatamente clara. O ponto a ser considerado acerca de relações pressuposicionais para a autorreferência é que toda sentença que pressupõe a si mesma não pode ser falsa, mas apenas verdadeira ou nem verdadeira nem falsa. Uma outra consequência estranha disto é que algumas sentenças claramente falsas e inócuas não poderão ser avaliadas como falsas, por exemplo: “esta própria sentença está escrita em inglês”.

*value gaps*, que preserva a equivalência entre o **PB** e o **PTE** (o semântico e o lógico), mas evita a acusação de inconsistência reinterpretando o sentido de “restringir a bivalência”.

### 3.2 A Solução de Kripke para o Paradoxo do Mentiroso

A solução que Kripke apresenta para o Paradoxo do Mentiroso está inserida em um conjunto de considerações que ele reuniu para um simpósio acerca da verdade em 1975, com o título *Outline of a Theory of Truth*. Embora ele não apresente os seus resultados de maneira sistematicamente formal, suprimindo do texto, sobretudo, as provas dos teoremas relevantes (já que ele se concentra mais na semântica do que na sintaxe da linguagem<sup>91</sup>), ainda assim, uma apresentação fiel da sua solução exige a consideração de aspectos bastante técnicos. Mas ao mesmo tempo, os seus resultados estão direcionados às questões filosóficas que temos discutido até aqui, e isto precisa ser bem destacado. Vamos dividir a seção presente em três subseções: em 3.2.1 apresentaremos a solução de Kripke de uma maneira bastante informal e simplificada, de modo a permitir uma compreensão geral, e mesmo para facilitar a leitura da subseção seguinte. Em 3.2.2 apresentaremos alguns dos resultados formais que acompanham a solução de Kripke<sup>92</sup>, tornando precisos os conceitos utilizados. Em 3.2.3 retomaremos os resultados de Kripke inserindo-os no quadro geral dos problemas que temos tratado até agora.

#### 3.2.1 Explicação informal da solução de Kripke

Assim como as Soluções Alternativas em geral, a solução de Kripke para o Paradoxo do Mentiroso consiste em restringir a lógica clássica de modo a evitar que as linguagens semanticamente fechadas derivem versões de **PM**, mais especificamente restringindo o **PB**, admitindo *truth value gaps*. Mas a solução de Kripke, veremos, produz razões genuínas para justificar tal restrição, e, além disto, produz também um

---

<sup>91</sup> Feferman, S. (1982) construiu um sistema axiomático que suporta os modelos de Kripke. Esta teoria é também conhecida como **KF**. Soames também deu algumas contribuições a esta teoria (1999) gerando a teoria **KFS**. Uma interpretação alternativa é fornecida por Maudlin (2004) e conhecida como **FM**. Neste trabalho não faremos distinção entre estas teorias ou mencionaremos diferenças de interpretação, apenas consideraremos as repercussões do texto de Kripke de uma maneira geral e bastante simplificada.

<sup>92</sup> Mas como já foi dito, por “resultados formais” não entendemos a apresentação sistemática de uma teoria, mas apenas um esboço da sua semântica.

modelo eficiente para expressar esta restrição. Na medida em que Kripke assume *gaps*, pode tomar o predicado verdade como um predicado comum para toda a linguagem, e pode evitar os constrangimentos da hierarquia de linguagens.

Na solução de Kripke, a distinção usual em Filosofia da Linguagem entre sentenças e proposições é retomada. Embora, tenhamos inicialmente excluído esta distinção, vamos utilizá-la a partir de agora. Uma sentença é um conjunto bem formado de expressões que utilizamos (enunciamos) em determinados contextos a fim de expressar uma proposição. Uma sentença particular difere de qualquer outra sentença particular, embora elas possam exemplificar a mesma sentença geral. A ocorrência particular da sentença “a neve é branca” escrita no quadro difere da ocorrência particular da sentença “a neve é branca” falada por um indivíduo, e cada ocorrência particular é única, mas ambas, neste caso, são ocorrências particulares de uma mesma sentença geral<sup>93</sup>. A proposição é o que é expresso por uma sentença. Dois usos assertivos de uma mesma sentença em diferentes contextos podem produzir diferentes proposições. Se Tarski diz em 25/03/1929 “eu estou cansado”, ele expressa uma proposição. Se Maudlin afirma em 25/03/2010 “eu estou cansado”, ele também expressa uma proposição, mas que é diferente da proposição expressa por Tarski, embora ambos tenham utilizado a mesma sentença (a mesma sentença tipo). Se Maudlin quer expressar a mesma proposição que Tarski expressou, ele deve afirmar outra sentença: “Tarski estava cansado em 25/03/1929”.

Kripke entende que toda sentença que expressa uma proposição é sempre verdadeira ou falsa, mas que há sentenças que não expressam uma proposição, não são verdadeiras nem falsas. Mas ele também entende que a falha de um enunciado em expressar uma proposição não é uma falha de significado. Suponha a sentença “o atual rei da França é careca”. Esta sentença não expressa uma proposição, mas há circunstâncias especificáveis sob as quais esta sentença tem condições de verdade bem determinadas, e assim, expressaria uma proposição. Portanto, ela possui significado. Kripke afirma que ele está sob a influência de Strawson, com a diferença de que Strawson não afirma que “o atual rei da França é careca” possui significado<sup>94</sup>. Kripke está procurando caracterizar uma falha que não é uma falha de significado, mas que interfere na constituição da proposição, interfere no que é expresso pela sentença. A

---

<sup>93</sup> Sentenças gerais são classes de sentenças particulares. Os termos mais corretos são *sentence types* (sentenças tipo) e *sentence tokens* (sentenças enquanto ocorrências particulares).

<sup>94</sup> Cf. Kripke (1975, pp.63-64).

idéia é que algumas sentenças possuem significado, mas nas circunstâncias atuais não há condições que as tornem verdadeiras ou falsas. Tais sentenças não são verdadeiras nem falsas, são *gaps*<sup>95</sup>.

Mas os fenômenos para os quais Kripke quer formular uma teoria de *truth value gaps* não são os fenômenos usuais que motivam tal teoria, como as falhas pressuposicionais, as falhas de referência ou os erros categoriais. Kripke procura uma teoria que justifique *gaps* para as sentenças que envolvem círculos viciosos com os predicados de verdade. Como veremos adiante, esta teoria irá permitir caracterizar tipos específicos de sentenças patológicas, como as sentenças *infundadas* e as sentenças paradoxais.

As considerações de Kripke partem de uma elucidação do significado intuitivo dos predicados de verdade na linguagem. Segundo Kripke:

Nós queremos capturar uma intuição de algo da seguinte espécie. Suponha que nós estamos explicando a palavra “verdadeiro” para alguém que não a conhece ainda. Nós podemos dizer que nós estamos no direito de asserir (ou rejeitar) de qualquer sentença que ela é verdadeira precisamente sob as circunstâncias em que nós podemos asserir (ou rejeitar) a sentença ela mesma. (Kripke, 1975, p. 65).<sup>96</sup>

Portanto, se queremos explicar quando nós atribuímos o predicado de verdade, devemos pressupor que o nosso interlocutor possui meios para determinar quais sentenças ele deve asserir ou rejeitar. Se o nosso interlocutor conhece o significado da palavra “neve” e sabe atribuir o predicado “é branca” e o predicado “é verde” (o que obviamente depende de ele conhecer fatos do mundo), ele sabe que está em condições de asserir “a neve é branca” e está em condições de rejeitar “a neve é verde”. Sob

---

<sup>95</sup> Kripke não se dedica muito a esta distinção entre não possuir significado e não expressar uma proposição. O que ele parece entender é: uma sentença possui significado se ao menos há circunstâncias especificáveis (mesmo que não atuais) sob as quais a sentença tem condições de verdade, enquanto que uma sentença expressa uma proposição se ela tem condições de verdade atualmente. Uma sentença não possui significado se não há circunstâncias especificáveis possíveis em que ela possui condições de verdade. Seguindo o exemplo que Kripke aponta, “o atual rei da França não é careca” possui significado, pois no *contexto* em que há um rei na França as condições de verdade da sentença são claramente determináveis, mas ela não expressa uma proposição, já que no *contexto* atual as condições de verdade da sentença não são determináveis. Kripke, entretanto, não fornece uma teoria sobre o significado, e esta interpretação que oferecemos é apenas especulativa.

<sup>96</sup> *We wish to capture an intuition of somewhat the following kind. Suppose we are explaining the word ‘true’ to someone who does not yet understand it. We may say that we are entitled to assert (or deny) of any sentence that it is true precisely under the circumstances when we can assert (or deny) the sentence itself.*

exatamente as mesmas condições, ele pode asserir “‘a neve é branca’ é verdadeira” e pode rejeitar “‘a neve é verde’ é verdadeira”. Generalizando este processo, poderemos tornar claro para o nosso interlocutor como utilizar o predicado de verdade.

Mas e se uma sentença envolve nela mesma o predicado “verdadeiro”? Suponha a sentença “alguma sentença impressa no *New York Daily News*, 7 de Outubro de 1971, é verdadeira”. Nosso interlocutor sabe que está em condições de asserir que esta sentença é verdadeira se, e somente se, está em condições de asserir a sentença ela mesma. Mas ele só está em condições de asserir esta sentença se ele está em condições de asserir que alguma sentença impressa no *New York Daily News* é verdadeira, e por sua vez, só está em tais condições se ele está em condições de asserir alguma sentença impressa no *New York Daily News*. O processo de atribuição dos predicados de verdade depende de podermos retroceder até chegarmos a sentenças que em última instância não envolvem mais predicados de verdade. Pois, se isto não ocorre, nosso interlocutor não está em condições nem de asserir nem de rejeitar que tal sentença é verdadeira. Quando isto ocorre dizemos que tal sentença é *infundada*. Uma sentença é *fundada*<sup>97</sup> se ela recebe um valor de verdade neste processo, e ela é *infundada* se ela não recebe um valor de verdade neste processo. Um exemplo de uma sentença *infundada* é “esta própria sentença é verdadeira”<sup>98</sup>. Não estamos em condições de asserir ou rejeitar que esta sentença é verdadeira, já que não estamos em condição de asserir ou rejeitar a sentença ela mesma, pois isto envolveria asserir ou rejeitar que ela é verdadeira.

A idéia de Kripke pode ser ilustrada pela seguinte situação<sup>99</sup>. Imagine que vamos escrever dois grandes livros, o Verdadeiro e o Falso, construindo a extensão destes dois predicados para uma linguagem. Para isto, dispomos de um domínio **D**, que inclui os objetos e as sentenças denotados por nomes da nossa linguagem. Para uma explicação mais simples, vamos supor inicialmente que nosso domínio é finito e bem pequeno: **D** = {1. neve, 2. grama, 3. “a neve é branca”, 4. “a grama é verde”, 5. “a neve é verde”, 6. “‘a neve é branca’ está impressa no *Daily News*”, 7. “‘a neve é branca’ é verdadeira”, 8. “‘a neve é branca’ é falsa”, 9. “algo impresso no *Daily news* é verdadeiro”, 10. “‘algo impresso no *Daily news* é verdadeiro’ é verdadeira”, 11. “esta própria sentença é falsa”, 12. “esta própria sentença é verdadeira”}<sup>100</sup>. A nossa linguagem, portanto, tem os

<sup>97</sup> *Grounded* (cf. Kripke, 1975, p. 66).

<sup>98</sup> O nome comum para esta sentença na literatura é *Truth Teller*, o contador da verdade. Vamos abreviá-la como “**CV**”.

<sup>99</sup> Esta ilustração é uma adaptação de JC Beall (*Appendix*, in *Revenge of the Liar*, 2007).

<sup>100</sup> Nomeamos os objetos do nosso domínio com números para facilitar a menção a eles daqui pra frente.

seguintes predicados: “é branco”, “é verde”, “está impresso no *Daily News*”, “é verdadeiro” e “é falso”. A idéia é construir os dois livros dando a interpretação dos predicados de acordo com a finalidade do livro. Vamos construir estes livros listando sentenças em capítulos, e a regra para entrar nestes capítulos será dada por uma definição indutiva. Antes de iniciarmos a construção de ambos os livros, devemos dar a extensão dos outros predicados da linguagem: “é branco” = {neve}, “é verde” = {grama}, “está impresso no *Daily News*” = {“a neve é branca”}. Vamos supor que esta interpretação prévia dos outros predicados é dada no prefácio do livro, é a base para a construção. Neste prefácio “é verdadeiro” =  $\{\emptyset\}$  e “é falso” =  $\{\emptyset\}$ . Isto espelha o fato de que no nível mais básico do processo não asserimos a verdade ou falsidade de sentenças, mas apenas as sentenças que não envolvem predicados de verdade. A partir do capítulo I, cada um dos livros constrói respectivamente a extensão de um dos nossos predicados de verdade (“verdadeiro” e “falso”) em capítulos. As regras para a construção dos capítulos sucessivamente são as seguintes:

a) Uma sentença está no capítulo **n** do livro Verdadeiro sse ela está no capítulo **n – 1** deste livro ou ela é uma sentença tal que o nome na sentença denota um objeto que está na extensão do seu predicado em **n – 1**.

b) Uma sentença está no capítulo **n** do livro Falso sse ela está no capítulo **n – 1** deste livro ou ela é uma sentença tal que o nome na sentença denota um objeto que está na antiextensão<sup>101</sup> de seu predicado em **n – 1**<sup>102</sup>.

O capítulo I do Verdadeiro vai ter como elementos as sentenças que estão na extensão de “é verdadeiro” do prefácio mais as sentenças da linguagem cujo objeto denotado pela expressão de sujeito na frase esteja na extensão do seu predicado no prefácio. Bom, no prefácio a extensão de “é verdadeiro” e “é falso” estão vazias, logo nenhuma sentença envolvendo predicados de verdade entra no capítulo I. As sentenças que entram no capítulo um do Verdadeiro são 3, 4 e 6, já que a extensão de seus predicados é dada no prefácio e os objetos mencionados pertencem a tais extensões. A sentença que entra no capítulo I do Falso é 5. Assim que construímos o capítulo I, temos uma nova interpretação para “é verdadeiro” e “é falso”. No prefácio as extensões destes predicados eram vazias, agora, temos como extensão dos predicados de verdade,

<sup>101</sup> A antiextensão de um predicado é o conjunto das coisas que não satisfazem aquele predicado. A antiextensão de “é verde” = {neve}. A antiextensão dos predicados também é fornecida no prefácio.

<sup>102</sup> Neste domínio simplificado, não ocorrem sentenças envolvendo conectivos sentenciais. Para um domínio mais complexo, certamente será necessário suplementar estas regras com cláusulas para os diferentes conectivos e quantificadores.

respectivamente, as sentenças verdadeiras e falsas do capítulo I. A extensão do predicado “é verdadeiro” em I é {3, 4, 6}, e a extensão de “é falsa” em I é {5}. No capítulo II, temos: Verdadeiro = {3, 4, 6, 7, 9}, e Falso = {5, 8}. Os elementos que entram no capítulo II são todos aqueles que entram no capítulo I mais aquelas sentenças cujo nome denota um objeto na extensão do seu predicado em I. 3, 4 e 6 entram no capítulo II porque já estavam no capítulo I. 7 e 9 entram no capítulo II porque no capítulo I o predicado “é verdadeiro” inclui 3 na sua extensão. 8 entra no capítulo II do Falso porque no capítulo I o predicado “é falso” inclui 3 como antiextensão. As extensões de Verdadeiro e Falso aumentam, pois a partir do capítulo I, algumas sentenças foram sendo acrescentadas à extensão de “é verdadeiro” e “é falso”, logo, sentenças que envolvem estes predicados podem resultar verdadeiras ou falsas, entrando em capítulos consecutivos dos livros. No capítulo III, teremos: Verdadeiro = {3, 4, 6, 7, 9, 10}, e Falso = {5, 8}. Na medida em que os capítulos se seguem, mais sentenças vão entrando nos livros, mas nenhuma sentença muda de um livro para outro ou sai de ambos os livros. Uma sentença como 10 pode não pertencer nem ao capítulo I do Verdadeiro nem ao capítulo I do Falso, mas no capítulo III ela pertence ao Verdadeiro.

O que acontece é que, em algum capítulo, o conjunto das sentenças da nossa linguagem que podem resultar verdadeiras ou falsas é totalmente determinado, de modo que todo capítulo seguinte é igual a tal capítulo:  $\mathbf{n} = \mathbf{n} + 1 = \mathbf{n} + 2 = \mathbf{n} + 3$ , etc<sup>103</sup>. É fácil ver que no capítulo IV: Verdadeiro = {3, 4, 6, 7, 9, 10}, e Falso = {5, 8}, e também no capítulo V, VI, etc. Quando isto acontece, concluímos os nossos livros. As sentenças 11 e 12 não entraram em nossos livros. Sentenças como 11 e 12 não recebem valor de verdade neste processo já que, para todo capítulo, se perguntarmos se 12 pertence ao Verdadeiro, ela pertence a este capítulo se, e somente se, ela mesma está na extensão de Verdadeiro do capítulo anterior, e assim por diante, até chegarmos ao prefácio onde a extensão dos predicados de verdade é vazia, e o mesmo ocorre se perguntarmos se ela pertence ao Falso. Com a sentença 11, para cada capítulo do Verdadeiro, ela pertence a ele se, e somente se, ela mesma pertence ao capítulo anterior do Falso, e Para cada capítulo do Falso ela pertence a ele se, e somente se, ela mesma pertence ao capítulo anterior do verdadeiro, e assim, tal sentença igualmente resulta sem valor no processo.

Vamos colocar de maneira um pouco mais clara o que ocorre nesta situação. Distinguímos i) a extensão dos predicados “é verdadeira” e “é falsa” de ii) o que resulta

---

<sup>103</sup> Na seção seguinte veremos que isto corresponde ao ponto fixo *minimal* da função que nos dá os conjuntos das sentenças verdadeiras e falsas da linguagem.

como verdadeiro ou falso. A interpretação dos predicados no prefácio determina a extensão dos predicados de verdade do capítulo I. A extensão dos predicados de verdade do capítulo I diz o que é verdadeiro ou falso no prefácio. Mas a extensão dos predicados de verdade do capítulo I determina a extensão do capítulo II. E a extensão dos predicados de verdade do capítulo II diz o que é verdadeiro ou falso no capítulo I. Em cada capítulo dos livros, temos um conjunto das sentenças que resultam verdadeiras e um conjunto das sentenças que resultam falsas. Colocamos estes conjuntos respectivamente nas extensões dos predicados “é verdadeira” e “é falsa” e obtemos no capítulo seguinte um novo conjunto do que resulta verdadeiro e um novo conjunto do que resulta falso. Isto é equivalente a uma hierarquia de linguagens onde os predicados de verdade de uma linguagem são dados na sua metalinguagem. Mas para algum capítulo, as extensões dos predicados coincidem com os conjuntos das sentenças que resultam verdadeiras e falsas. Neste ponto, temos a interpretação que nos permite avaliar o que é verdadeiro e o que é falso na nossa linguagem como um todo. Quando isto ocorre a nossa linguagem é fechada (é semanticamente fechada, no sentido desaconselhado por Tarski), não há mais hierarquia. Neste capítulo temos uma interpretação dos predicados de verdade para uma linguagem unívoca. Resulta nesta linguagem que algumas sentenças não recebem valor de verdade.

O que este processo nos diz sobre o modo como os predicados de verdade se comportam na linguagem é que eles exibem relações de dependência semântica. Toda sentença que envolve um predicado de verdade depende semanticamente da sentença para a qual este predicado é atribuído. Se nós entendermos que os predicados de verdade funcionam como operadores (como a negação, por exemplo, que toma sentenças e forma novas sentenças com o valor inverso), então, isto fica mais claro. Suponha que “é verdadeiro” retorna verdadeiro sse a sentença que entra é verdadeira, e falso sse a sentença que entra é falsa. E que “é falso” retorna verdadeiro sse entra falso, e retorna falso sse entra verdadeiro. Então, o valor de uma sentença com predicados de verdade depende do valor da sentença componente. Se uma sentença com predicados de verdade não pode ser decomposta em partes que não envolvem predicados de verdade e que possuam um valor de verdade no nível básico (no prefácio), então esta sentença é *infundada*.

Enfim, a intuição que Kripke pretende capturar sobre os predicados de verdade é algo bastante semelhante ao esquema (T), porém com uma significativa diferença: nesta interpretação, uma sentença  $p$  é verdadeira sse  $\langle \text{Tr}(\langle p \rangle) \rangle$  é verdadeira, e  $p$  é falsa sse

$\langle \text{Tr}(\langle p \rangle) \rangle$  é falsa. Dissemos que alguém está em condições de asserir (ou rejeitar) que uma sentença é verdadeira sse está em condições de asserir (ou rejeitar) a sentença ela mesma. Até aqui as coisas parecem funcionar de maneira semelhante ao esquema (T). Porém, para algumas sentenças, não estamos em condições de asserir ou rejeitar a sentença ela mesma, de modo que, estas sentenças não possuem valor de verdade na nossa interpretação. A intuição que Kripke pretende capturar fornece as condições suficientes para a atribuição de valores verdade a algumas sentenças, mas não fornece condições que sejam conjuntamente necessárias e suficientes para a atribuição de valores para todas as sentenças. Os predicados de verdade não são exaustivos em relação ao domínio, pois há sentenças do domínio que não pertencem a  $\text{Tr} \cup \text{Fs}$ .

Os *gaps* decorrentes desta interpretação caracterizam linguagens onde os predicados de verdade são apenas parcialmente definidos, ou seja, não são definidos para todos os elementos do domínio. Assim sendo, quando uma sentença cai no *gap* não dizemos que ela é nem verdadeira nem falsa, ou tampouco dizemos que ela é nem verdadeira nem não verdadeira. Em primeiro lugar, Kripke não defende a distinção entre “falsa” e “não verdadeira”, já que o que ele pretende não é sustentar que sentenças sem valor de verdade são não verdadeiras. Kripke segue Tarski em assumir a equivalência entre “falsa” e “não verdadeira”, e o recurso que ele vai buscar para evitar os argumentos que levam da rejeição do **PB** e do **PTE** a uma contradição é outro, como veremos mais tarde<sup>104</sup>. Em segundo lugar, quando uma sentença cai no *gap*, o que podemos dizer, de acordo com a interpretação de Kripke, é que não estamos em condições de asserir a verdade da sentença nem de asserir a verdade da negação da sentença (ou rejeitar a verdade da sentença, ou asserir a falsidade da sentença). Dizer desta sentença que ela é nem verdadeira nem falsa nos comprometeria a asserir que ela não é verdadeira, ou rejeitar que ela é verdadeira, o que só seria autorizado se estivéssemos em condições de rejeitar a sentença ela mesma. Pela mesma razão, não dizemos que a sentença é não verdadeira. Estas sentenças não possuem valor de verdade no sentido forte de “não possuem valor de verdade”, não estamos em condições de asserir ou rejeitar delas qualquer predicado de verdade ou a negação de qualquer

---

<sup>104</sup> Como vimos antes, a tradição fregiana de *gaps* evita estes argumentos exatamente sustentando esta distinção.

predicado de verdade. Silenciamos-nos<sup>105</sup> a respeito de se elas são verdadeiras ou não. Os predicados de verdade não são definidos para elas.

Suponha uma sentença que não recebe valor no processo que ilustramos: “esta própria sentença é falsa”, também conhecida como **PM**. O que ocorre é que **PM** não possui valor, não podemos asserir a verdade de **PM** nem rejeitar a verdade de **PM**, pois não estamos em condições de asserir ou rejeitar **PM**. Assim, **PM** não possui valor, “**PM** é verdadeira” também não possui valor, “**PM** é falsa” e “**PM** não é verdadeira” também não possuem valor. Para expressar os *gaps* de Kripke, precisamos convencionar que o que quer que represente estes *gaps* significa ausência de valor de verdade no sentido forte de ausência.

Na medida em que assumirmos *truth value gaps*, precisamos de um conjunto de regras que nos permita rever a interpretação dos conectivos e dos quantificadores da linguagem. Na lógica clássica, a interpretação dos conectivos e quantificadores pressupõe que toda sentença é sempre ou verdadeira ou falsa. Como vamos escolher estas regras? Kripke afirma que não há problema em escolher outras lógicas que expressem *gaps* distintas da que ele utiliza<sup>106</sup>, desde que elas preservem certas características que mencionaremos mais tarde. Mas ele tende a preferir a lógica forte de Kleene, pois ela é mais compatível com a idéia de que o valor atribuído a sentenças que caem no *gap* é uma mera convenção para expressar uma ausência real de valor verdade, e não um terceiro valor de verdade. Kleene toma estas sentenças como **indefinidas**. A lógica forte de Kleene preserva a intuição de que algumas sentenças podem não expressar uma proposição e ainda entrarem como componentes de sentenças complexas, que por sua vez, expressam uma proposições<sup>107</sup>. Na lógica forte de Kleene, a sentença “esta própria sentença é falsa” não possui valor, mas a sentença “a neve é branca ou esta

---

<sup>105</sup> Mas aqui silenciar não é no sentido epistêmico de ser agnóstico a respeito do valor de uma sentença. Somos agnósticos a respeito do valor de uma sentença quando supomos que ela é certamente verdadeira ou falsa, mas não sabemos se é verdadeira ou se é falsa. No caso presente, não temos uma questão de ignorância (cf. Field, H. 2008, p. 70-71), mas de inaplicabilidade. Silenciamos-nos quanto ao valor de verdade de uma determinada sentença porque valores de verdade não são definidos para ela.

<sup>106</sup> Kripke menciona como lógicas alternativas para suportarem sua construção a lógica de *superavaliação* de van Fraassen e a lógica fraca de Kleene.

<sup>107</sup> Segundo Kripke, esta é mais uma razão para distinguir falha de significado de falha em expressar uma proposição. Kripke pensa que uma sentença complexa que envolve uma parte sem significado é ela mesma toda sem significado. Por outro lado, ele quer defender que se uma parte de uma sentença não expressa uma proposição, a sentença ainda pode expressar uma proposição (cf. Kripke 1975, p.64). Kripke responderia diferentemente às duas questões seguintes. Se o significado de parte de uma sentença falha, o significado da sentença como um todo deve falhar? Se a determinação de valores de verdade de parte de uma sentença falha, então a determinação de valores de verdade da sentença como um todo deve falhar? Kripke diria “sim” para a primeira pergunta, e diria “não” para a segunda. Mas se ele identificasse significado com proposição, ele teria de dar a mesma resposta para as duas perguntas.

própria sentença é falsa” é verdadeira, e, logo, expressa uma proposição. Pela regra clássica da disjunção, basta que um termo seja verdadeiro para a disjunção resultar verdadeira. Neste caso, um termo é verdadeiro e o outro não tem valor. Por esta interpretação, seguimos mais de perto a idéia de que *gaps* expressam uma falha de sentenças em expressarem proposições.

A interpretação forte de Kleene para os conectivos é a seguinte. Negação: a negação de uma sentença é verdadeira sse a sentença é falsa, e é falsa sse a sentença é verdadeira. Em todos os outros casos, ela é indefinida (ou seja, quando a sentença é indefinida). Disjunção: uma disjunção de duas sentenças é verdadeira sse pelo menos uma das sentenças é verdadeira, é falsa sse ambas as sentenças são falsas, e de outro modo ela é indefinida (ou seja, quando nenhuma é verdadeira e pelo menos uma não é falsa). Conjunção: uma conjunção entre duas sentenças é verdadeira sse ambas as sentenças componentes são verdadeiras, é falsa sse pelo menos uma é falsa, e é indefinida de outro modo (ou seja, se não é o caso que ambas são verdadeiras e se pelo menos uma não é falsa). Condicional: uma sentença condicional é verdadeira sse o antecedente é falso ou o conseqüente é verdadeiro, é falsa sse o antecedente é verdadeiro e o conseqüente é falso, e é indefinida de outro modo (ou seja, quando o antecedente é indefinido e o conseqüente é falso ou indefinido, e quando o antecedente é verdadeiro e o conseqüente é indefinido). Bicondicional: uma sentença bicondicional é verdadeira sse ambas são verdadeiras ou ambas são falsas, é falsa sse uma é verdadeira e a outra é falsa, e é indefinida de outro modo (ou seja, quando pelo menos uma é indefinida).

A interpretação dos quantificadores é a seguinte. Quantificador Universal: uma sentença com quantificador universal é verdadeira sse todas as instâncias (que resultam de substituir a variável por um nome de algum objeto do domínio) são sentenças verdadeiras, é falsa sse pelo menos uma instância é uma sentença falsa, e é indefinida de outro modo (ou seja, quando pelo menos uma instância não é uma sentença verdadeira e nenhuma instância é uma sentença falsa). Quantificador Existencial: uma sentença com quantificador existencial é verdadeira sse pelo menos uma instância é uma sentença verdadeira, é falsa sse todas as instâncias são sentenças falsas, e é indefinida de outro modo (ou seja, quando nenhuma instância é verdadeira e pelo menos uma instância não é falsa).

Além da interpretação dos conectivos e dos quantificadores, vale estipular ainda uma interpretação dos predicados de verdade, como se eles se comportassem como

operadores. Como dissemos, o predicado “é falsa” é equivalente ao predicado “não é verdadeira”. Basta que tenhamos uma interpretação para “é verdadeira”: uma sentença com o predicado “é verdadeira” é verdadeira sse a sentença componente é verdadeira, é falsa sse a sentença componente é falsa, e é indefinida de outro modo (ou seja, quando a sentença componente é indefinida).

Dadas as regras de interpretação, podemos avaliar o valor de sentenças complexas. A sentença “a neve é verde ou esta própria sentença é falsa” resulta indefinida. A sentença “se esta própria sentença é verdadeira, então a neve é branca” resulta verdadeira. A sentença “**PM** é verdadeira ou **PM** é falsa” é indefinida. E a sentença “todas as sentenças são verdadeiras ou falsas” é indefinida, pois no nosso domínio **D** temos pelo menos uma sentença que não é verdadeira (é indefinida) e nenhuma instância torna falsa a disjunção sob o escopo deste quantificador universal.

De maneira geral, apresentamos a idéia de Kripke acerca do comportamento dos predicados de verdade e sua caracterização das sentenças patológicas. Apresentamos ainda as regras de interpretação que regimentam o uso dos quantificadores, dos conectivos e do predicado verdade na presença de *truth value gaps*. Passemos agora a uma breve apreciação dos pontos formais que subjazem a estes resultados.

### 3.2.2 Explicação formal da solução de Kripke

O que faremos aqui é apenas tornar mais precisa a ilustração apresentada na subseção anterior nos termos utilizados por Kripke. Vamos falar sobre a construção de modelos de ponto fixo para os predicados de verdade. Estes modelos permitem que Kripke forneça uma definição parcial para os predicados de verdade, e permitem classificar as sentenças *infundadas* e as sentenças paradoxais, inclusive distinguindo umas das outras.

A idéia é distinguir, inicialmente, i) a extensão dos predicados “é verdadeira” e “é falsa” de ii) o que é verdadeiro e o que é falso. Os modelos devem satisfazer também três *desiderata*: 1) Todas as sentenças verdadeiras da linguagem que não envolvem predicados de verdade devem permanecer verdadeiras na nossa interpretação. 2) Uma sentença  $p$  é verdadeira sob a nossa interpretação sse  $\langle \text{Tr}(\langle p \rangle) \rangle$  é verdadeira sob a nossa interpretação. 3) Uma sentença  $p$  é falsa sob a nossa interpretação sse  $\langle \text{Tr}(\langle p \rangle) \rangle$  é falsa sob a nossa interpretação. Mas como já foi bem apontado por Tarski, se a linguagem é semanticamente fechada e se nós preservarmos a lógica clássica, então não podemos

sustentar consistentemente estes três *desiderata*. O objetivo de Kripke é construir uma interpretação dos predicados de verdade (ou seja, dar uma extensão para “é verdadeira” e “é falsa”) que preserve o fechamento semântico característico da linguagem natural e seja consistente. Esta interpretação será dada pela extensão que é um ponto fixo da função que dá o conjunto das sentenças verdadeiras e falsas da linguagem. Neste ponto fixo ocorre que algumas sentenças não possuem valor de verdade, e assim, as sentenças patológicas não ameaçam a consistência da linguagem.

Temos uma linguagem  $\mathbf{L}$  sem predicados de verdade, que possui um domínio  $\mathbf{D}$ , e cada predicado  $F(x)$  de  $\mathbf{L}$  tem um par  $\langle S_1, S_2 \rangle$  que dá a extensão ( $S_1$ ) e a antiextensão ( $S_2$ ) de  $F(x)$ . O predicado é verdadeiro dos elementos da sua extensão e falso dos elementos da sua antiextensão. Quando temos um predicado parcial, para alguns elementos do domínio ele é indefinido, estes elementos não pertencem a  $S_1 \cup S_2$ .

A seguir, adicionamos a  $\mathbf{L}$  o predicado  $\text{Tr}(x)$ , e obtemos a linguagem  $\mathcal{L}$  com um novo domínio  $\mathbf{D}$  que tem objetos e sentenças. Para o predicado  $\text{Tr}(x)$ , temos também um par  $\langle S_1, S_2 \rangle$  que é a interpretação de  $\text{Tr}(x)$  no domínio  $\mathbf{D}$ . Acontece que, dada uma interpretação de  $\text{Tr}(x)$ , um conjunto de sentenças verdadeiras e um conjunto de sentenças falsas são obtidos. E não necessariamente estes conjuntos coincidem com a extensão e a antiextensão de  $\text{Tr}(x)$ . Na ilustração que fizemos na subseção anterior, a extensão e a antiextensão de  $\text{Tr}(x)$  no prefácio são vazias, mas os conjuntos das sentenças que resultam verdadeiras e das sentenças que resultam falsas no prefácio não são vazios, pois utilizando a interpretação dada para os outros predicados mais as regras de Kleene, podemos avaliar todas as outras sentenças da linguagem que não envolvem essencialmente predicados de verdade. As sentenças que neste nível são avaliadas como verdadeiras ou falsas, podem, então, no nível seguinte, serem adicionadas como interpretação de  $\text{Tr}(x)$ . Para cada novo capítulo, um novo par  $\langle S_1, S_2 \rangle$  é fornecido como a interpretação de  $\text{Tr}(x)$ , e disto um novo conjunto de sentenças que resultam verdadeiras e um novo conjunto de sentenças que resultam falsas na linguagem são obtidos. Estes novos conjuntos obtidos são, por conseguinte, utilizados para interpretar  $\text{Tr}(x)$  e um novo par de conjuntos é obtido, assim por diante.

Vamos utilizar a função  $\varphi(x)$ , uma função sobre conjuntos de conjuntos, que toma pares do tipo  $\langle S_1, S_2 \rangle$  e retorna pares do tipo  $\langle S_1', S_2' \rangle$ ,  $\varphi(\langle S_1, S_2 \rangle) = \langle S_1', S_2' \rangle$ , onde  $S_1'$  é o conjunto das sentenças que resultam verdadeiras na linguagem em que  $S_1$  é a

extensão de  $\text{Tr}(x)$ , e  $S_2'$  é o conjunto dos objetos do domínio que não são sentenças<sup>108</sup> ou são sentenças que resultam falsas na linguagem em que  $S_2$  é a antiextensão de  $\text{Tr}(x)$ . A função  $\varphi(x)$  é uma função que nos fornece a interpretação de uma linguagem a partir de uma interpretação de  $\text{Tr}(x)$ , nos fornece a interpretação de  $\mathcal{L}(\langle S_1, S_2 \rangle)$ , ou seja, ela nos dá o que é verdadeiro e o que é falso na linguagem em que a interpretação de  $\text{Tr}(x)$  é  $\langle S_1, S_2 \rangle$ .  $\varphi(x)$ , porém, não é definida para todas as sentenças do domínio. Há sentenças do domínio que não entram em nenhuma interpretação  $\langle S_1, S_2 \rangle$ . Consequentemente,  $S_1' \cup S_2'$  não esgota o domínio. Algumas sentenças do domínio não são verdadeiras nem falsas.

Bem, o que acontece é que a função  $\varphi(x)$  possui pontos fixos. Em geral, um ponto fixo de uma função qualquer  $f(x)$  é um valor de  $x$  que torna  $f(x) = x$ . Existem pares do tipo  $\langle S_1, S_2 \rangle$  que tornam  $\varphi(\langle S_1, S_2 \rangle) = \langle S_1, S_2 \rangle$ , e assim,  $S_1 = S_1'$  e  $S_2 = S_2'$ . Nestes pontos fixos, a extensão do predicado verdade coincide com as sentenças verdadeiras da linguagem, e a antiextensão do predicado verdade coincide com as sentenças falsas da linguagem. Quando a interpretação do predicado verdade coincide com a interpretação da linguagem (a avaliação que fornece quais sentenças são verdadeiras e quais são falsas), podemos dizer que esta linguagem é semanticamente fechada, o predicado verdade é unívoco na linguagem (a linguagem possui seu próprio predicado verdade), e não precisamos mais ascender a uma outra linguagem para fornecer a interpretação do seu predicado verdade.

Kripke afirma a existência de vários pontos fixos da função  $\varphi(x)$ . Há o ponto fixo *minimal*, os pontos fixos *maximais*, e o ponto fixo *intrínseco*<sup>109</sup>. Para definirmos estas expressões que aqui classificam diferentes pontos fixos, precisamos primeiro definir a relação “estender uma interpretação de  $\text{Tr}(x)$ ”. Esta relação é uma relação que se dá entre interpretações de  $\text{Tr}(x)$  (entre pares do tipo  $\langle S_1, S_2 \rangle$ ). Simbolizamos esta relação

<sup>108</sup> Para evitar confundir as sentenças patológicas para as quais os predicados de verdade são indefinidos com outros tipos de objetos do domínio para os quais a função  $\varphi(x)$  também é indefinida, Kripke estipula que os objetos do domínio que não são sentenças estão na antiextensão de  $\text{Tr}(x)$ . Mas não vamos manter isto com rigor neste trabalho.

<sup>109</sup> No seu artigo (1975), Kripke constrói estes pontos fixos mudando a interpretação dos predicados. Mas a existência destes pontos fixos da função  $\varphi(x)$  pode ser obtida rigorosamente, mediante provas e definições da teoria de conjuntos. Heck apresenta estas provas no texto *Kripke's Theory of Truth*, disponível para leitura no seu endereço web.

com “ $\geq$ ”<sup>110</sup>. Uma interpretação de  $\text{Tr}(x)$ ,  $\langle S_1', S_2' \rangle$ , estende uma interpretação de  $\text{Tr}(x)$ ,  $\langle S_1, S_2 \rangle$ , (ou seja,  $\langle S_1', S_2' \rangle \geq \langle S_1, S_2 \rangle$ ) sse  $S_1 \subseteq S_1'$  e  $S_2 \subseteq S_2'$ . Ou seja, uma interpretação  $\langle S_1', S_2' \rangle$  estende uma interpretação  $\langle S_1, S_2 \rangle$  sse toda sentença que pertence a  $S_1 \cup S_2$  também pertence a  $S_1' \cup S_2'$ . E como a função  $\varphi(x)$  é monotônica<sup>111</sup>, se  $\langle S_1', S_2' \rangle \geq \langle S_1, S_2 \rangle$ , então  $\varphi(\langle S_1', S_2' \rangle) \geq \varphi(\langle S_1, S_2 \rangle)$ . Retomando a ilustração dos livros (o Verdadeiro e o Falso), cada capítulo estende os capítulos anteriores. Pode haver sentenças indefinidas em uma interpretação de  $\text{Tr}(x)$  de um capítulo que se tornam verdadeiras ou falsas no capítulo seguinte, mas nenhuma sentença que é verdadeira ou falsa em um capítulo passa a ser indefinida ou muda de valor (de verdadeiro para falso ou vice versa) em capítulos posteriores. Esta é uma propriedade muito importante da relação “estender uma interpretação de  $\text{Tr}(x)$ ”<sup>112</sup>, já que se isto não ocorresse, poderíamos construir pontos fixos inconsistentes. Se uma sentença resultasse verdadeira em uma interpretação da linguagem e falsa em uma outra interpretação que estendesse a primeira, no ponto fixo, a mesma sentença resultaria verdadeira e falsa, já que o ponto fixo estende as interpretações anteriores.

Na construção dos pontos fixos, o que vamos observar é o seguinte: um ponto fixo estende as interpretações anteriores e tem como elementos todas as sentenças que são verdadeiras ou falsas na linguagem. Mas podemos construir diferentes pontos fixos dependendo da interpretação que fornecemos para  $\text{Tr}(x)$  no nível mais básico (no prefácio do livro). O ponto fixo *minimal* começa com a extensão e a antiextensão de  $\text{Tr}(x)$  vazias. Este ponto fixo é *minimal* porque ele é estendido por todos os outros pontos fixos, é um subconjunto de todos os outros pontos fixos. Começamos com uma linguagem básica  $\mathcal{L}_0$  (o prefácio), cuja interpretação do predicado  $\text{Tr}(x)$  é o par  $\langle \emptyset, \emptyset \rangle$ . De  $\varphi(\langle \emptyset, \emptyset \rangle)$  obtemos um conjunto de sentenças verdadeiras e um conjunto de sentenças falsas  $\langle S_1, S_2 \rangle$ , e então temos uma linguagem  $\mathcal{L}_1$  cuja interpretação de  $\text{Tr}(x)$  é  $\langle S_1, S_2 \rangle$ . De  $\varphi(\langle S_1, S_2 \rangle)$  obtemos um novo par  $\langle S_1', S_2' \rangle$ , que, por sua vez, é a

<sup>110</sup> Este símbolo normalmente expressa a relação “maior ou igual” entre números. Mas Kripke utiliza o símbolo para significar a relação “estender uma interpretação” apenas quando o insere entre pares ordenados do tipo  $\langle S_1, S_2 \rangle$ .

<sup>111</sup> Uma função monotônica é uma função que preserva a ordem. Se  $x < y$ , então  $f(x) < f(y)$ .

<sup>112</sup> Hartry Field denomina esta propriedade como *Roach Motel Property* (cf. Hartry Field, 2008, p. 61). A idéia é que no processo de construção das interpretações de  $\text{Tr}(x)$ , as sentenças podem entrar, mas não podem sair.

interpretação de  $\text{Tr}(x)$  em  $\mathcal{L}_2$ . Este processo continua, até que para um determinado par  $\langle S_{1,\sigma}, S_{2,\sigma} \rangle$  em uma determinada linguagem  $\mathcal{L}_\sigma$ ,  $\varphi(\langle S_{1,\sigma}, S_{2,\sigma} \rangle) = \langle S_{1,\sigma}, S_{2,\sigma} \rangle$ . Em  $\mathcal{L}_\sigma$ , temos o ponto fixo *minimal*. Este ponto fixo é como aquele que apresentamos na subseção anterior. Nele as sentenças *infundadas* não possuem valor de verdade.

Os pontos fixos *maximais* são aqueles que obtemos quando começamos atribuindo o maior número possível de sentenças para a interpretação de  $\text{Tr}(x)$  no nível básico. Sentenças como “esta própria sentença é verdadeira” (**CV**) podem ser atribuídas consistentemente à extensão ou à antiextensão de  $\text{Tr}(x)$  (não deriva contradição supor que ela é verdadeira ou que ela é falsa), mas esta atribuição é arbitrária. Estes pontos fixos são *maximais* porque não há um ponto fixo que estenda estes pontos fixos e do qual eles sejam subconjuntos próprios, ainda que não haja um único ponto fixo *maximal*, pois cada um deles envolve uma escolha arbitrária no nível básico. Há, por exemplo, um ponto fixo *maximal* em que a sentença “esta própria sentença é verdadeira” é verdadeira e outro em que ela é falsa. Na construção do ponto fixo *maximal*, começamos com uma interpretação de  $\text{Tr}(x)$  em  $\mathcal{L}_0$  que é diferente de  $\langle \emptyset, \emptyset \rangle$ , colocando, por exemplo, sentenças *infundadas*, que não resultam em contradição, na extensão ou na antiextensão de  $\text{Tr}(x)$  arbitrariamente. Assim, quando obtemos o ponto fixo *maximal*, atribuímos valor a mais sentenças do que no ponto fixo *minimal*, este ponto fixo estende o ponto fixo *minimal*. Mas se  $\langle S_1, S_2 \rangle$  é um ponto fixo *maximal*, então não existe uma interpretação de  $\text{Tr}(x)$ ,  $\langle S_1^*, S_2^* \rangle$ , diferente de  $\langle S_1, S_2 \rangle$ , tal que  $\langle S_1^*, S_2^* \rangle$  estende  $\langle S_1, S_2 \rangle$ , e  $\langle S_1^*, S_2^* \rangle$  é um ponto fixo.

O ponto fixo *intrínseco* é o ponto fixo que inclui na interpretação de  $\text{Tr}(x)$  todas as sentenças que têm sempre o mesmo valor de verdade em todos os pontos fixos que elas têm algum valor de verdade. Diferente do ponto fixo *maximal* que atribui valor a todas as sentenças *infundadas* que não resultam em contradições, este ponto fixo atribui valor a algumas sentenças *infundadas*, aquelas que parecem possuir um valor intrínseco quando possuem algum valor. Este ponto fixo não inclui a sentença “esta própria sentença é verdadeira”, mas inclui sentenças como “esta própria sentença é verdadeira ou falsa”. O único valor que podemos atribuir a ela (consistentemente) em todas as interpretações que ela tem um valor é “verdadeira”. Ambas estas sentenças são *infundadas* (tanto **CV** quanto “esta própria sentença é verdadeira ou falsa”), logo, elas resultam indefinidas no ponto fixo *minimal*, mas a primeira pode ser avaliada arbitrariamente como verdadeira ou como falsa em pontos fixos *maximais*, enquanto a

segunda é sempre verdadeira em todo ponto fixo onde ela tem um valor, ela é verdadeira no ponto fixo *intrínseco*.

Há ainda um conjunto de sentenças importantes que devem ser observadas. Suponha uma sentença *infundada*  $p$ , mas que não leva a contradições (como “esta própria sentença é verdadeira” (**CV**)). Vimos que **CV** está na extensão de  $\text{Tr}(x)$  de alguns pontos fixos *maximais* e na antiextensão de  $\text{Tr}(x)$  de outros pontos fixos *maximais*. Mas **CV** não está nem na extensão nem na antiextensão de  $\text{Tr}(x)$  no ponto fixo *minimal*, pois ela é *infundada*. Agora, suponha a sentença  $\langle \text{CV} \vee \sim \text{CV} \rangle$ . Pelas regras de Kleene, ela também é indefinida no ponto fixo *minimal*. Mas ela é sempre verdadeira em todos os pontos fixos *maximais*, já que sempre que ela tem um valor, ela é sempre verdadeira. Se **CV** é verdadeira em um ponto fixo *maximal*,  $\langle \text{CV} \vee \sim \text{CV} \rangle$  é verdadeira, se **CV** é falsa em um ponto de fixo *maximal*,  $\langle \text{CV} \vee \sim \text{CV} \rangle$  é igualmente verdadeira. Para qualquer atribuição arbitrária de valores que escolhermos para **CV**,  $\langle \text{CV} \vee \sim \text{CV} \rangle$  é sempre verdadeira. Mas daí não se segue que esta sentença é verdadeira no ponto fixo *intrínseco*, esta avaliação seria não-composicional. Uma sentença na forma de uma disjunção pertence a uma interpretação sse pelo menos um dos seus disjuntos pertence a esta interpretação. Como nem **CV** nem  $\langle \sim \text{CV} \rangle$  pertencem à extensão ou à antiextensão do ponto fixo *intrínseco*, então  $\langle \text{CV} \vee \sim \text{CV} \rangle$  também não está no ponto fixo intrínseco<sup>113</sup>.

Tendo definido os diferentes pontos fixos, Kripke pode formular uma caracterização bastante precisa das sentenças patológicas. Uma sentença é *infundada* sse ela não possui valor de verdade no ponto fixo *minimal*. A sentença “esta própria sentença é verdadeira”, a sentença “a sentença **CV** é verdadeira ou é falsa” são sentenças *infundadas*. Mas, uma sentença é paradoxal sse ela não possui valor de verdade em nenhum ponto fixo. Elas não possuem valor no ponto fixo *minimal* porque elas também são *infundadas*, mas mesmo nos pontos fixos em que a interpretação de  $\text{Tr}(x)$  é expandida ao máximo, elas não possuem valor, ou nestes pontos fixos haveria uma valoração das sentenças da linguagem que levaria à inconsistência. Tais sentenças não permitem uma atribuição de valores de verdade. Dadas as regras de Kleene mais os princípios que regem o predicado de verdade (a versão de Kripke para o esquema (**T**)), se nós atribuirmos um valor para **PM** em alguma interpretação, então nós também

---

<sup>113</sup> É interessante notar que se nós seguíssemos uma lógica não-composicional (como a *superavaliação* de van Fraassen) aqui, poderíamos inserir esta sentença no ponto fixo *intrínseco*.

atribuímos o valor contraditório para **PM** na interpretação seguinte, de tal maneira que em qualquer ponto fixo em que **PM** tem um valor, **PM** e sua negação pertencem à classe das sentenças verdadeiras. Outro exemplo de sentença paradoxal é: “se esta sentença é verdadeira, então  $2 + 2 = 5$ ”. Se ela é verdadeira em algum ponto fixo, então “ $2 + 2 = 5$ ” é verdadeira neste ponto fixo. Mas “ $2 + 2 = 5$ ” é falsa em todo ponto fixo, é uma sentença *fundada* e falsa da linguagem, segundo o primeiro *desiderata* que enunciamos. Por outro lado, se a sentença condicional resulta falsa de alguma interpretação de  $\text{Tr}(x)$ , então, como ela é um condicional, pelas regras de Kleene, o antecedente é verdadeiro, portanto, ela é verdadeira. Assim, no ponto fixo, ela também é verdadeira se ela é falsa. Sentenças paradoxais são indefinidas e não recebem valor de verdade em nenhum ponto fixo.

Mostramos que para uma linguagem em que interpretamos  $\text{Tr}(x)$ , há um valor para  $x$  que espelha o fechamento semântico. Nesta linguagem, temos um predicado de verdade que é, ao mesmo tempo, o predicado de verdade da linguagem ela mesma. Kripke também apresenta uma espécie de hierarquia. Como vimos, uma interpretação de  $\text{Tr}(x)$  em  $\mathcal{L}_0$  produz um novo par  $\langle S_1, S_2 \rangle$ , que é a interpretação de  $\text{Tr}(x)$  em  $\mathcal{L}_1$ . Assim, a linguagem  $\mathcal{L}_1$  é a linguagem que expressa verdade para  $\mathcal{L}_0$ ,  $\mathcal{L}_2$  a linguagem que expressa verdade para  $\mathcal{L}_1$ , e assim por diante. Mas a hierarquia de Kripke termina. Em algum nível, a interpretação do predicado de verdade é única, e então, não há mais hierarquia. As interpretações relevantes para a teoria de Kripke são dadas por estes pontos fixos, onde a hierarquia é eliminada, onde a linguagem é semanticamente fechada. Aliás, a hierarquia de Kripke pode ser estendida a níveis não finitos, introduzindo níveis limites. Ou seja, se tivermos uma linguagem cujo domínio torne a hierarquia um processo interminável, ainda sim, em princípio, os teoremas que provam a existência dos pontos fixos são mantidos. Suponha uma hierarquia infinita. Temos  $\mathcal{L}_0 = \mathcal{L}(\langle S_{1,0}, S_{2,0} \rangle)$ , e temos  $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3 \dots$  até que chegamos a um ordinal limite  $\omega$  (não finito).  $\mathcal{L}_\omega = \mathcal{L}(\langle S_{1,\omega}, S_{2,\omega} \rangle)$ , onde  $S_{1,\omega}$  é a união de todo  $S_{1,\alpha}$  para  $\alpha$  finito, e  $S_{2,\omega}$  é a união de todo  $S_{2,\alpha}$  para  $\alpha$  finito. Daí em diante, temos  $\mathcal{L}_{\omega+1}, \mathcal{L}_{\omega+2}, \mathcal{L}_{\omega+3}, \dots, \mathcal{L}_{\omega+\omega}$ , etc. E assim continuamos. Temos uma hierarquia que em princípio não termina. Para todo

nível limite  $\lambda$ ,  $\mathcal{L}_\lambda = \mathcal{L}(\langle S_{1,\lambda}, S_{2,\lambda} \rangle)$ , onde  $S_{1,\lambda} = \bigcup_{\alpha < \lambda} S_{1,\alpha}$ , e  $S_{2,\lambda} = \bigcup_{\alpha < \lambda} S_{2,\alpha}$ <sup>114</sup>. Ou seja, podemos, em princípio, ter uma hierarquia que não termina. Existirão níveis limites que constituem a união da interpretação de  $\text{Tr}(x)$  para todo nível finito menor que eles. E o ponto é que, mesmo para estas linguagens, ainda é verdadeiro que há valores para a função  $\varphi(x)$  que tornam  $\varphi(x) = x$ .

O resultado relevante de Kripke é que mesmo para linguagens com um domínio que implicaria uma hierarquia não finita, sempre há valores para  $x$  que formam pontos fixos na função  $\varphi(x)$ . Ou seja, mesmo para estas linguagens, o processo estabiliza, de modo que há um par  $\langle S_1, S_2 \rangle$ , que quando introduzido como interpretação de  $\text{Tr}(x)$ , fornece a interpretação de verdade para a linguagem de maneira unívoca, é o predicado de verdade da linguagem ela mesma. Diferentemente das hierarquias de linguagem de Tarski, portanto, a hierarquia de Kripke constrói a interpretação de um predicado de verdade que no ponto fixo é um único predicado para toda a linguagem, e não vários predicados relativos a cada linguagem. Na ilustração que mostramos na subseção anterior, parecia intuitivo supor que o processo estabilizava, mas para linguagens com um domínio não finito esta suposição pode, em princípio, não ser tão clara. O resultado relevante de Kripke, por isto mesmo, é tão elucidativo. Mesmo para linguagens não finitas, é certo que há valores que produzem pontos fixos na função que dá a interpretação da linguagem, e nestes pontos fixos, podemos classificar as sentenças patológicas da maneiras que fizemos aqui, temos uma semântica fechada, e podemos evitar as consequências indesejáveis previstas pelo diagnóstico de Tarski.

### 3.2.3 Resultados da solução de Kripke

Vamos retomar agora os resultados da avaliação de Kripke para as sentenças patológicas, a partir de suas considerações sobre o funcionamento dos predicados de verdade da linguagem, descritos pelos modelos de pontos fixos. Vimos na primeira seção deste capítulo que há duas maneiras de interpretar as ocorrências de *truth value gaps* sem que isto implique em contradição. A primeira delas, consiste em distinguir os predicados “é falso” e “não é verdadeiro”, o que também envolve distinguir **PB** e **PTE**.

---

<sup>114</sup> Para uma linguagem limite  $\lambda$ , temos a extensão  $\lambda$  de  $\text{Tr}(x)$ , que é a união de toda extensão de  $\text{Tr}(x)$  finita  $\alpha$  menor que  $\lambda$ , e a antiextensão  $\lambda$  de  $\text{Tr}(x)$ , que é a união de toda antiextensão de  $\text{Tr}(x)$  finita  $\alpha$  menor que  $\lambda$ .

Esta não é a opção de Kripke. A opção de Kripke é entender *truth value gaps* como uma restrição à lógica clássica em um sentido bastante específico.

Das sentenças patológicas que classificamos, algumas não possuem valor no ponto fixo *minimal*, outras não possuem valor em nenhum ponto fixo. Quando uma sentença não possui valor, convencionamos expressar tal coisa dizendo que esta sentença é indefinida. Mas diferentemente da interpretação usual de *gaps*, não afirmamos que tais sentenças são nem verdadeiras nem falsas, ou que elas são não verdadeiras. Pelas regras de Kleene mais as regras para  $\text{Tr}(x)$ , quando a sentença **CV** é indefinida,  $\langle \sim \text{CV} \rangle$  também é indefinida,  $\langle \text{Tr}(\langle \text{CV} \rangle) \rangle$  é indefinida,  $\langle \sim \text{Tr}(\langle \text{CV} \rangle) \rangle$  é indefinida, e  $\langle \sim \text{Tr}(\langle \sim \text{CV} \rangle) \rangle$  (que é equivalente a  $\langle \sim \text{Fs}(\langle \text{CV} \rangle) \rangle$ ) também é indefinida. Ou seja, se uma sentença é indefinida, é indefinido que ela é verdadeira, é indefinido que ela é falsa, e é indefinido que ela não é verdadeira nem falsa ( $\langle \sim \text{Tr}(\langle \text{CV} \rangle) \wedge \sim \text{Fs}(\langle \text{CV} \rangle) \rangle$  é indefinida). Deste modo, quando dizemos que uma sentença é indefinida, na interpretação de Kripke, não estamos em condições de asserir ou rejeitar a sentença ela mesma, ou que ela é verdadeira, e também não estamos autorizados a asserir que ela é não-verdadeira. Nós dizemos que ela é indefinida no sentido de que a função do predicado verdade não é definida para ela.

Assim, a restrição que é implicada pelos *gaps* de Kripke não é uma negação do **PB**. Não estamos autorizados a afirmar a verdade da negação da bivalência. Mas nem por isto somos obrigados a afirmar a verdade da bivalência. Se o **PB** afirma que toda sentença da linguagem é sempre verdadeira ou falsa, e se nossa linguagem tem como elementos do domínio sentenças patológicas (paradoxais e *infundadas*), então o que a restrição de Kripke pode dizer é que o **PB** ele mesmo é expresso por uma sentença indefinida. Não há nenhuma avaliação em nenhum ponto fixo que torna  $\langle \text{Tr}(x) \vee \text{Fs}(x) \rangle$  falsa para algum  $x$ , mas algumas avaliações de algumas sentenças tornam instâncias desta função indefinidas. Por isto, a quantificação universal desta função é uma sentença indefinida. Este sentido de restrição não acarreta nenhuma contradição, pois não estamos dizendo que **PB** é falso. É interessante observar que se não houve sentenças paradoxais no nosso domínio (apenas sentenças *fundadas* e *infundadas*), então poderíamos tomar o **PB** como verdadeiro nos pontos fixos *maximais*. As mesmas considerações aqui valem para o **PTE**<sup>115</sup>.

---

<sup>115</sup> Aqui estamos falando do princípio lógico, pois como já dissemos, na abordagem de Kripke “falsa” é equivalente a “não verdadeira”, e portanto, **PB** e **PTE** semântico são o mesmo princípio.

O esquema (**T**) também é restrito na solução de Kripke, embora isto também não implique negar o esquema, ou tampouco perder totalmente a intuição capturada por ele. Obviamente, se uma sentença  $p$  é indefinida, então  $\langle \text{Tr}(\langle p \rangle) \leftrightarrow p \rangle$  também é indefinida pelas regras de Kleene. Mas, para toda sentença que possui um valor de verdade, o esquema é mantido. Podemos asserir  $p$  sse podemos asserir  $\langle \text{Tr}(\langle p \rangle) \rangle$ , e podemos rejeitar  $p$  sse podemos rejeitar  $\langle \text{Tr}(\langle p \rangle) \rangle$ . Aliás, podemos recuperar parcialmente tanto o **PB** quanto o esquema (**T**) em um enunciado que será sempre verdadeiro em qualquer interpretação, desde que tornemos explícita a condição da nossa restrição: para toda sentença da linguagem, se ela possui um valor de verdade no ponto fixo *minimal*, então (**T**) e **PB** são sempre verdadeiras. As sentenças da nossa linguagem que são patológicas tornam o antecedente do condicional anterior falso, portanto, ele é verdadeiro. Para todas as outras sentenças, (**T**) e **PB** são sempre verdadeiras.

Além da garantia de que estes resultados não implicam uma revisão da lógica clássica grave o suficiente para ameaçar sua consistência, devemos também levar em conta a fecundidade das considerações de Kripke enquanto solução para os paradoxos. A solução de Kripke permite manter uma distinção fina e precisa de patologias distintas, como a das sentenças *infundadas* e a das sentenças paradoxais. A solução também permite uma aplicabilidade parcialmente abrangente, já que se aplica aos predicados de verdade de maneira geral, por identificar “é falso” com “não é verdadeiro”<sup>116</sup>. Mas a solução também permite separar as ocorrências inócuas daquelas sentenças que podem resultar em patologias. É o caso do paradoxo que mostramos a respeito de Nixon e Jones. Jones diz: “Todas as declarações de Nixon sobre *Watergate* são falsas”, e Nixon diz: “Todas as declarações de Jones sobre *Watergate* são verdadeiras”. Pela Solução Ortodoxa, estas sentenças sequer possuem uma atribuição de níveis legítima, e como os níveis devem ser prefixados às sentenças, não podemos nos sensibilizar para as ocorrências inócuas destas sentenças, como dissemos antes. Quando, por exemplo, Nixon tivesse enunciado alguma outra sentença sobre o *Watergate*, e esta sentença fosse verdadeira, então as duas sentenças acima seriam falsas (e *fundadas*) e não haveria

---

<sup>116</sup> Por identificar “é falso” com “não é verdadeiro”, a solução de Kripke sequer está sujeita às conhecidas versões reforçadas do Mentiroso como **PMR**, e se aplica a todas as versões conhecidas de Paradoxo do Mentiroso envolvendo predicados de verdade. Um ponto interessante a ser investigado é se, tal como a Solução Ortodoxa, baseada nos resultados de Tarski, também se pode generalizar os resultados de Kripke para outros conceitos semânticos. Se isto é o caso, então há pontos fixos também para as funções que dão a interpretação destes conceitos na linguagem, podemos construir uma linguagem semanticamente fechada para eles, e paradoxos não ocorrem, porque estes predicados são também parciais. Neste caso, poderíamos dizer que a solução de Kripke é significativamente abrangente. Mas este ponto é apenas especulativo.

nenhum paradoxo. A sentença de Jones é uma quantificação universal, e como tal, é verdadeira sse tudo que Nixon diz sobre *Watergate* é falso, é falsa sse pelo menos uma das coisas que Nixon diz sobre *Watergate* é verdadeira, e é indefinida de outro modo. Portanto, se supusermos que Nixon disse pelo menos uma sentença verdadeira sobre *Watergate*, então, a sentença de Jones é falsa, e assim, a sentença de Nixon “Todas as declarações de Jones sobre *Watergate* são verdadeiras” é falsa. Embora, as duas sentenças envolvam o predicado de verdade de uma maneira aparentemente viciosa, há circunstâncias sob as quais as regras de interpretação dos quantificadores que usamos permitem uma atribuição de valores para as sentenças, e elas não resultam mais indefinidas.

### 3.3 Vantagens e Desvantagens

Neste capítulo introduzimos a Solução Alternativa como uma revisão da lógica clássica que permite preservar o fechamento semântico da linguagem sem que isto implique inconsistência, evitando os paradoxos que temos discutido. As soluções alternativas compartilham o diagnóstico de Tarski. Vimos também que há mais de uma maneira de interpretar estas revisões ou restrições da lógica clássica. Mencionamos na primeira seção as interpretações usuais de *truth value gaps* e os problemas que podem ser apresentados para elas. A seguir, consideraremos alguns pontos da Solução Alternativa, as suas vantagens e desvantagens. Para isto vamos nos concentrar principalmente na solução de Kripke.

Na medida em que a Solução Alternativa parte do mesmo diagnóstico que a Solução Ortodoxa, e nós consideramos este um bom diagnóstico, devemos assumir que a Solução Alternativa também possui um bom ponto de partida, um bom diagnóstico. A diferença é que a Solução Ortodoxa parte deste diagnóstico para a imposição de uma determinação no uso dos conceitos semânticos (sobretudo o conceito de verdade), e a Solução Alternativa não impõe uma determinação no uso do conceito, mas reúne justificativas e indumentárias semânticas que permitam tratar das linguagens sem impor as determinações da Solução Ortodoxa. Enquanto o paciente sujeito ao tratamento ortodoxo deve obedecer a um rigoroso conjunto de proibições e mudar radicalmente a sua rotina para sempre, o paciente sujeito ao tratamento alternativo vai poder manter sua rotina comum, desde que passe por uma cirurgia delicada que irá desligar alguns mecanismos e reconstruí-los de uma maneira parcial, mas sem a garantia de que estes

mecanismos não serão, de alguma forma, modificados. A Solução Ortodoxa parte do diagnóstico de Tarski para impor regras de uso dos predicados semânticos. A Solução Alternativa parte do diagnóstico de Tarski e da necessidade de não impor determinações aos usos dos predicados de verdade para restringir a lógica clássica sem que isto implique em contradições.

Além de partir de um bom diagnóstico, a Solução Alternativa está também apoiada em boas razões filosóficas e formais. Como vimos, Kripke apresenta razões bastante convincentes para a admissão de *gaps* na linguagem, além de uma fina distinção de fenômenos específicos que motivam esta admissão. Os *gaps* já são justificados na linguagem natural por uma série de fenômenos independentes das considerações a respeito do Paradoxo do Mentiroso (como as falhas de referência e de pressuposição), mas na abordagem de Kripke eles são ainda justificados por uma razão adicional que abrange os fenômenos das sentenças *infundadas* e paradoxais. Além disto, a solução de Kripke também suporta os casos de *gaps* com considerações formais precisas e fecundas, tornando a caracterização das sentenças um resultado efetivo.

A Solução Alternativa cumpre o requisito da *especificidade*, já que a restrição ao **PB** elimina uma premissa necessária do paradoxo. Ela desautoriza a afirmação de que **PM** é verdadeira ou falsa (desautoriza também, no caso de Kripke, a afirmação de que **PM** é verdadeira ou não é verdadeira). Quando uma sentença é indefinida, não podemos asserir que ela é verdadeira ou falsa, e também não podemos asserir ou rejeitar a sentença ela mesma. Logo, a existência destas sentenças no domínio da nossa linguagem não leva a contradições. Nesta abordagem, **PB** e **(T)** não são mais tautologias, e a derivação dos paradoxos está bloqueada.

A Solução Alternativa também cumpre (embora, em princípio, apenas parcialmente) o requisito da *completude*. Na medida em que identifica “é falso” com “não é verdadeiro”, a Solução Alternativa (Kripke) permite construir uma explicação que vale sempre tanto para as versões mais simples do Paradoxo do Mentiroso quanto para as versões reforçadas. Na realidade, por esta razão, é mais significativo dizer que as versões reforçadas sequer são colocadas para esta abordagem, já que a versão reforçada do Mentiroso (**PMR**) tem o mesmo significado que a versão simples (**PM**). E de uma maneira geral, na medida em que a Solução Alternativa, restringindo a lógica clássica, pode preservar o fechamento semântico da linguagem, então, ela espelha os traços característicos da linguagem natural. Diferentemente da posição de Tarski, aqui podemos manter uma expectativa em relação à linguagem natural. A Solução

Alternativa vale tanto para linguagens formais quanto para linguagens que no mínimo se assemelham essencialmente com a linguagem natural. Neste sentido, a solução é bastante abrangente<sup>117</sup>.

A Solução Alternativa não está fundada em *postulações ad hoc*. A admissão de *truth value gaps* é estabelecida por boas razões. Não é o caso de avaliar algumas sentenças como indefinidas pela simples razão de que de outro modo elas produziriam inconsistências na linguagem. A razão para atribuir-lhes ausência de valores é independentemente motivada. No caso de Kripke isto fica ainda mais claro, porque nós podemos distinguir sentenças *infundadas* de sentenças paradoxais, e as primeiras não produzem modelos inconsistentes quando avaliadas como verdadeiras ou falsas, mas são igualmente indefinidas no ponto fixo *minimal*.

Na medida em que não proíbe a autorreferência, a Solução Alternativa, também não compartilha com a Solução Ortodoxa a característica de produzir *estragos excessivos*. Não há qualquer imposição contra círculos de uma maneira geral. A idéia, pelo contrário, é permitir círculos como um traço característico da linguagem, e deixar que as circunstâncias determinem se a atribuição de valores no processo resulta em sentenças patológicas de algum tipo ou não. Como vimos, no caso “Nixon e Jones”, algumas sentenças que seriam precipitadamente desautorizadas pela hierarquia ortodoxa resultam em sentenças absolutamente inofensivas quando avaliadas pelos modelos de Kripke sob certas circunstâncias.

Por último, e talvez a mais importante vantagem da Solução Alternativa, é que ela *conserva intuições*, ou pelo menos conserva uma intuição fundamental que é abandonada pela Solução Ortodoxa, o fechamento semântico da linguagem. Na Solução Ortodoxa somos levados a uma hierarquia de linguagens e a uma hierarquia de predicados de verdade, de modo que não temos um predicado de verdade que é único para toda a linguagem, e nenhuma linguagem possui o seu próprio predicado de verdade. Nos modelos de Kripke, em qualquer ponto fixo, temos um valor para o par <extensão, antiextensão> do predicado verdade que coincide com as sentenças verdadeiras e falsas da linguagem, ela mesma. Esta linguagem possui um predicado de verdade que é o único da linguagem. Ela não apenas preserva uma característica

---

<sup>117</sup> Embora, como já dissemos, resta ver se os resultados de Kripke podem ser generalizados para os outros conceitos semânticos. Em princípio, o que temos são considerações particularmente direcionadas para os predicados de verdade.

essencial da linguagem natural, como também espelha uma intuição comum, que é a de que só há um predicado de verdade e não vários.

Entre as desvantagens que podemos levantar contra a Solução Alternativa, mais especificamente contra Kripke, a mais importante é levantada pelo próprio Kripke, e ficou também conhecida como Problema da Vingança. Deixaremos esta para considerarmos por último. As outras desvantagens tornam-se menores tanto pelas vantagens da solução de Kripke quanto pela relevância do Problema da Vingança. Não queremos dizer com isto que estas desvantagens menores não levantam sérios problemas, mas que, tendo em vista o problema levantado pela Vingança, e mesmo a fecundidade da teoria em outros aspectos, elas acabam sendo esquecidas pela literatura. Vamos apenas mencionar rapidamente estas desvantagens menores.

Em primeiro lugar, podemos dizer que não é claro que a solução de Kripke seja totalmente abrangente. Como dissemos, não há uma resposta clara para a questão de se os resultados podem ser generalizados para outros conceitos semânticos, e assim, resolver paradoxos semelhantes que não envolvem predicados de verdade. Ou seja, na Solução Ortodoxa, paradoxos de círculos viciosos envolvendo outros conceitos semânticos podem ser eliminados pela hierarquia de linguagens. Esta hierarquia não permite que os conceitos sejam utilizados de maneira viciosa. Não podemos, todavia, defender uma linguagem semanticamente fechada que eliminasse a hierarquia para estes conceitos, até aqui. Não sabemos se há um ponto fixo na função que estabelece a interpretação da linguagem para estes conceitos. Logo, não sabemos se a solução de Kripke pode ser estendida para outros paradoxos semânticos envolvendo círculos viciosos. De maneira geral, a solução obtida pela admissão de *truth value gaps* é diretamente relacionada apenas com patologias da atribuição de predicados de verdade. Ao menos neste sentido, a Solução Ortodoxa é, em princípio, mais abrangente.

Podemos mencionar ainda um ponto que fica mais ou menos impreciso na solução de Kripke. Ele não fornece uma teoria do significado. É certo que ele não oferece sequer uma teoria sistematizada e formalizada para os predicados de verdade, mas apenas conjuntos de considerações que esboçam uma teoria. As considerações que ele apresenta para a noção de significado, todavia, são ainda menos satisfatórias. Não fica claro o que Kripke entende por significado, e por quais razões melhores explicitadas ele distingue “falha de significado” de “falha em expressar uma proposição”. Ele apenas afirma esta distinção. Para outras formulações da Solução Alternativa, um problema semelhante se coloca. Nas outras abordagens (Martin e van Fraassen), é comum

classificar estes paradoxos do tipo Mentiroso como falhas de significado que geram sentenças nem verdadeiras nem falsas. Mas tanto do ponto de vista formal quanto do ponto de vista filosófico, não é possível classificar a patologia que ocorre em **PM** juntamente com os outros tipos de patologias do significado. Do ponto de vista formal, como vimos, **PMR** não envolve a suposição do **PB**, mas do **PTE**. E assim, sentenças que envolvem outros tipos de falhas de significado podem ser classificadas como “não verdadeiras”, mas **PMR**, que é essencialmente semelhante a **PM**, não pode. Do ponto de vista filosófico, não há razões para classificar **PM** ou **PMR** como patologias do mesmo tipo das falhas de significado mais comuns, elas não são casos em que há falhas de referência, nem casos de falhas pressuposicionais ou erros categoriais. Em suma, na falta de uma teoria do significado, as soluções alternativas encontram dificuldades em caracterizar o tipo de falha semântica ocorrente em sentenças que produzem paradoxos do tipo Mentiroso.

Além destas desvantagens, há também o desconforto inerente às tentativas de restringir a lógica clássica. Já mencionamos os problemas decorrentes para as interpretações de *truth value gaps* que entendem tal restrição como negação de certos princípios. Para as soluções alternativas usuais, várias distinções pouco fecundas deverão ser formuladas. Mas Kripke não defende estas distinções (entre **PB** e **PTE**, por exemplo). Para ele, restringir o **PB** não implica negar o **PTE** ou afirmar a verdade da negação do **PTE**. Estes princípios irão resultar em sentenças indefinidas no ponto fixo *minimal*. Mas isto não parece confortável o suficiente. Não basta que a solução evite negar tautologias da lógica clássica. Parece haver um esforço no sentido de preservar como verdadeiras o máximo delas. Aqui vemos, certamente, o desconforto da solução de Kripke. Se não houver sentenças paradoxais no domínio, o **PTE** é sempre verdadeiro em todos os pontos fixos *maximais*. E se assumirmos uma semântica não composicional, **PTE** é verdadeiro no ponto fixo *intrínseco* (o que ocorre se utilizarmos uma lógica do tipo *superavaliação*). Vemos que Kripke tenta de algum modo recuperar os princípios fundamentais da lógica, mas de maneira insuficiente, pois no domínio relevante para esta discussão há sentenças paradoxais, e assim, o **PB** resulta indefinido nos pontos fixos *maximais*, e também resulta indefinido no ponto fixo *intrínseco*, a menos que adotemos uma semântica não-composicional. A solução de Kripke acaba restringindo **PB**, **PTE** e **(T)**, que são princípios que espelham coisas muito intuitivas. Vale sempre lembrar que pelo menos não somos obrigados a afirmar que eles sejam falsos.

O Problema da Vingança é um resultado da própria semântica de Kripke que pode ser considerado a mais séria objeção à sua teoria. O que acontece é que quando caracterizamos sentenças como **PM** ou **CV** como indefinidas, significando com isto que elas não possuem valor de verdade, em certo sentido, estamos afirmando que elas não são verdadeiras. Ora, se elas não possuem valor de verdade, elas não são verdadeiras. Mas, quando afirmamos que **PM** não é verdadeira, estamos afirmando **PM** ela mesma, porque **PM** = “esta própria sentença é falsa (o que é o mesmo que ‘não é verdadeira’)”. Queremos poder afirmar que **PM** não é verdadeira, mas se o fizermos, afirmamos uma sentença que resulta indefinida pela nossa própria semântica. O ponto é: não podemos asserir as sentenças que expressam a solução de Kripke, não podemos dizer com sentenças verdadeiras que as sentenças patológicas não são verdadeiras. O Problema da Vingança reflete a limitação de expressividade da teoria de Kripke. Convencionamos que os *gaps* em Kripke tinham um sentido de ausência de valor no sentido mais forte. Quando dizemos que uma sentença é indefinida, queremos dizer com isto que não podemos atribuir ou negar dela qualquer predicado de verdade. Mas a teoria precisa poder expressar com sentenças verdadeiras que as sentenças patológicas não possuem valor. E pela regra do que podemos afirmar como verdadeiro nesta teoria, tais sentenças não estão autorizadas. Logo, não podemos expressar a própria teoria.

Suponha que nestes casos queremos dizer que tais sentenças são indefinidas, introduzindo “indefinida” como um predicado da linguagem. Poderíamos então, formular o seguinte paradoxo: “esta própria sentença é falsa ou indefinida”. Se ela não possui valor no sentido forte, então, ela é indefinida, mas isso é parte do que ela diz de si mesma, logo, ela é verdadeira. Por outro lado, se ela é verdadeira, ela é falsa ou não possui valor, e se ela é falsa, ela é verdadeira. O problema aqui é que se nós introduzirmos “indefinida” na linguagem, queremos com isto tornar verdadeiras sentenças que atribuem este predicado a sentenças realmente indefinidas, queremos tornar verdadeiras sentenças como “**PM** é indefinida”, mas se o fizermos, abrimos o caminho para novos paradoxos. Por outro lado, se introduzirmos “indefinida” de tal maneira que “**PM** é indefinida” resulte também em uma sentença indefinida, não podemos mais expressar nossa teoria com sentenças verdadeiras.

Não podemos dizer com uma sentença que resulte verdadeira na nossa linguagem que **PM** não é verdadeira em nenhum ponto fixo da interpretação de  $Tr(x)$ . E também não podemos introduzir um predicado na linguagem que expresse com sentenças verdadeiras que há sentenças que não possuem valor de verdade no sentido mais forte.

Algumas soluções foram levantadas para este problema, mas nenhuma delas parece satisfatória. Segundo Kripke:

(...) há asserções que nós podemos fazer sobre a linguagem objeto que nós não podemos fazer na linguagem objeto. Por exemplo, sentenças do tipo Mentiroso não são verdadeiras na linguagem objeto, no sentido de que o processo indutivo nunca as torna verdadeiras; mas nós estamos impedidos de dizer isto na linguagem objeto pela nossa interpretação da negação e do predicado de verdade (...) A necessidade de ascender a uma metalinguagem pode ser uma das fraquezas da presente teoria. O fantasma da hierarquia de Tarski ainda está conosco. (Kripke, 1975, pp.79-80).<sup>118</sup>

Kripke sugere introduzir um novo predicado de verdade da metalinguagem, para o qual toda sentença da linguagem que não está na sua extensão está na sua antiextensão. As sentenças patológicas da linguagem seriam não verdadeiras no sentido de não estarem na extensão deste predicado da metalinguagem. Neste caso, poderíamos formular também na metalinguagem as semânticas dos conceitos de paradoxo, sentenças *fundadas* e *infundadas*. O problema óbvio aqui é que, em última instância, Kripke não criou uma alternativa genuína à hierarquia de linguagens da Solução Ortodoxa.

Outros autores (como Parsons, 1984, e Maudlin, 2004) tentaram fornecer meios para resolver este impasse a partir de considerações a respeito de regras de asserções. A idéia geral em ambos os autores, é que nós devemos conceber algo como um ato ilocucionário alternativo a asserir, um ato de asserção contrário a asserir, algo como rejeitar a sentença. Asserir uma sentença é afirmar a verdade da sentença. Mas rejeitar uma sentença (nestas teorias) não é afirmar a verdade da sua negação, e sim, recusar tanto a verdade da sentença quanto da sua negação. Esse sentido de rejeição poderia ser expresso na própria linguagem com um operador de negação forte “ $\neg$ ” que retorna uma sentença verdadeira quando aplicado a uma sentença falsa ou indefinida, e retorna uma sentença falsa quando aplicado a uma sentença verdadeira. Dessa maneira, expressaríamos que sentenças como **PM** não possuem valor de verdade simplesmente

---

<sup>118</sup> (...) *there are assertions we can make about the object language which we cannot make in the object language. For example, Liar sentences are not true in the object language, in the sense that the inductive process never makes them true; but we are precluded from saying this in the object language by our interpretation of negation and the truth predicate (...)* The necessity to ascend to a metalanguage may be one of the weakness of the present theory. The ghost of the Tarski hierarchy is still with us.

rejeitando que elas sejam verdadeiras, ou asserindo  $\langle \neg \text{Tr}(\langle \mathbf{PM} \rangle) \rangle$ . Poderíamos dizer com sentenças verdadeiras que  $\mathbf{PM}$  não é verdadeira, e inclusive; dizer com sentenças verdadeiras que  $\mathbf{PM}$  não é verdadeira nem falsa: com a sentença  $\langle \neg \text{Tr}(\langle \mathbf{PM} \rangle) \wedge \neg \text{Fs}(\langle \mathbf{PM} \rangle) \rangle$ . O problema desta solução é que, ao introduzir “ $\neg$ ” na linguagem ou construir um conjunto de regras para “rejeição”, abrimos o caminho para novas versões do Mentiroso que envolveriam “ $\neg$ ”, como “esta própria sentença não é verdadeira (no sentido forte da negação)”, ou paradoxos semelhantes envolvendo conceitos que expressem regras de asserção.

Aliás, um operador que funcionasse como “ $\neg$ ” não preservaria a monotonicidade de  $\varphi(x)$ , ou tornaria a relação “estender uma interpretação” uma relação que produziria pontos fixos inconsistentes. Suponha que uma interpretação  $\langle S_1, S_2 \rangle$  avalia uma determinada sentença  $\langle p \rangle$  como indefinida. Neste caso,  $\langle S_1, S_2 \rangle$  avalia  $\langle \neg p \rangle$  como verdadeira. Mas pode haver uma interpretação  $\langle S_1', S_2' \rangle$  que estende  $\langle S_1, S_2 \rangle$  e avalia  $\langle p \rangle$  como verdadeira e  $\langle \neg p \rangle$  como falsa. Se “ $\neg$ ” preserva a monotonicidade de  $\varphi(x)$ , então, no ponto fixo,  $\langle \neg p \rangle$  é verdadeira e falsa. Ainda que  $\langle S_1', S_2' \rangle$  estenda  $\langle S_1, S_2 \rangle$ ,  $\varphi(\langle S_1', S_2' \rangle)$  não estende  $\varphi(\langle S_1, S_2 \rangle)$ , e portanto, as regras estipuladas para “ $\neg$ ” não podem ser utilizadas na nossa linguagem, pois elas tornam a função  $\varphi(x)$  não-monotônica ou produzem pontos fixos inconsistentes. Como no caso anterior, podemos no máximo introduzir “ $\neg$ ” em uma metalinguagem, o que é um resultado desconfortável<sup>119</sup>.

Enfim, embora a Solução Alternativa, sobretudo a teoria de Kripke, possua vários pontos de elucidação a respeito dos predicados de verdade e permita classificar as sentenças patológicas de uma maneira bastante satisfatória, algum limite de expressividade acompanha esta solução. De alguma maneira, as consequências indesejáveis de restringir a lógica clássica parecem acompanhar as tentativas de solução alternativas, ainda que estas restrições sejam cuidadosamente minimizadas.

---

<sup>119</sup> Maudlin apresenta soluções para estes impasses, mas que envolvem considerações específicas sobre a teoria de asserções que ele apresenta. Não é o caso discuti-las aqui.

## Conclusão

Apresentamos neste trabalho uma breve apreciação das principais vias de solução para o Paradoxo do Mentiroso discutidas na literatura. As teorias, tentativas de solução, ou meras sugestões que têm sido apresentadas a respeito do Mentiroso são atravessadas pela disputa entre a Solução Ortodoxa e a Solução Alternativa. Mesmo assim, é muito difícil e, até certa medida, inapropriado reunir os textos que se encaixam nesta disputa de modo a caracterizar suas posições e avaliá-las com justiça e ao mesmo tempo fecundidade. O que acontece é que alguns dos autores estão inseridos em problemas independentes e concentrados em questões que envolvem o paradoxo por razões isoladas e não necessariamente comuns a um problema central que apareça para todos.

O Mentiroso ocorre na obra de Tarski por razões muito específicas e tangenciais em relação aos seus objetivos centrais. O problema para Tarski é o de construir uma definição materialmente adequada e formalmente correta do conceito de verdade, e tal coisa não é possível se uma linguagem pode expressar a sua própria metalinguagem. O Paradoxo do Mentiroso aparece aqui como um tipo de fenômeno que sucede ao fechamento semântico, justificando a introdução da hierarquia de linguagens. Tarski não está de maneira nenhuma interessado em caracterizar uma versão geral do Paradoxo do Mentiroso, incluindo as versões da linguagem natural. Na verdade, pelo diagnóstico de Tarski, uma linguagem que se comporte como as linguagens semanticamente fechadas de uma maneira incorrigível (o que parece ser o caso da linguagem natural) é inconsistente. Mas ele sequer estende esta conclusão para a linguagem natural. Estes resultados e estas conclusões dependem de podermos especificar com precisão a estrutura da linguagem acerca da qual estamos tratando. Para a linguagem natural estes resultados todos são meramente vagos. A preocupação de Tarski é exclusivamente direcionada às linguagens formais.

Por outro lado, é tentador explorar as consequências do diagnóstico de Tarski para a linguagem natural ou para as versões do Paradoxo do Mentiroso nas linguagens semanticamente fechadas que se assemelhem à linguagem natural. A formulação mais direta da aplicação dos resultados de Tarski para a linguagem natural, concluiria que a linguagem natural é inconsistente. Não é isto, todavia, que os autores consideram como a conclusão da Solução Ortodoxa. Considerar a linguagem natural inconsistente, além das consequências desconfortáveis e indesejáveis que esta tese envolve por ela mesma,

não é uma solução para o paradoxo. É como assumir que o problema não tem solução, como declarar ao paciente: “lamento, mas você está doente”. O diagnóstico de Tarski, portanto, tem motivado a formulação de uma abordagem para o Paradoxo do Mentiroso que Kripke chama de Solução Ortodoxa. Não há um autor que claramente defenda a Solução Ortodoxa, embora, como vimos, alguns autores defendam algo parecido (por exemplo, Charles Parsons, 1974). Por isto mesmo, os autores que atacam a Solução Ortodoxa costumam atacar a teoria de Tarski. Como vimos, isto não é correto.

A Solução Alternativa também não está isenta de disputas de interpretação. Em geral, a Solução Alternativa é caracterizada pela defesa do fechamento semântico da linguagem e pela autorização de revisões da lógica clássica. Mas também vimos que há mais de uma maneira de interpretar estas revisões da lógica clássica, principalmente maneiras diferentes de interpretar a ocorrência de *truth value gaps*. Com o objetivo de tornar claro estes pontos, introduzimos no capítulo 3 um interlúdio sobre *truth value gaps*. Mostramos nesta seção alguns problemas nas interpretações usuais de *gaps*.

Podemos dizer que, de maneira geral, a disputa entre Solução Ortodoxa e Solução Alternativa é uma disputa entre Tarski e Kripke, estabelecidas as devidas ressalvas. Tentamos reorganizar o quadro de soluções a partir destes autores e apresentar de maneira simplificada as vantagens e desvantagens de ambas as posições. Para fim de nossa conclusão, cabe recuperarmos esse emparelhamento entre as duas vias de solução, retomando resumidamente as vantagens e desvantagens de cada uma.

A Solução Ortodoxa deve ser destacada, em primeiro lugar, pela sua abrangência. Ainda que a adaptação dos resultados de Tarski para a linguagem natural seja muito questionável, a hierarquia de predicados semânticos é suficiente para eliminar todos os paradoxos semânticos que envolvam círculos viciosos. A Solução Ortodoxa não pode ser acusada de formular uma solução específica para um paradoxo, uma solução que não tem nada a dizer sobre paradoxos essencialmente semelhantes, mas que envolvem outros conceitos. Mas além de abrangente, devemos levar em consideração que a Solução Ortodoxa também é formalmente confortável, porque ela não exige qualquer revisão da lógica clássica, pelo contrário, ela preserva coisas importantes como o **PB**, o esquema (**T**) e o **PTE** da lógica.

Por outro lado, devemos destacar que a Solução Ortodoxa é, até certo ponto, muito artificial. Considerando que ela pode ser estendida à linguagem natural, a acusação de artificialidade é ainda mais legítima, a imposição de uma hierarquia de predicados na linguagem natural claramente é um recurso artificial, meramente formulado para

impedir a derivação dos paradoxos, e não elucida o funcionamento destes predicados na linguagem. Além disso, mesmo enquanto uma solução formulada apenas para linguagens puramente formais, esta solução é artificial e *ad hoc*. Se o defensor da Solução Ortodoxa argumenta que a solução é apenas um método que determina como devemos construir linguagens que possuem conceitos semânticos, então, ele supostamente estaria livre da acusação de artificialidade. Porém, se ele pretende que a sua linguagem espelhe de alguma maneira o significado real dos conceitos utilizados, ele deve justificar a hierarquia de linguagens com argumentos independentes da derivação de paradoxos. Observamos que este não é o caso.

A hierarquia de predicados imposta pela Solução Ortodoxa não é apenas artificial e *ad hoc*, mas também pode ser considerada demasiadamente abrangente, por *produzir estragos excessivos*. É como arrancar a cabeça do paciente para que ele não se perturbe mais com o seu nariz. Não podemos mais, sob hipótese alguma, expressar sentenças em que ocorrem autorreferências. Isto é ruim porque, como nós vimos, algumas sentenças que envolvem autorreferência são totalmente inofensivas, como “esta própria sentença está escrita em português”, ou pelo menos não derivam paradoxos, como “esta própria sentença é verdadeira”. Além disto, a hierarquia de predicados é pré-fixada, de modo que a Solução Ortodoxa também acaba sendo insensível às circunstâncias empíricas que podem determinar se uma sentença resulta ou não em paradoxos. É o caso que discutimos a respeito das sentenças de Nixon e Jones. Na Solução Ortodoxa, sequer temos como introduzir uma atribuição consistente de níveis para os seus predicados.

Por fim, a Solução Ortodoxa não está em harmonia com o significado intuitivo do conceito de verdade. Ela implica uma proliferação de linguagens e de predicados que se distancia muito do modo como supomos que seria uma teoria intuitiva. Uma teoria que explique o significado do conceito de verdade deveria falar de **um** predicado e não de **vários** predicados. Certamente, a linguagem natural é semanticamente fechada e possui o seu próprio predicado de verdade, de modo que, no mínimo enquanto uma solução para os paradoxos da linguagem natural, a Solução Ortodoxa pode ser considerada uma solução contraintuitiva.

As Soluções Alternativas compartilham a preservação do fechamento semântico da linguagem ao custo de revisões da lógica. Neste sentido, elas compartilham todas as vantagens decorrentes de não compartilharem das desvantagens da Solução Ortodoxa. Uma solução que se caracteriza por tratar o predicado de verdade como um predicado comum a toda a linguagem e que se aplique à linguagem ela mesma está em harmonia

com o significado intuitivo do conceito de verdade. Além disso, essa solução não leva a estragos excessivos e não precisa impor uma hierarquia artificial e *ad hoc*.

A solução de Kripke possui uma sofisticada caracterização dos fenômenos de dependência semântica dos predicados de verdade, elucidando o funcionamento destes predicados e tornando a ocorrência de *truth value gaps* satisfatoriamente justificada. A caracterização dos tipos de patologias que envolvem os predicados de verdade é fina o suficiente, não apenas para justificar os *gaps*, mas também para distinguir ocorrências inofensivas das ocorrências patológicas de autorreferências. Podemos inclusive distinguir sentenças *infundadas* de sentenças paradoxais. A solução de Kripke permite explicar o caso das sentenças de Nixon e Jones. Sob determinadas circunstâncias elas possuem valor de verdade e são *fundadas*, mas sob outras circunstâncias elas não possuem valor de verdade.

Por outro lado, as Soluções Alternativas não são confortáveis do ponto de vista formal. Como vimos, a negação do **PB** leva à afirmação da verdade de uma contradição. Aqueles que defendem que há *truth value gaps* na linguagem precisam desviar-se da acusação de inconsistência. A interpretação comum de *gaps* afirma de fato a negação do **PB**, mas ao mesmo tempo afirma a distinção entre **PB** e **PTE** (tanto o semântico quanto o lógico). Esta distinção permite expressar que sentenças com falhas de valor são não-verdadeiras e evita a acusação de inconsistência, porém alguma dificuldade em estabelecer que estes princípios sejam sempre verdadeiros persiste<sup>120</sup>. Enquanto solução para o Mentiroso, esta interpretação de *truth value gaps* é bastante insatisfatória, pois ela perde seu *status* de expressividade tão logo a sentença **PM** é avaliada como nem verdadeira nem falsa. Para aqueles que pretendem avaliar toda sentença como verdadeira ou não-verdadeira, uma versão mais forte do Mentiroso pode ser formulada, o **PMR**. Assim que **PMR** é colocada ao defensor dos *gaps* que distingue **PB** de **PTE**, ele deve de alguma maneira ceder na capacidade expressiva de sua avaliação. Se ele disser que **PMR** não é verdadeira nem falsa, temos duas situações. Na primeira, ele admite que ao afirmar “tal sentença é nem verdadeira nem falsa”, quando de fato ela é nem verdadeira nem falsa, então o que é dito é verdadeiro. Neste caso, se ela é nem verdadeira nem falsa, ela é verdadeira, já que isto é parte do que ela diz. Na segunda, ele admite que ao afirmar “tal sentença é nem verdadeira nem falsa”, quando de fato ela é nem verdadeira nem falsa, então o que é dito também é nem verdadeiro nem falso.

---

<sup>120</sup> No caso de van Fraassen isto é feito assumindo uma semântica não-composicional.

Neste caso, não pode afirmar que **PMR** não é verdadeira nem falsa. Por um lado, a sentença é avaliada de uma maneira contraditória, por outro, ela não pode ser avaliada como o defensor deste tipo de *gaps* pretende avaliá-la.

Algo bastante semelhante acontece com a interpretação dos *gaps* de Kripke. Kripke entende que uma sentença não possui valor de verdade quando ela não recebe valor em nenhum ponto fixo da função que estabelece a interpretação da linguagem. Ele não sustenta uma distinção entre **PB** e **PTE** semântico, e admite que em uma linguagem cujo domínio tenha sentenças paradoxais, o **PB** e o **PTE** da lógica também não possuem valor em nenhum ponto fixo. Isto, contudo, não significa que as respectivas negações destes princípios sejam afirmadas como verdadeiras pela teoria, elas também não possuem valor, são indefinidas. Kripke entende falha de valor no sentido mais forte. Uma versão mais forte do Mentiroso como **PMR** sequer aparece para a sua abordagem. Para ele “é falsa” é o mesmo que “não é verdadeira”, logo, **PMR** é a mesma sentença de **PM**. Ambas não possuem valor. O problema que surge nesta abordagem é um problema de expressividade semelhante àquele que ocorreu antes para a interpretação de van Fraassen. Trata-se do Problema da Vingança. Nesta abordagem, não podemos expressar com sentenças verdadeiras que **PM** não é verdadeira, ou que **PM** é indefinida. Se abrimos o caminho para introduzir um recurso que permita expressar tais sentenças verdadeiramente, abrimos também o caminho para versões reforçadas do Mentiroso.

Estas limitações de expressividade para as teorias que assumem *gaps* parecem decorrer da necessidade de uma metalinguagem clássica, na qual podemos afirmar que **PB** e **PTE** são sempre verdadeiros. O fantasma de Tarski persegue os defensores do fechamento semântico. A abertura semântica, porém, é um suicídio da Solução Alternativa. Se ela é uma opção aceitável, estamos certamente retornando ao projeto da Solução Ortodoxa.

Além de se mostrar formalmente desconfortável, a Solução Alternativa também apresenta outros pontos de desvantagem em relação à Solução Ortodoxa. A Solução Ortodoxa em princípio é mais abrangente do que a Solução Alternativa, já que a aceitação de *truth value gaps* só resolve os paradoxos que envolvem predicados de verdade<sup>121</sup>. Outro problema é que as Soluções Alternativas, em geral, precisam justificar a falha das sentenças patológicas como algum tipo de falha semântica, e não parece

---

<sup>121</sup> Claro que tanto a teoria de Kripke quanto as outras teorias que aceitam *gaps* podem desenvolver a semântica de predicados parciais para incluir outros paradoxos semânticos. Mas como vimos, na teoria de Kripke, mais especificamente, resta ver se os resultados que obtemos para o predicado de verdade analogamente podem ser obtidos para outros predicados. Sobre isto não podemos ter ainda uma garantia.

ocorrer uma falha de significado nas sentenças que envolvem patologias do tipo Mentiroso. Alguns autores tentam aproximar a falha semântica do Mentiroso com falhas de referência, falhas pressupucionais ou mesmo falhas de significado. Mas estas tentativas não são muito bem sucedidas, porque não há bons argumentos para inserir os fenômenos do tipo Mentiroso na mesma classe destas patologias. Não há uma falha de referência ou de pressuposição nas sentenças do tipo **PM**, e elas também não são sentenças sem significado ou que envolvam erros categoriais. Kripke tentou caracterizar um tipo específico de patologia para estas sentenças. Ele entende que estas sentenças (*infundadas* e paradoxais) são sentenças que não possuem valor no ponto fixo *minimal*, sentenças que apresentam círculos viciosos sob a relação de dependência semântica. Não são sentenças sem significado, mas são sentenças que não expressam proposições, não possuem condições de verdade determinadas. A abordagem de Kripke parece caminhar na direção de uma boa classificação das patologias do tipo Mentiroso, mas falta aperfeiçoar esta solução com a adição de uma teoria do significado que torne mais precisa e bem justificada a distinção entre “falha de significado” e “falha em expressar uma proposição”.

Em suma, a Solução Ortodoxa apresenta vantagens do ponto de vista formal, é mais abrangente, e não envolve a caracterização de um tipo específico de falha semântica. Porém, pesa negativamente o fato de que ela não preserva importantes intuições, é dificilmente adaptável à linguagem natural, carece de justificção, e é exageradamente abrangente, no sentido de produzir mais estragos do que o necessário para eliminar os paradoxos. Por sua vez, a Solução Alternativa preserva melhor importantes intuições, possui boas razões para a caracterização das sentenças patológicas (por isso não é artificial), e não é exageradamente abrangente, já que não desautoriza sentenças que envolvem círculos inofensivos. Esta solução, contudo, é formalmente desconfortável (tanto porque não é satisfatoriamente expressiva, quanto porque não preserva **PB** e **PTE**), além disso, falta uma teoria do significado que suporte a classificação das sentenças patológicas.

O embate entre Solução Ortodoxa e Solução Alternativa apenas reafirma o impasse a que somos levados pelo paradoxo, e reafirma o quão genuíno é o Paradoxo do Mentiroso. Diante do confronto entre estas vias de solução, somos levados a focar principalmente as consequências indesejáveis de ambas. Na Solução Ortodoxa, abandonamos a premissa que diz que há uma sentença que diz de si mesma que é falsa. Na Solução Alternativa, abandonamos a premissa que diz que toda sentença é sempre

verdadeira ou falsa. Ao considerarmos estas vias de solução em seus detalhes, tornamos mais conhecidos e melhores explicados os resultados indesejáveis de abandonar suas respectivas premissas. Neste sentido, na pior das hipóteses, o quadro que resulta desta discussão é uma sofisticação do diagnóstico que Tarski apresentou para o Mentiroso. Conhecendo as desvantagens e as vantagens de ambas as soluções, temos uma melhor elucidação das condições impostas pelo problema. Seguindo o diagnóstico de Tarski, como tem sido feito, tornamos mais detalhado o impasse entre assumirmos ou não o fechamento semântico da linguagem.

Ao fim, nos parece que não há razões suficientes para dizermos qual destas soluções está correta. Talvez possamos argumentar que a solução de Kripke está mais próxima de uma solução correta, pela razão de que ele se preocupa ao mesmo tempo em preservar o fechamento semântico (o que é um traço característico da linguagem natural e está de acordo com a noção intuitiva de verdade) e restringir o mínimo possível da lógica clássica. Kripke, como dissemos, trata os *truth value gaps* de maneira meramente convencional (uma convenção para tratar sentenças que não expressam proposições), insistindo que uma sentença sem valor deve ser tratada como um objeto do domínio para o qual o predicado de verdade não é definido. E, neste sentido, a sua permissividade em relação a *gaps* parece afetar de maneira menos drástica a lógica clássica. Além disto, os modelos fornecidos pela solução de Kripke são fecundos e permitem a classificação de vários tipos de fenômenos específicos sob a relação de dependência semântica. Mas de qualquer maneira, esta permissividade afeta a lógica o suficiente para perdermos importantes princípios e para acabar revelando graves limitações de expressividade na própria teoria. Uma objeção que poderíamos também levantar tanto às Soluções Alternativas quanto à Solução Ortodoxa é a ausência de considerações que possam ser compartilhadas pelos paradoxos estruturalmente semelhantes que não envolvem conceitos semânticos, paradoxos que envolvem círculos viciosos e não envolvem conceitos semânticos, como por exemplo, o Paradoxo de Russell ou o Paradoxo Burali Forti, mencionados no começo deste trabalho. As soluções que apresentamos não têm nada a dizer sobre estes paradoxos, embora suas semelhanças sejam fortemente indisputáveis, o que tende a nos induzir a pensar que enquanto caracterizações da circularidade (que é um traço essencial nos paradoxos) estas soluções são meramente acidentais.

Várias razões indicam que ainda não temos uma boa solução para o Mentiroso. Sabemos que um paradoxo genuíno tende a deixar alguma insatisfação, mesmo depois

que uma boa solução é apresentada. Esperamos que uma solução ao menos permita que continuemos a utilizar os conceitos envolvidos de uma maneira natural e saudável.

Uma análise adequada de um paradoxo deve diagnosticar a fonte do problema que o paradoxo revela, e a partir daí nos ajudar a refinar os conceitos envolvidos, tornando-os verdadeiramente coerentes. Mas deveria fazê-lo de tal modo que nas situações normais as ocupações possam proceder conforme o usual. Isto é o que aconteceu, por exemplo, na teoria de conjuntos e na teoria da relatividade. Mas isto não aconteceu ainda no caso dos paradoxos semânticos, ou assim iremos afirmar. (Barwise, J. 1987, p.4).<sup>122</sup>

O novo impasse colocado pela escolha entre as duas vias de solução para o paradoxo é: ou somos levados a um uso artificial e limitado do predicado, ou podemos utilizá-lo de uma maneira mais natural, mas reconsiderando princípios fundamentais que antes sustentávamos a respeito dele, além de nos resignarmos a uma significativa perda de capacidade expressiva.

Em princípio, não temos razões suficientes para aceitar como solução definitiva nenhuma das vias aqui apresentadas. Mas daí também não se segue que temos razões para acreditar que não pode haver uma solução satisfatória para este problema. Pode ser o caso que amanhã uma solução seja apresentada com argumentos tão convincentes e que satisfaça tão bem os requisitos para uma solução, que a reação natural seja não considerar mais o Mentiroso um paradoxo, mas algo como uma falácia ou uma redução ao absurdo, provando a falsidade de alguma premissa. Por hora, entretanto, tendo em vista as condições a que se submetem as soluções genuinamente bem elaboradas, o Mentiroso preserva de alguma maneira as características de um problema aparentemente sem solução. A despeito desta aparência, é razoável que ainda sejamos orientados pelo critério wittgensteiniano:

Para uma resposta que não se pode formular, tampouco se pode formular a questão. O enigma não existe. Se uma questão se pode em geral levantar, a ela também se pode responder. (Wittgenstein L. proposição 6.5, *Tractatus*. 1921. trad. 2001, p. 279).

---

<sup>122</sup> *An adequate analysis of a paradox must diagnose the source of the problem the paradox reveals, and thereby help us refine the concepts involved, making them truly coherent. But it should do so in such a way that in normal situations business can proceed as usual. This is what has happened, for example, in set theory and relativity theory. But it has not happened in the case of the semantical paradoxes, or so we will claim.*

## Bibliografia

- ABÍLIO RODRIGUES. (2005). *Sobre a Concepção de Verdade de Tarski*. Abstracta. 2: 1. pp. 24-61.
- ARTHUR PAP. (1960). *Types and meaninglessness*. *Mind*, New Series, vol. 69. No. 273. pp. 41-54.
- BARWISE, J. and ETCHEMENDY, J. (1987). *The Liar. An Essay on Truth and Circularity*. Oxford U. P.
- BEALL JC. (2007). *Revenge of the Liar. New Essays on the Paradox*. Oxford University Press.
- BOOLOS, George. (1971). *The Iterative Conception of Set*. Em *The Journal of Philosophy*, Vol. 68, No. 8. pp. 215-231.
- BRANQUINHO, J., MURCHO, D., GOMES, N. G. (2006) *Enciclopédia de Termos Lógico-Filosóficos*. Martins Fontes.
- BURGE, T. (1979). *Semantical Paradox*. In *Recent Essays on Truth and the Liar Paradox*. Edited by Robert L. Martin (1984). Oxford U. P.
- CHATEAUBRIAND, O. (2001). *Tarski's Semantic Conception of Truth*. In *Logical Forms*, chapter 7, Part I. Campinas. Coleção CLE. pp. 213-259.
- CHARLES PARSONS. (1974). *The Liar Paradox*. In *Recent Essays on Truth and the Liar Paradox*. Edited by Robert L. Martin (1984). Oxford U. P.
- FEFERMAN, S. (1982). *Toward Useful Type-Free Theories, I*. In *Recent Essays on Truth and the Liar Paradox*. Edited by Robert L. Martin (1984). Oxford U. P.
- FIELD, H. (2008). *Saving Truth from Paradox*. Oxford University Press.
- FRAASSEN, V. (1968). *Presupposition, Implication, and Self-Reference*. *The Journal of Philosophy*, vol. 65, No. 5. pp. 136-152.
- \_\_\_\_\_, (1966). *Singular Terms, Truth-Value Gaps, and Free Logic*. *The Journal of Philosophy*, Vol. 63, No. 17. pp. 481-495.
- FREGE. (1892). *Sobre o Sentido e a Referência*. In *Lógica e Filosofia da Linguagem*. Seleção, introdução, tradução e notas de Paulo Alcoforado. UFRJ. Ed. CULTRIX. SP. (1978).
- HAACK, S. (1978) “*Filosofia das Lógicas*” Trad. Cezar Augusto Mortari e Luiz Henrique de Araújo Dutra. Ed. Unesp. (2002).
- HECK, R. G. Jr. *Formal Backgrounds for Theories of Truth*. Disponível no endereço web do autor: <http://frege.brown.edu/heck/philosophy/pdf/notes/FormalBackground.pdf> Acessado em 27/05/2010

\_\_\_\_\_, *Kripke's Theory of Truth*. Disponível no endereço web do autor: <http://frege.brown.edu/heck/philosophy/pdf/notes/KripkesTheoryOfTruth.pdf>. Acessado em 27/05/2010.

HERZBERGER, H. J. (1982). *Naïve Semantics and The Liar Paradox*. *The Journal of Philosophy*, vol. 79, no 9. pp 479-497.

KIRKHAM, R. L. (1992). *Teorias da Verdade – uma introdução crítica*. Trad. Alessandro Zir. Unisinos 2003.

KRIPKE, S. (1975). *Outline of a Theory of Truth*. In *Recent Essays on Truth and the Liar Paradox*. Edited by Robert L. Martin (1984). Oxford U. P.

MAUDLIN, T. (2004). *Truth and Paradox. Solving the Riddles*. Oxford University Press.

MARTIN, R. L. (1984). *Introduction In Recent Essays on Truth and the Liar Paradox*. Edited by Robert L. Martin (1984). Oxford U. P.

\_\_\_\_\_, (1967). *Toward a solution to the Liar Paradox*. *The Philosophical Review*, vol 76. pp 279-376.

MCGEE, V. (1991). *Truth, Vagueness, and Paradox. An Essay on the Logic of Truth*. Hackett Publishing Company.

QUINE, Willard van O. (1966) “Ways of paradox and Other Essays” Random House. New York.

REINHARDT, W. (1986) *Some Remarks on Extending and Interpreting Theories with a Partial Predicate for Truth*. *Journal of Philosophical Logic*. Vol. 15, pp. 219-251.

RESCHER, Nicholas. (2001) “Paradoxes”. Open Court. Chicago and La Salle, Illinois.

RUSSELL. (1950). *Da Denotação*. In *Os Pensadores*. Trad. Pablo Rubén Mariconda. Abril Cultural, SP. 1978.

\_\_\_\_\_, (1908) *Mathematical Logic as Based on the Theory of Types*. *American Journal of Mathematics*, Vol. 30, No. 3. pp. 222-262.

SAINSBURY, R. M. (1995) “Paradoxes”. Cambridge University Press.

SANTOS, R. (2008). *A Solução de Kripke para o Paradoxo do Mentiroso no Quadro da Lógica Clássica*, disponível pelo endereço na web: [www.filosofia.uevora.pt/paradoxos/Ricardo20080125kripke.pdf](http://www.filosofia.uevora.pt/paradoxos/Ricardo20080125kripke.pdf). Acessado em 27/05/2010.

\_\_\_\_\_, (2003). *A Verdade de um Ponto de Vista Lógico-Semântico*. Fundação Calouste Gulbenkian. Fundação para a Ciência e a Tecnologia.

SIMMONS, K. (2002). “*Semantical and Logical Paradox*,”. In *A Companion to Philosophic Logic*. Edited by Dale Jacquette. Blackwell Companions to Philosophy.

SOAMES, S. (1999). *Understanding Truth*. Oxford U.P.

\_\_\_\_\_, (1989). *Presupposition*. In *Handbook of Philosophical Logic*, vol. IV. pp. 553-616. Edited by Gabbay and Guentner. D. Reidel Publishing Company.

SORENSEN, R. (2002). *Philosophical Implication of Logical Paradox*. In *A Companion to Philosophical Logic*. Edited by Dale Jacquette. Blackwell Companions to Philosophy.

STRAWSON, P. (1950) “*Sobre Referir*”. Trad. Balthazar Barbosa Filho. *Escritos Lógico-Linguísticos. Ensaio. Os Pensadores*. Abril Cultural.

TARSKI, A. (1933). *O Conceito de Verdade nas Linguagens Formalizadas*. Trad. C. A. Mortari e L. H. A. Dutra. Em *A Concepção Semântica da Verdade. Textos Clássicos de Tarski*. Unesp. 2006.

\_\_\_\_\_, (1944). *A Concepção Semântica da Verdade e os Fundamentos da Semântica*. Trad. L. H. A. Dutra. Em *A Concepção Semântica da Verdade. Textos Clássicos de Tarski*. Unesp. 2006.

\_\_\_\_\_, (1969). *Verdade e Demonstração*. Trad. Jesus de Paula Assis. Em *A Concepção Semântica da Verdade. Textos Clássicos de Tarski*. Unesp. 2006.

TERENCE PARSONS. (1984). *Assertion, Denial, and The Liar Paradox*. *Journal of Philosophical Logic*. 13:2, pp. 137-152.

TILES, Mary. (1989) *The Philosophy of Set Theory. An Historical Introduction to Cantor's Paradise*. Dover Publications. New York.

WILLIAMSON, T. (1992). *Vagueness and Ignorance*. In *Proceedings of the Aristotelian Society, Supplementary Volumes*. Vol 66. pp 145-177.