

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS
DEPARTAMENTO DE FILOSOFIA

RAFAEL DOS SANTOS ONGARATTO

O PARADOXO DO MENTIROSO
uma abordagem deflacionista

PORTO ALEGRE
2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS
DEPARTAMENTO DE FILOSOFIA

RAFAEL DOS SANTOS ONGARATTO

O PARADOXO DO MENTIROSO

uma abordagem deflacionista

Trabalho de Conclusão de Curso de graduação
apresentado ao Departamento de Filosofia do
Instituto de Filosofia e Ciências Humanas da
Universidade Federal do Rio Grande do Sul, como
requisito parcial para a obtenção do título de
Bacharel em Filosofia.

Orientador: Paulo Francisco Estrella Faria

PORTO ALEGRE

2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

REITOR

Carlos André Bulhões Mendes

VICE-REITORA

Patricia Pranke

DIRETOR DO INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

Hélio Ricardo do Couto Alves

VICE-DIRETOR DO INSTITUTO DE FILOSOFIA E CIÊNCIAS HUMANAS

Alex Niche Teixeira

CHEFE DA BIBLIOTECA DE CIÊNCIAS SOCIAIS E HUMANIDADES

Fabiana Hennies Brigidi

CIP - Catalogação na Publicação

Ongaratto, Rafael dos Santos
O Paradoxo do Mentiroso: uma abordagem
deflacionista / Rafael dos Santos Ongaratto. -- 2021.
50 f.
Orientador: Paulo Francisco Estrella Faria.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Instituto
de Filosofia e Ciências Humanas, Bacharelado em
Filosofia, Porto Alegre, BR-RS, 2021.

1. deflacionismo. 2. paradoxo do mentiroso. 3.
Kripke. 4. Tarski. 5. Ramsey. I. Francisco Estrella
Faria, Paulo, orient. II. Título.

Rafael dos Santos Ongaratto

O PARADOXO DO MENTIROSO
uma abordagem deflacionista

Trabalho de Conclusão de Curso de graduação apresentado ao Departamento de Filosofia do Instituto de Filosofia e Ciências Humanas da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Bacharel em Filosofia.

Porto Alegre, 30 de novembro de 2021

Resultado: Aprovado

BANCA EXAMINADORA:



Sílvia Altmann

Departamento de Filosofia

Universidade Federal do Rio Grande do Sul (UFRGS)



Rogério Passos Severo

Departamento de Filosofia

Universidade Federal do Rio Grande do Sul (UFRGS)

"The logical paradoxes [...] have been around for a long time now. Yet no solution has been found."

Graham Priest, The Logic of Paradox

"The necessity to ascend to a metalanguage may be one of the weaknesses of the present theory. The ghost of the Tarski hierarchy is still with us."

Saul Kripke, Outline of a Theory of Truth

"[...] nobody, as far as I know, has noted that deflationists, unlike virtually everyone else, seem to have a non-ad hoc reason for maintaining that the given sentences are meaningless."

Jeffrey Beall, A neglected deflationist approach to the liar

RESUMO

Atualmente, há uma gama de teorias da verdade que procuram resolver o paradoxo do mentiroso, cada qual abandonando algumas das pressuposições básicas da que o engendram. A discussão contemporânea inaugural desse debate foi realizada por um defensor da lógica clássica: Alfred Tarski. Em seu diagnóstico, o paradoxo é derivado da conjunção das seguintes suposições: (I) a linguagem-objeto é uma linguagem *semanticamente fechada*, e (II) as leis básicas da lógica valem nessa linguagem. Enquanto Tarski negou (I), a maioria de seus sucessores buscou reformar a lógica, negando (II). O presente texto realiza um estudo comparativo entre os textos de Tarski, Kripke, e autores deflacionistas, para determinar se as soluções alternativas ao paradoxo são capazes de solucioná-lo mais satisfatoriamente do que a solução clássica. O objetivo deste estudo é, primeiramente, estabelecer um mapa conceitual de algumas das principais posições lógico-filosóficas em resposta ao paradoxo do mentiroso. Em segundo lugar, meu objetivo é defender uma resposta deflacionista ao paradoxo do mentiroso, que busca inspiração na formulação deflacionista de Frank Ramsey.

Palavras-chave: deflacionismo, Kripke, paradoxo do mentiroso, Ramsey, Tarski.

ABSTRACT

Currently, there is a variety of theories of truth that attempt to solve the liar paradox, each one giving up some of the basic assumptions that engender it. The first contemporary discussion of this debate has been made by a defender of classical logic: Alfred Tarski. In his diagnostic, the paradox is derived from the conjunction of the following suppositions: (I) the object-language is *semantically closed*, and (II) the basic laws of logic are valid in this language. The present text presents a comparative study of Tarski's and Kripke's writings, as well as of deflationist authors, to determine whether non-classical solutions to the paradox are able to solve it more satisfactorily than the classical solution. The objective of this study is, firstly, to establish a conceptual mapping of the main logico-philosophical positions in response to the paradox of the liar. Secondly, my objective is to defend a deflationist response to the liar paradox, which is inspired by Ramsey's formulation of deflationism.

Keywords: deflationism, liar paradox, Kripke, Ramsey, Tarski.

SUMÁRIO

INTRODUÇÃO	8
1 ALFRED TARSKI E O DIAGNÓSTICO CLÁSSICO	10
1.1 Linguagens Formalizadas	12
1.2 Esquema T.....	14
1.3 O Paradoxo do Mentiroso Formalizado	15
1.4 A solução hierárquica de Tarski.....	16
1.5 Objeções à solução clássica	19
1.6 Respostas às objeções.....	21
2 A SOLUÇÃO INORTODOXA DE KRIPKE.....	22
2.1 A crítica kripkeana à abordagem ortodoxa.....	22
2.2 As condições para uma teoria da verdade.....	25
2.3 A solução kripkeana	27
2.4 Objeções a Kripke	31
3 A SOLUÇÃO DEFLACIONISTA.....	34
3.1 A concepção clássica de verdade.....	34
3.2 A concepção deflacionista da verdade	35
3.3 A solução de inspiração ramseyana.....	37
3.4 Críticas ao deflacionismo	39
4 CONCLUSÃO	46
REFERÊNCIAS BIBLIOGRÁFICAS.....	48

INTRODUÇÃO

O debate sobre o paradoxo do mentiroso, embora ainda em curso, já apresenta certo grau de maturidade: uma variedade de posições lógico-filosóficas está disponível, cada qual desenvolvendo sistemas formais cada vez mais criativos para contorná-lo. Assim, é importante, a partir desse estágio, realizar um trabalho de *seleção*; primeiro, colhemos o máximo de grãos que pudermos. Assim, no trabalho a seguir, explorarei diferentes respostas oferecidas ao paradoxo do mentiroso na história, e procurarei desenvolver uma resposta alternativa com base na teoria deflacionista da verdade: nossa jornada, no entanto, começa pelo diagnóstico clássico.

O ‘diagnóstico clássico’ refere-se à solução oferecida por Tarski em suas exposições relacionadas à teoria da verdade. Resumidamente, sua proposta é construir uma linguagem formalizada em que não seja possível gerar frases como ‘Esta frase é falsa’, ou seja, frases que afirmem algo sobre seu próprio valor de verdade. O problema com algumas dessas frases é que elas parecem ser tanto verdadeiras quanto falsas. Consideremos a frase

(1) '1' é falsa.

A frase (1) é verdadeira ou falsa? Suponhamos que (1) seja verdadeira, ou seja, aquilo que (1) *afirma* é verdadeiro; logo, (1) é falsa. Suponhamos, por outro lado, que (1) seja falsa. Nesse caso, é o caso que ' '1' é falsa', ou seja, aquilo que (1) afirma é verdadeiro. Logo, (1) é verdadeira. Assim, (1) é verdadeira se e somente se (1) é falsa. Ora, como toda frase é ou verdadeira ou falsa, segue-se que a sentença (1) é verdadeira *e* falsa¹.

A maneira pela qual Tarski barrou essas frases foi negando a *universalidade* à linguagem-objeto em questão, ou seja, essa linguagem não pode se referir a e falar sobre *quaisquer* objetos irrestritamente, como é o caso da linguagem natural. Na linguagem natural, podemos falar, por exemplo, de cachorros: um animal de quatro patas, alguns grandes, alguns pequenos. Também podemos falar de 'cachorros': uma palavra de 9 ocorrências de letras e 6 tipos de letras; também podemos falar de "cachorros": o nome do nome de cachorros. E assim por diante. Essa característica da linguagem natural permite a construção de frases como 'Esta frase é falsa', em que a expressão 'Esta frase' refere-se à frase 'Esta frase é falsa'. Essa irrestrrição aplica-se não só a nomes, mas também aos predicados que afirmam a verdade de uma frase. Por exemplo, 'A neve é branca' é verdadeira' é uma frase que afirma a verdade de

¹ Neste primeiro momento, eu apresentei o paradoxo do mentiroso propositalmente na linguagem natural, pois a construção *formal* de um predicado de verdade não é trivialmente equivalente ao uso comum do predicado ‘é verdadeiro’.

outra frase ('A neve é branca') por meio do nome dessa frase ("A neve é branca"), ou seja, essa frase predica algo de outra frase, diferentemente de 'A neve é branca', que predica a brancura de um objeto (a neve). A partir dessa ideia, Tarski estabeleceu uma *hierarquia* de linguagens, em que uma linguagem de nível 0, por exemplo, contém apenas predicados sobre objetos, como 'A neve é branca', ao passo que a linguagem de nível 1 permite também atribuir propriedades às próprias sentenças do nível 0, com construções como 'A neve é branca' é verdadeira'. Desse modo, uma determinada linguagem não pode falar sobre si mesma, ou seja, não são permitidas construções em que se tome, na linguagem de nível n , uma sentença Φ de nível n e se afirme " Φ é verdadeira', nem " Φ é falsa'. Assim, barra-se o paradoxo do mentiroso: supondo, por exemplo, que a frase (1) seja de nível n , não é possível que a frase, ao afirmar a falsidade de uma frase, esteja se autorreferindo, pois nesse caso uma frase de nível n estaria predicando a falsidade de uma frase de nível n . Embora seja preciso abandonar a noção de um predicado *global* de 'verdade', consegue-se manter intactas as leis lógicas básicas.

Uma das principais objeções feitas à solução tarskiana é que a hierarquia de linguagens não captura o funcionamento da linguagem natural. Em português, por exemplo, podemos ouvir alguém falando alguma frase em que acreditamos e afirmar "O que ele disse é verdadeiro", seja qual for essa frase. Nessa medida, a resposta de Kripke procura fazer jus a essa objeção, dispendendo mais atenção às características efetivas da linguagem natural: Kripke apresenta uma linguagem em que há fechamento semântico (ainda que parcial), ou seja, é possível atribuir, em alguma medida, *universalidade* a essa linguagem.

Essa solução alternativa, não obstante, também possui pontos fracos. A sua proposta de construir uma linguagem em que o predicado de verdade é definido *parcialmente* – isto é, em que o predicado-verdade (e sua anti-extensão) não se aplica a certas frases – é ameaçado por um retorno à hierarquia de linguagens de Tarski, o que derrotaria seus propósitos iniciais de construir uma linguagem semanticamente fechada. Por fim, a solução deflacionista que pretenderei desenvolver também sofre de ameaças de inexpressabilidade, mas, como tentarei argumentar, a restrição de frases contendo o uso substancial do predicado de verdade é bem justificada.

1 ALFRED TARSKI E O DIAGNÓSTICO CLÁSSICO

Em 1933, em *O Conceito de Verdade em Linguagens Formalizadas*², Alfred Tarski focaliza seus esforços em uma tarefa simples, mas difícil: definir o predicado 'ser verdadeiro':

O presente artigo é quase completamente devotado a um único problema – a *definição de verdade*. Seu desafio é construir – em relação a uma dada linguagem – uma *definição materialmente adequada e formalmente correta do termo 'frase verdadeira'*. (TARSKI, 1956[1933], p. 152, tradução minha)

Os detalhes de tal construção ultrapassam o escopo do presente trabalho; o que pretendo mostrar nesta seção é de que modo Tarski, a partir de suas considerações sobre o paradoxo do mentiroso na linguagem natural, mobiliza-se para construir uma linguagem formalizada que não permita derivá-lo.

O primeiro passo de Tarski³ – e o passo natural – ao procurar uma definição de verdade é avaliar se o uso comum de tal termo é *consistente*, ou seja, o uso de frases com o predicado 'é verdadeiro' ou 'é falso' deve ser tal que não seja possível derivar contradições de nenhuma frase que seja da forma 'Φ é verdadeiro' nem 'Φ é falso', portanto, Tarski se volta para a linguagem natural no §1 de *O Conceito de Verdade em Linguagens Formalizadas*.⁴ No entanto, como visto na introdução, a linguagem natural permite construções como "Esta frase é falsa", o que demonstra sua inconsistência.⁵ Quais são as características da linguagem natural que permitem seu surgimento? Segundo Tarski, as características presentes na linguagem que, conjuntamente, permitem derivar o paradoxo do mentiroso são as seguintes (TARSKI, 1944, p. 348): (I) a linguagem em questão é *semanticamente fechada*, ou seja, ela permite que, dentro dessa linguagem, construam-se frases que afirmem a verdade ou a falsidade de quaisquer frases dessa linguagem; (II) nessa linguagem, as leis da lógica clássica são válidas - em especial, o princípio do terceiro excluído e o princípio de explosão.

Além disso, há um elemento adicional da construção de uma definição de verdade que precisa ser levado em conta. Uma primeira tentativa de Tarski de formular uma caracterização de verdade informal remonta à caracterização de Aristóteles, segundo a qual (III) "uma frase

² Ver Tarski (1956, p. 152-278). Um trabalho posterior de Tarski, cuja função é expor os resultados do precedente de maneira menos técnica – que também irei utilizar – é *A Concepção Semântica de Verdade* (TARSKI, 1944).

³ Idem, p. 154-165.

⁴ Ver Tarski (1956, p. 154, tradução minha): "Para o propósito de introduzir o leitor ao assunto, uma consideração – talvez apenas transitória – do problema de definir a verdade em uma linguagem coloquial é desejável."

⁵ Scott Soames procura argumentar, contrariamente ao que Tarski pretendeu, que poderíamos tentar solucionar o paradoxo do mentiroso na própria linguagem natural, estabelecendo uma hierarquia, por exemplo, de frases em português (SOAMES, 1998, p. 151). No entanto, não entrarei nessa disputa. Por ora, seguirei o argumento conforme Tarski o expõe.

verdadeira é uma que afirma que o estado de coisas é tal, e o estado de coisas é tal. " (TARSKI, 1956, p. 155, tradução minha). Ao utilizar essa caracterização inicial de verdade na linguagem coloquial, somos capazes de derivar o paradoxo do mentiroso. Basicamente, o que (III) afirma é que, por exemplo, "O livro está sobre a mesa" ser uma frase verdadeira depende de o livro estar sobre a mesa, ou seja, o estado de coisas que é descrito pela frase deve ocorrer. De maneira esquemática, isso significa que "O livro está sobre a mesa" é verdadeira se, e somente, o livro está sobre a mesa.

Podemos denominar tal afirmação generalizada como o "esquema-T⁶ informal". Embora Tarski reserve o termo "esquema-T" apenas para linguagens formais, estender o uso para linguagens coloquiais nos permite construir de maneira esquemática o argumento cuja conclusão é o paradoxo do mentiroso.

Dessa maneira, a derivação do paradoxo na linguagem natural pode ser montada da seguinte forma:

- (1) '1' é falsa [de (I)]
- (2) '1' é verdadeira sse 1 [esquema-T informal]
- (3) '1' é verdadeira sse '1' é falsa [1, 2, substituição 1/'1' é falsa]
- (4) '1' é verdadeira ou '1' é falsa [princípio do terceiro excluído]
- (5) '1' é verdadeira e '1' é falsa [3, 4, lógica proposicional]
- (6) O papai noel existe [5, princípio de explosão]

Assim, temos uma primeira avaliação daquilo que permite o surgimento do paradoxo do mentiroso na linguagem natural – supondo que a linguagem natural seja como a caracterizada por Tarski: primeiramente, deve ser possível gerar frases do tipo (1) – a condição (I). Em segundo lugar, o esquema-T informal é válido para todas as frases. A partir de (1) e (2), é possível derivar (3), um bicondicional que afirma que uma frase é verdadeira se, e somente se, ela é falsa. Não temos ainda, no entanto, uma contradição. Para derivá-la, é preciso aceitar (4), que é uma instância do princípio do terceiro excluído (doravante denominado 'PTE'), que afirma que toda frase é verdadeira ou falsa, ou seja, não pode ser o caso que uma frase não seja nem verdadeira nem falsa:

(PTE) Toda frase é verdadeira ou falsa.

Assim, assumindo o PTE e (3), podemos derivar (5); por fim, temos uma contradição flagrante. Poderíamos parar por aí, simplesmente afirmando que contradições não podem ser verdadeiras em nosso sistema, pois isso seria absurdo. No entanto, qual é o problema *global*

⁶ O "esquema-T" é uma tradução de "T-schema", que abrevia a expressão "Truth schema".

de se aceitar contradições em um sistema lógico clássico? O cerne do problema se encontra no fato de que o sistema lógico permite derivar o princípio de explosão (doravante denominado ‘PE’), que afirma que *toda* frase pode ser derivada logicamente a partir de uma contradição:

(PE) Se uma contradição é verdadeira, então qualquer frase é verdadeira.

Portanto, se (5) for verdadeira, posso provar, a partir de PE, (6), assim como quaisquer outras frases da linguagem. Desse modo, o sistema torna-se trivial, pois toda frase é verdadeira, o que torna a linguagem inútil para propósitos cognitivos⁷. Logo, é possível entender a preocupação central de Tarski em evitar que sua definição de verdade seja montada em uma linguagem que derive o paradoxo do mentiroso.

Não obstante, a avaliação de Tarski em relação à linguagem natural é oscilante. Por um lado, em 1933, ele sugere que a linguagem natural é inconsistente:

Um traço característico da linguagem coloquial (em contraste com várias linguagens científicas) é sua universalidade. [...] Essas antinomias parecem fornecer uma prova de que toda linguagem que é universal no sentido acima e para a qual as leis da lógica normais valem deve ser inconsistente. (TARSKI, 1956[1933], p. 164-5, tradução minha)

Por outro lado, em 1944, o autor suspende o juízo em relação à definição de verdade na linguagem coloquial, por sua suposta indefinição estrutural:

Nossa linguagem cotidiana certamente não tem uma estrutura exatamente especificada. Não sabemos precisamente quais expressões são frases e sabemos ainda menos quais são asseríveis. Logo, o problema da consistência não possui um significado exato relativamente a essa linguagem. Podemos, no máximo, apenas arriscar achar que uma linguagem cuja estrutura foi exatamente especificada e que se assemelha à nossa linguagem cotidiana tanto quanto possível seria inconsistente. (TARSKI, 1944, p. 349, tradução minha)

De qualquer modo, seja para construir uma linguagem que não seja universal, seja para poder especificar a estrutura de suas expressões, uma resposta satisfatória ao paradoxo do mentiroso, segundo Tarski, só pode ser feita em uma linguagem formalizada.⁸

1.1 Linguagens Formalizadas

⁷ Como apontado para mim pelo professor Rogério Severo, é importante ressaltar, nesse caso, o aspecto *cognitivo* da linguagem, pois a trivialidade pode ser apreciada para outros propósitos de uma determinada linguagem, seja em usos expressivos, usos recreativos, etc. Em *Begriffsschrift*, por exemplo, Frege está ciente das limitações de sua linguagem como restrita a *certos propósitos*: "Essa conceitografia, da mesma maneira, é um dispositivo inventado para certos propósitos científicos e não se deve condená-la por não ser adequada para outros." (FREGE, 1967, p. 6, tradução minha).

⁸ Ver Tarski (1956, p. 165, tradução minha): "Pelas razões dadas na seção precedente, eu agora abandono a tentativa de resolver nosso problema para a linguagem cotidiana e me restrinjo doravante inteiramente a *linguagens formalizadas*".

Linguagens formalizadas, segundo Tarski, são linguagens em que é possível determinar *exaustivamente* as formas possíveis das expressões que ocorrem em seu vocabulário, ou seja, dada uma sequência de sinais, é possível indicar, a partir de sua forma sintática, se tal sequência é uma expressão da linguagem ou não. Além disso, dentre tais expressões, algumas são classificadas inequivocamente como *frases*, isto é, expressões que possuem um valor de verdade. Ao contrário da linguagem natural, uma frase em uma linguagem formalizada é *inteiramente* determinada por sua forma sintática. Por exemplo, na linguagem natural, pode haver dúvida sobre o significado da expressão "Todo menino ama uma menina". Essa frase pode significar que (i) há *uma* menina que todo menino ama, ou (ii) para todo menino, há alguma menina que ele ama. Em uma linguagem formalizada, no entanto, não há ambiguidades desse tipo. Na linguagem de predicados com quantificadores, por exemplo, a frase acima pode ser aproximada inequivocamente como (i) $\exists y \forall x Axy$ ou (ii) $\forall x \exists y Axy$.

Essas características podem ser sistematizadas como (a) "para cada uma dessas linguagens, uma lista ou descrição é dada em termos estruturais de todos os *sinais a partir dos quais as expressões da linguagem são formadas*", e (b) "dentre todas as possíveis expressões que podem ser formadas por esses sinais, aquelas chamadas sentenças são distinguidas por meio de propriedades puramente estruturais" (TARSKI, 1956, p. 166). Assim, a especificação completa das expressões de uma linguagem formalizada pode ser estabelecida por meio de uma enumeração das formas dos diferentes tipos de expressão e das constantes lógicas, ao passo que a especificação das frases normalmente é estabelecida por uma definição recursiva: primeiramente, as fórmulas básicas (atômicas) são estabelecidas. Em segundo lugar, mostram-se como as demais fórmulas (complexas) são geradas a partir das atômicas. Como exemplo, considere a introdução à lógica proposicional de Raymond Smullyan:⁹

Usaremos como nossos conectivos lógicos os seguintes símbolos:

- | | | | |
|-----|-------------------------|-----|------------------------------------|
| (1) | \sim [leia-se "não"], | (2) | \wedge [leia-se "e"] |
| (3) | \vee [leia-se "ou"], | (4) | \rightarrow [leia-se "implica"]. |

Esses símbolos são chamados, respectivamente, símbolos de negação, da conjunção, da disjunção e da implicação. Os três últimos são chamados coletivamente *binários*; o primeiro (\sim), conectivo *unário*.

Nossos outros símbolos serão:

- (i) Um conjunto enumerável $p_1, p_2, \dots, p_n, \dots$ de símbolos chamados *variáveis proposicionais*.
- (ii) Os dois símbolos (e) são respectivamente chamados *parêntese esquerdo* e *parêntese direito* [...]. A noção de *fórmula* é dada, então, por meio das seguintes

⁹ Quando eu passar a apresentar o paradoxo do mentiroso de maneira formal, recorrerei à notação de Smullyan em seu livro *Lógica de Primeira Ordem*.

regras recursivas que nos possibilitam obter novas fórmulas a partir de fórmulas já construídas:

F_0 : Toda variável proposicional é uma fórmula.

F_1 : Se A é uma fórmula, $\sim A$ é uma fórmula.

F_2, F_3, F_4 : Se A e B são fórmulas, $(A \wedge B)$, $(A \vee B)$, $(A \rightarrow B)$ são fórmulas. (SMULLYAN, 2009, p. 5).

Tratando-se do paradoxo do mentiroso em linguagens formalizadas, não obstante, ainda resta um problema: o que significa dizer que uma linguagem determinada é inconsistente? Certamente, isso não quer dizer que no conjunto de fórmulas dessa linguagem há *tanto uma fórmula p quanto sua negação*. Nesse caso, a lógica proposicional seria inconsistente pelo simples fato de haver, para toda frase p , a sua contraparte $\sim p$. A sugestão de Martin Pleitz é entender que Tarski afirma a inconsistência de uma linguagem derivativamente (PLEITZ, 2017, p. 88): uma linguagem formalizada é apenas um conjunto de expressões reunidas por certo critério; um *sistema* de linguagem formalizada, por outro lado, é uma linguagem formalizada acrescida de um sistema dedutivo; por fim, uma *teoria formal* é um sistema de linguagem formalizada acrescido de axiomas. A atribuição de consistência ou inconsistência, segundo Pleitz, é corretamente utilizada quando atribuída a teorias formais, pois, nesse caso, pode-se dizer que uma teoria formal é inconsistente se os axiomas derivam, como teoremas, uma frase e sua negação.¹⁰ Linguagens formalizadas, por outro lado, são inconsistentes apenas em um sentido derivado: uma linguagem formalizada é inconsistente se sua teoria formal é inconsistente.

1.2 Esquema T

Tendo feito a distinção entre linguagens formalizadas e linguagens informais, podemos apresentar a condição anteriormente denominada de 'esquema T' em um contexto formal. A partir de sua exposição informal da definição de verdade no § 1, Tarski procura transpor seus resultados para linguagens formalizadas. Em um contexto formal, é garantido que cada expressão seja determinada inequivocamente e, para cada expressão, há um nome que se refira a ela. Assim, o autor pode apresentar uma definição *parcial* de verdade, ou seja, uma condição *necessária* para quaisquer definições de verdade propostas serem corretas:

¹⁰ Embora Martin Pleitz enfatize o papel dos axiomas na derivação de contradições, existem sistemas lógicos sem axiomas, como o sistema de dedução natural, em que os teoremas são derivados puramente por regras de inferência. Nesse sentido, é mais acurado afirmar que a inconsistência se atribui a uma teoria formal quando ela possui axiomas; caso ela seja um sistema baseada apenas em regras de inferência, faz mais sentido atribuir inconsistência ao sistema da linguagem.

CONVENÇÃO T. Uma definição formalmente correta do símbolo 'Tr', *formulada na metalinguagem, será chamada de uma definição adequada de verdade se ela possui as seguintes consequências:*

(α) *todas as frases que são obtidas da expressão 'x \in Tr se, e somente se, p', substituindo pelo símbolo 'x' um nome estrutural-descritivo de qualquer frase da linguagem em questão e para o símbolo 'p' a expressão que forma a tradução dessa frase na metalinguagem;*

(β) *a frase 'para todo x, se x \in Tr, então x \in S' [...] (TARSKI, 1956, p. 188, tradução minha)*

Isto é, dada uma definição proposta γ de verdade, γ deve necessariamente derivar (α) e (β). A segunda condição apenas afirma que, se uma frase pertence ao conjunto das frases verdadeiras, então essa frase pertence ao conjunto de frases, ou seja, toda frase verdadeira é uma frase. A primeira condição, por outro lado, é uma asserção do esquema T generalizado. É importante ressaltar a diferença entre o esquema T conforme exposto em (α) e o que anteriormente denominei de "esquema T informal". No caso do esquema T de (α), podemos especificar uma classe de termos 'x', que são nomes estruturais-descritivos¹¹ de certas frases p, pois a metalinguagem é tal que, para cada frase da linguagem-objeto (que está determinada inequivocamente), há um nome correspondente na metalinguagem. Dessa maneira, podemos formular o esquema T como

(Esquema T) $\forall s, p (Ts \leftrightarrow p)$ ¹², onde 's' é um nome para a frase 'p'

1.3 O Paradoxo do Mentiroso Formalizado

A partir das considerações feitas nas seções anteriores, podemos formular o paradoxo do mentiroso de maneira formalizada:

(p ₁) Fp ₁	[frase do mentiroso]
(p ₂) Tp ₁ \leftrightarrow p ₁	[esquema T aplicado à p ₁]
(p ₃) Tp ₁ \leftrightarrow Fp ₁	[p ₁ , p ₂ , Lei de Leibniz]
(p ₄) Fp ₁ \rightarrow Tp ₁	[p ₃ , lógica proposicional]
(p ₅) Tp ₁	[p ₁ , p ₄ , modus ponens]
(p ₆) Tp ₁ \wedge Fp ₁	[p ₁ , p ₅ , introdução do ' \wedge ']

¹¹ Esses nomes são estruturais-descritivos por conta de sua referência ser estabelecida por meio da caracterização (na metalinguagem) da estrutura formal da frase (na linguagem-objeto).

¹² Leia-se 'Tp' como 'p é verdadeira'. Ver Smullyan (2017, p. 18): "**Definição.** Sob qualquer interpretação, a fórmula sinalizada TX é dita *verdadeira* se X é verdadeira, e *falsa* se X é falsa, e a fórmula sinalizada FX é dita *verdadeira* se X é falsa, e *falsa* se X é verdadeira." Embora Smullyan utilize "T" e "F" como termos da linguagem-objeto, aqui os estaremos utilizando como termos da metalinguagem.

Formulado dessa maneira, o paradoxo do mentiroso ainda não está inteiramente formalizado, pois p_1 ainda depende de "um fato semântico não-trivial externo à dedução" (PLEITZ, 2017, p. 96, tradução minha), que é a identidade entre p_1 e Fp_1 . Isto é, p_6 não é derivado como teorema (para isso, as premissas deveriam ser todas axiomas ou dedutíveis de axiomas), pois a frase p_1 arbitrariamente determina autorreferência. O que precisamos é uma maneira formal de construir uma frase equivalente a p_1 que se baseie apenas nas características explícitas da linguagem formal em questão.

Felizmente, há uma maneira de construir uma frase autorreferencial que não dependa de um fato externo ao sistema lógico, que se baseia no *Lema Diagonal*:

Aquilo com que o Lema Diagonal contribui para a transferência do paradoxo do Mentiroso a um contexto formal é uma contraparte formal (ou análogo) de autorreferencialidade frasal, que não depende de nenhuma informação semântica externa. (PLEITZ, 2017, p. 98, tradução minha).

Isto é, o lema diagonal garante que haja uma frase que predique algo de si mesma (nesse caso, a falsidade). Desse modo, garantimos (p_1) – ou alguma formulação equivalente – em um contexto formal.¹³

Recapitulando o diagnóstico de Tarski, ele identifica duas suposições centrais ao argumento cuja conclusão é uma contradição: (I) a suposição de que a linguagem em questão é semanticamente fechada, ou seja, a suposição de que há nomes para todas as frases dessa linguagem, e, além disso, que os predicados de verdade e falsidade podem ser aplicados irrestritamente; (II) a suposição de que as leis clássicas da lógica valem nessa linguagem.

Assim, até este ponto, a lição de Tarski consiste em ter *identificado* as pressuposições subjacentes ao paradoxo do mentiroso e chegado à conclusão que é preciso descartar *alguma* delas:

[...] as suposições (I) e (II) são essenciais [para a geração do paradoxo]. Já que toda linguagem que satisfaz ambas suposições é inconsistente, nós temos que rejeitar pelo menos uma delas. (TARSKI, 1944, p. 349, tradução minha)

1.4 A solução hierárquica de Tarski

¹³ Por conta do presente trabalho ser uma exposição geral do diagnóstico de Tarski, não há espaço para desenvolver o tema da diagonalização. Brevemente, o método de Tarski consiste em estabelecer uma correspondência numérica entre uma frase qualquer "E" e sua frase *diagonalizada*, $\forall v_1(v_1 = e \rightarrow E)$, em que "e" é o número de Gödel de "E" (SMULLYAN, 1994, p. 82). Desse modo, é possível estabelecer uma frase autorreferencial.

Tendo identificado as premissas necessárias para a geração do paradoxo do mentiroso, Tarski opta por conservar as leis básicas da lógica em detrimento da universalidade da linguagem:

Seria supérfluo ressaltar aqui as consequências de rejeitar a suposição (II), isto é, de mudar a nossa lógica (supondo que isso seja possível) mesmo em suas partes mais elementares e fundamentais. Assim, nós consideramos apenas a possibilidade de rejeitar a suposição (I). Dessa maneira, nós decidimos *não utilizar nenhuma linguagem que seja semanticamente fechada no sentido dado*. (TARSKI, 1944, p. 349, tradução minha)

Desse modo, as linguagens em que Tarski pretende buscar uma definição de verdade são tais que o predicado 'é verdadeiro' não se aplica de maneira irrestrita a quaisquer frases. É com o propósito de restringir o uso desse predicado que Tarski introduz a distinção entre uma *linguagem-objeto* e uma *metalinguagem*. Por um lado, ao determinar uma linguagem para a qual se procura definir o predicado 'é verdadeiro', essa linguagem é a linguagem *sobre a qual* se fala, ou seja, a linguagem-objeto. Por outro lado, como não é permitido afirmar, *nessa* linguagem, que suas frases são verdadeiras (ou falsas), determina-se outra linguagem, a metalinguagem, que possa mencionar todas as frases da linguagem-objeto, além de ser capaz de afirmar a verdade ou falsidade dessas frases. A metalinguagem, por isso, é a linguagem *na qual* se fala.

Por exemplo, poderíamos delimitar uma linguagem L cujas frases incluíssem frases sobre estados de coisas existentes no mundo.¹⁴ Nessa linguagem, estariam incluídas as frases 'O gato está miando', 'O livro está debaixo da cama', etc. No entanto, não poderíamos afirmar, em L, "'O gato está miando' é verdadeiro", "'O livro está debaixo da cama' é falso", pois essas afirmações não estariam incluídas no poder expressivo de L. Por outro lado, poderíamos considerar uma metalinguagem L*, em que houvesse nomes para todas as frases de L, além de ser possível aplicar, a esses nomes, o predicado 'é verdadeiro' ou 'é falso'. Em L*, poderíamos afirmar frases como "'O gato está miando' é verdadeiro". No entanto, poderíamos reiterar tal operação para qualquer frase? Nesse caso, seria possível afirmar que "'O gato está miando' é verdadeiro' é verdadeiro". Tarski, não obstante, ao rejeitar a possibilidade de se adotar uma linguagem semanticamente fechada, nega a possibilidade de reiterar indefinidamente o predicado 'é verdadeiro' para quaisquer frases em uma determinada linguagem.

Dessa maneira, é possível evitar o paradoxo do mentiroso na linguagem-objeto, pois ela não contém seu próprio predicado de verdade, assim como na metalinguagem, pois ela

¹⁴ Esse exemplo é meramente ilustrativo, não pretendo me comprometer com quaisquer tipos de metafísicas realistas.

também não contém seu próprio predicado de verdade.¹⁵ Como consequência, o predicado ‘é verdadeiro’ é um predicado ambíguo: ele pode significar ‘é verdadeiro₁’, no caso em que ocorre na linguagem-objeto; ‘é verdadeiro₂’, no caso em que o predicado ocorre na metalinguagem; de maneira geral, ‘é verdadeiro_n’ é um predicado que ocorre na linguagem de nível n da hierarquia de linguagens. Nenhuma linguagem da hierarquia é semanticamente fechada, pois, a cada nível i , a linguagem não aplica o predicado ‘é verdadeiro_i’ para frases de nível i .¹⁶

Soames explica de maneira mais detalhada como a hierarquia de linguagens consegue barrar o paradoxo do mentiroso:

1. A frase (1) não é verdadeira_n.

P1. 'A frase (1) não é verdadeira_n' é verdadeira_n sse a frase (1) não é verdadeira_n.

Nessa interpretação, (P1) não é uma instância correta do esquema-T, já que a frase que substitui ambas ocorrências de 'S' no esquema – a saber, a frase (1) – não é uma frase de E_n. Assim, ela pode ser seguramente rejeitada. (SOAMES, 1998, p. 153, tradução minha)

Isto é, o aparente esquema-T produzido por (1) é uma frase mal formada, e, portanto, não pode ser utilizada no argumento do paradoxo do mentiroso. Para aplicarmos o predicado 'é verdadeiro' à frase (P1) corretamente, devemos atentar ao seu nível: como ela utiliza o predicado ‘é verdadeiro_n’, é necessário afirmar ‘A frase (1) não é verdadeira_n’ é verdadeira_{n+1}. Nesse caso, (P1) seria reformulado como

P1*. ‘A frase (1) não é verdadeira_n’ é verdadeira_{n+1} sse a frase (1) não é verdadeira_n.

Nesse caso, também não seria possível derivar o paradoxo, como explica Soames:

Essa é uma instância correta do esquema T e é, portanto, aceitável. Seu lado direito é tanto verdadeiro_{n+1} quanto verdadeiro_{n+2}; seu lado esquerdo é verdadeiro_{n+2}. Como resultado, a frase como um todo é verdadeira_{n+2}. Com essa formulação, nós derivamos

C2. A frase (1) é verdadeira_{n+1} e a frase (1) não é verdadeira_n. (SOAMES, 1998, p. 152, tradução minha)

Ao derivar (C2), não há nenhuma contradição (logo, nenhum paradoxo), pois o predicado ‘verdadeiro_{n+1}’ é diferente do predicado ‘verdadeiro_n’. Seria como afirmar que a grama é verde e a grama não é dura. São predicados diferentes e independentes, portanto, a afirmação de um e a negação de outro são compatíveis.

¹⁵ Pelo mesmo raciocínio, poderíamos tomar a metalinguagem L* como uma linguagem-objeto, e adotar outra linguagem, L**, como a metalinguagem de L*, e concluir que L** não contém seu próprio predicado de verdade, e assim por diante. Ver Tarski (1944, p. 350, tradução minha): "Deveria ser notado que os termos 'linguagem-objeto' e 'metalinguagem' possuem apenas um significado relativo. Se, por exemplo, nós estamos interessados na noção de verdade aplicada a frases, não da nossa linguagem-objeto original, mas da própria metalinguagem, a última se torna automaticamente a linguagem-objeto de nossa discussão; e, para definir verdade para essa linguagem, nós temos que ir para uma nova metalinguagem – digamos, para uma metalinguagem de um nível superior. Dessa maneira, nós chegamos a toda uma hierarquia de linguagens."

¹⁶ Para uma explicação paralela, ver Soames (1998, p. 152).

1.5 Objeções à solução clássica

Embora, como vimos, o paradoxo do mentiroso seja efetivamente bloqueado pela solução hierárquica de Tarski, há algumas limitações oriundas de sua resposta. Em primeiro lugar, imaginemos a seguinte situação: João e Maria são amigos há muito tempo, o que ocasionou uma confiança mútua incondicional. Em certo dia, João e Maria são encontrados por Manuel dormindo em uma floresta sombria, cheios de doces e guloseimas. Manuel pergunta para João o que eles fizeram, e João, cansado demais para responder, afirma ‘Tudo que a Maria disser é verdade’, e volta a seu sono profundo.¹⁷

A dificuldade que a solução de Tarski apresenta, nesse caso, é que a frase ‘Tudo que a Maria disser é verdade’ abrevia a frase ‘Tudo que a Maria disser é verdade_n’, para algum n determinado. Isto é, a frase que João afirma pertence a um nível determinado. No entanto, quando João afirma que tudo que Maria disser é verdade – sem antecipar quais frases ela irá utilizar -, sempre há a *possibilidade* de Maria afirmar algo cujo nível é m , sendo $m > n$. No entanto, segundo a solução hierárquica de Tarski, isso não é possível, pois o predicado ‘é verdadeiro_n’ só pode ser predicado de frases cujo nível é inferior a n (SOAMES, 1998, p. 157).

Outra objeção possível à solução de Tarski envolve a noção de uma *totalidade completa* de verdades, em oposição a uma totalidade parcial. Na solução tarskiana, são barradas afirmações como ‘Nenhuma frase é verdadeira e falsa’, ‘Todas as frases verdadeiras são verdadeiras’, ou seja, não há a possibilidade de uma noção de verdade não especificada para um determinado nível. Como afirma Grim,

[...] uma hierarquia tarskiana efetivamente proíbe qualquer noção global de verdade ou conhecimento. Uma vez que 'verdade' é substituída por uma série infinitamente fragmentada de predicados de verdade, 'verdade₀', 'verdade₁', 'verdade₂', ... , por exemplo, não nos resta nenhuma maneira gramatical de sequer formular (40) ou quaisquer outras leis lógicas básicas.

Toda frase é verdadeira ou falsa. (40)

[...] Em uma hierarquia tarskiana, nenhuma das noções de onisciência, um conjunto de todas as verdades, ou um conjunto consistente maximal A correspondente ao mundo atual podem ser especificadas. (GRIM, 1991, p. 32, tradução minha)

¹⁷ Esse exemplo é uma variação de um exemplo oferecido por Soames, em que uma testemunha asse a confiabilidade de outra pessoa: "Por exemplo, alguém com confiança na integridade e competência de uma testemunha que foi diante do júri pode dizer

1. Toda frase dita pela testemunha ao júri é verdadeira." (SOAMES, 1998, p. 157, tradução minha).

Em resumo, a solução de Tarski consegue construir linguagens formais em que seja possível evitar a geração do paradoxo do mentiroso. No entanto, conforme argumentam Grim e defensores dessa segunda objeção, adotar tal hierarquia de linguagens implica abandonar, em princípio, a possibilidade de alcançar uma noção de verdade que possa ser aplicada a quaisquer níveis da hierarquia; em outras palavras, a resolução do paradoxo teria o custo de tornar a noção de uma totalidade completa de verdades um mero ideal irrealizável.

Uma terceira objeção foi um ponto caro a Kripke e a Priest, pois é a partir dessa consideração sobre a resposta tarskiana que ambos partem para formular suas próprias respostas. A objeção é que Tarski teria tratado muito ligeiramente do paradoxo do mentiroso na linguagem natural. Em seu artigo de 1933, a linguagem natural é considerada apenas no §1, com o objetivo de mostrar que *nessa* linguagem podemos gerar uma frase contraditória:

[...] é, presumivelmente, justamente essa universalidade da linguagem cotidiana que é a fonte primária das antinomias semânticas [...] Essas antinomias parecem oferecer uma prova de que toda linguagem que é universal no sentido acima, e para a qual as leis normais da lógica valem, deve ser inconsistente. (TARSKI, 1956[1933], p. 164, tradução minha)

Um ponto de acordo, que penso ser indisputável, é que a solução hierárquica de Tarski não pode funcionar para a linguagem natural, pois a linguagem natural claramente não possui uma hierarquia de tal tipo.¹⁸ Assim, o problema com a solução tarskiana é que ela descarta de maneira ligeira que seja possível evitar uma antinomia na linguagem natural. No entanto, essa deveria ser a maior preocupação (pelo menos do filósofo, não do lógico), pois isso implicaria que o uso do predicado-verdade na vida cotidiana é contraditório. Se essa fosse uma conclusão verdadeira, teríamos que considerar seriamente abandonar o uso desse termo. Seria possível, no entanto, manter uma linguagem sem que seja possível de qualquer modo usar o predicado-verdade?¹⁹

¹⁸ Ainda que eu pense ser indisputável, Soames procura solucionar o paradoxo na linguagem natural a partir de uma hierarquia dentro dela (SOAMES, 1998, p. 152). Essa solução, entretanto, parece-me artificial, pois na vida cotidiana utilizamos o predicado-verdade sem nos questionarmos quanto ao 'nível' da frase a que nos referirmos (por exemplo, quando uma mãe defensora de seu filho na escola afirma – sem saber ao certo aquilo que seu filho disse - "Meu filho não mente! Tudo que ele falou é verdade").

¹⁹ É importante ressaltar, neste ponto, que o fato do paradoxo do mentiroso surgir em linguagens formalizadas e na linguagem natural implica que qualquer solução geral do paradoxo deve manter seu ataque forte em ambos os fronts: se uma parte estiver deficiente, ainda há margem para afirmar que o paradoxo se sustenta em um determinado domínio. Ver Cardoso (2010, p. 21): "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."

É por conta dessa consequência que antes de afirmar a inconsistência da linguagem natural devemos buscar as alternativas de solução possíveis que possam preservá-la ao máximo sem desfigurá-la por completo.

1.6 Respostas às objeções

Em relação à primeira objeção, podemos imaginar uma resposta na seguinte linha: quando João afirma "Tudo que Maria disser é verdade", a expansão completa dessa frase é "Dado o conjunto de frases Γ que Maria afirmar, existe uma frase de nível $n+1$, tal que n é o nível maximal do conjunto Γ , que afirma "Tudo que Maria disser é verdadeiro $_{n+1}$ " ".

Em relação à segunda objeção, poderíamos responder o seguinte: embora não seja possível a noção de uma verdade *global*, é possível afirmar frases como "Para todo nível n , toda frase é verdadeira $_n$ ou falsa $_n$ ", "Para todo nível n , nenhuma frase é verdadeira $_n$ e falsa $_n$ ", ou seja, ainda que não possamos afirmar verdades *não-niveladas*, podemos afirmar verdades *em um nível indeterminado*. Existe, contudo, uma limitação acerca dessa resposta: ainda que possamos afirmar tais verdades, não é possível predicar a sua verdade. Seja a frase (i) "Para todo nível n , toda frase é verdadeira $_n$ ou falsa $_n$ ". O que aconteceria se tentássemos afirmar (ii) "'Para todo nível n , toda frase é verdadeira $_n$ ou falsa $_n$ " é verdadeira $_k$ "? Ora, como a frase (i) varia sobre todos os níveis, ela também fala sobre o nível k . Logo, não podemos predicar a verdade $_k$ de (i), mas sim a verdade $_{k+1}$. Pelo mesmo raciocínio, no entanto, a frase (i) também fala sobre o nível $k+1$, o que mostra que para qualquer i , não é possível predicar a verdade $_i$ de (i).

Em relação à terceira objeção, penso que Tarski, cuja preocupação maior era formular uma linguagem formal para as ciências dedutivas, responderia do seguinte modo: acho interessante a preocupação em resguardar a linguagem natural de antinomias, portanto, o esforço de buscar resolvê-las é elogiável. No entanto, não consigo conceber a linguagem natural a não ser da maneira como eu caracterizei, ou seja, como *universal* e logicamente clássica. Dadas essas duas características, não há como não concluirmos que a linguagem natural – se for possível formalizá-la - é inconsistente (TARSKI, 1956[1933], p. 165).

2 A Solução Inortodoxa de Kripke

2.1 A crítica kripkeana à abordagem ortodoxa

Uma proposta de solução alternativa ao paradoxo do mentiroso foi formulada por Kripke em seu artigo *Outline of a theory of truth* (KRIPKE, 1975). Seu trabalho não apenas é interessante por sua resposta positiva, mas também por suas razões para rejeitar a solução tarskiana, que Kripke denomina de "abordagem ortodoxa":

Até agora, a única abordagem aos paradoxos semânticos que foi desenvolvida detalhadamente é a que denominarei de "abordagem ortodoxa", que leva às celebradas hierarquias de linguagens de Tarski. (KRIPKE, 1975, p. 694, tradução minha)

É importante ter essa observação em mente ao longo desta seção, pois a partir dela fica claro que Kripke pretende propor uma solução que se afaste ao máximo de uma resposta que envolva uma hierarquia de linguagens no sentido tarskiano, ou seja, uma hierarquia de linguagens semanticamente abertas.

Em primeiro lugar, no entanto, é importante distinguir entre dois tipos de solução que remetem ao trabalho de Tarski. Ao utilizarmos a expressão "solução tarskiana", a solução a que tal expressão se refere pode ser relativa a (a) uma hierarquia de linguagens semanticamente abertas em linguagens *formais* ou a (b) uma hierarquia de linguagens na linguagem *natural*. O trabalho de Tarski restringe-se a oferecer a resposta (a), pois, como visto, "o problema da consistência não possui um significado exato relativamente a essa linguagem" (TARSKI, 1944, p. 349). Isto é, seria errôneo atribuir ao trabalho de Tarski a tentativa de solucionar o paradoxo do mentiroso, por exemplo, por meio da construção de uma hierarquia de frases do português.

Tendo feito essa ressalva, a opção (b) ainda pode ser formulada. O suposto defensor de (b) teria que assumir que as expressões da linguagem natural são suficientemente diferenciadas das não-expressões, e que cada expressão possui um nível determinado. Não obstante, concedendo essas suposições, ele poderia ao menos desenvolver sua solução sem ser refutado em princípio, como fez Tarski ao considerar tal posição.

Contrariamente a Tarski, cuja preocupação inicial era desenvolver uma linguagem formalizada para as ciências dedutivas, Kripke pretende que sua resposta possa ser aplicada à linguagem natural, ou seja, de uma perspectiva kripkeana, é necessário que a linguagem construída possua as características da linguagem natural, caso contrário, a resposta ao

paradoxo do mentiroso terá errado o alvo. Ao fazer uma objeção à abordagem ortodoxa, Kripke deixa claro esse objetivo:

[...] **no**ssa linguagem contém apenas uma palavra 'verdadeiro', não uma sequência de frases distintas 'verdadeiro_n', aplicadas a frases de níveis maiores e maiores. (KRIPKE, 1975, p. 695, tradução minha, grifo meu)

Com base nessa estratégia, Kripke consegue descartar a abordagem ortodoxa, pois tal resposta não se adequa materialmente às condições da linguagem natural. A objeção propriamente dita captura uma 'intuição' que é perdida na construção ortodoxa, pois cada nível $n+1$ da hierarquia contém um predicado 'verdadeiro_n' cuja extensão são as frases verdadeiras do nível n , ou seja, cada 'verdadeiro_i' ($0 \leq i$) possui uma extensão única, o que significa que há uma equivocidade do significado de 'verdadeiro'. Isto é, ao afirmarmos uma frase simples como 'O que ele disse é verdadeiro', a frase torna-se ambígua pelo fato de que 'verdadeiro' é uma abreviação de 'verdadeiro_i' para algum i . Até mesmo um defensor da abordagem ortodoxa concede tal ponto, como mostra Soames:

[...] a ideia hierárquica também pode ser aplicada a linguagens naturais como um diagnóstico dos paradoxos a que eles dão origem. [...] Em vez de conter uma única palavra (inequívoca) *verdadeira* (aplicada a frases), o inglês é concebido como contendo uma hierarquia infinita de predicados de verdade (de frases): *verdadeira₀*, *verdadeira₁*, *verdadeira₂*, e assim por diante. Além disso, a própria língua inglesa é tratada como uma hierarquia infinita de linguagens:

E_0 : Língua inglesa menos predicados semânticos

E_1 : E_0 mais *verdadeira₀*, onde *verdadeira₀* se aplica a todas as verdades de E_0 e não se aplica a nada mais.

E_2 : E_1 mais *verdadeira₁*, onde *verdadeira₁* se aplica a todas as verdades de E_1 e não se aplica a nada mais, e assim por diante. (SOAMES, 1998, p. 152, tradução minha)

Por si só, tal objeção não se mostra derradeira, pois se poderia responder que a linguagem natural, *embora não aparente*, possui uma hierarquia de predicados semânticos, e a solução ortodoxa explicita tal característica. No entanto, outros problemas surgem que tornam uma hierarquia infinita de linguagens problemática.

Para a solução hierárquica funcionar, deveria ser possível destacar da linguagem *apenas* as frases semânticas paradoxais, ou seja, a inaplicabilidade do predicado de verdade a certas expressões deveria se restringir àquelas frases que geram contradições (no presente caso, às frases que afirmam sua própria falsidade). No entanto, a solução de Tarski bloqueia certas frases que são perfeitamente aceitáveis na linguagem natural, pois elas não geram paradoxos. Por exemplo, o caso do 'contador de verdades' (em inglês, *truth teller*) é um caso de autorreferência aplicada ao predicado verdade:

(1) 1 é verdadeira.

Será que (1) gera algum paradoxo? Suponhamos que (1) seja verdadeira. Nesse caso, o (1) afirma é o caso, ou seja, (1) é verdadeira. Suponhamos, por outro lado, que (1) seja falsa. Nesse caso, é falso que (1) é verdadeira, ou seja, (1) é falsa. Portanto, na linguagem natural, tal frase não gera nenhuma contradição, apesar de não ser possível determinar em que situações (1) seria verdadeira ou falsa (KRIPKE, 1975, p. 693).

Também não é possível formular frases que seriam expressamente falsas:

(2) O papai Noel existe e (2) é falsa.

Nesse caso, o problema é que a segunda parte da conjunção ('(2) é falsa') deveria ser uma expressão de nível superior ao de (2). No entanto, a segunda parte da conjunção é parte de (2), ou seja, não há como o nível de parte de (2) ser superior ao de (2).²⁰

Por fim, Kripke oferece um contraexemplo forte contra a teoria tarskiana. Apresento uma versão modificada do exemplo para ilustrar de maneira mais simplificada seu ponto.²¹ Suponhamos que somos juízes trabalhando sobre um caso de homicídio cujas únicas evidências são os testemunhos de João e Pedro. João e Pedro são guardas noturnos do prédio que fica ao lado do local onde ocorreu o suposto assassinato. João, não confiando em Pedro, afirma, dentre outras coisas, 'Tudo que Pedro disser em seu testemunho é falso'. Ao ouvir tal afronta de seu colega de serviço, no entanto, Pedro acrescenta a seu próprio testemunho 'E tudo que meu colega João disser em seu testemunho é falso'.

Nós, enquanto juízes, devemos buscar uma versão coerente dos fatos. Assim, depois de pensar algum tempo sobre como encaixar os testemunhos, chegamos à seguinte conclusão: se a afirmação de João for falsa, então nem tudo que Pedro afirma é falso, ou seja, pode ser o caso que a afirmação subsequente de Pedro seja verdadeira. Supondo que tal afirmação seja verdadeira, então o testemunho de João é completamente inválido, pois tudo que ele afirmou é falso. Ou poderia igualmente ser o caso que a afirmação de Pedro é falsa, caso no qual, se a afirmação de João fosse verdadeira, o testemunho de Pedro se invalidaria. Por outro lado, poderia ocorrer que algumas coisas que João disse são verdadeiras, assim como que algumas coisas que Pedro disse são verdadeiras. Nesse caso, a afirmação de João sobre o testemunho de Pedro seria falsa, assim como a afirmação de Pedro sobre João.

²⁰ O nível de uma conjunção (A e B) pode ser calculado do seguinte modo: nível (A e B) = max {nível A, nível B}, ou seja, se a expressão '(2) é falsa' possuir nível n , então nível (2) = max {nível 'o papai Noel existe', n }. Supondo que o nível de 'o papai Noel existe' seja m , nível (2) = max { m , n }. Se $m \leq n$, então nível (2) = n , ou seja, a expressão '(2) é falsa' não possui nível superior a (2). Por outro lado, se $m > n$, então nível (2) = m , caso no qual '(2) é falsa' também não possui nível superior a (2). Uma definição recursiva do grau de uma frase pode ser conferida em Gupta (1982, p. 11).

²¹ Para conferir o exemplo originalmente exposto, ver Kripke (1975, p. 695 – 6).

Em resumo, o juiz de tal caso, ao ouvir tais testemunhos, concluiria que uma dessas possibilidades é verdadeira: (i) Pedro está falando a verdade, e o testemunho inteiro de João é inválido, ou (ii) João está falando a verdade, e o testemunho inteiro de Pedro é inválido, ou (iii) as acusações de João e Pedro entre si são ambas falsas. Embora o juiz não consiga determinar *a priori* qual dentre essas opções é a verdadeira, ele ao menos pode concluir que há alguma atribuição consistente de valores de verdade aos testemunhos de Pedro e João, ou seja, tais afirmações não geram paradoxos em qualquer situação (apenas no caso em que ambas fossem verdadeiras).

Um juiz tarskiano, por outro lado, faria o seguinte raciocínio: suponhamos que a frase afirmada por João tenha nível n . Nesse caso, a frase de Pedro, envolvendo o valor de verdade da frase de João, teria que possuir um valor m , tal que $m > n$. Por outro lado, como a frase de João envolve o valor de verdade da frase de Pedro, a frase de João deve possuir um valor n , tal que $n > m$, o que é absurdo, pois nenhum número é maior e menor do que outro número. O juiz tarskiano, dessa maneira, enredar-se-ia em tal paradoxo até perder sua paciência e arquivar o caso como um de seus *insolubilia*.

2.2 As condições para uma teoria da verdade

Concomitantemente à exposição da crítica à abordagem ortodoxa, Kripke oferece uma série de características que uma teoria da verdade deve possuir para ser adequada em relação ao comportamento do predicado de verdade na linguagem natural. Na linguagem natural, não podemos simplesmente bloquear todas as frases que contenham autorreferências semânticas, pois, como visto em (1) e (2), nem sempre tais frases geram paradoxos. No caso de não gerarem paradoxos, tais frases possuem um sentido específico e um valor de verdade determinado em alguns casos. Essa característica da linguagem seria sua 'arriscabilidade', ou seja, as frases precisam poder arriscar serem paradoxais, para serem bloqueadas apenas no caso positivo de gerarem contradições:

As versões do paradoxo do Mentiroso que usam predicados empíricos já apontam para um aspecto importante do problema: *muitas, provavelmente a maioria, de nossas afirmações comuns sobre verdade e falsidade tendem, se os fatos empíricos são extremamente desfavoráveis, a exibir características paradoxais.* (KRIPKE, 1975, p. 691, tradução minha)

Isto é, uma solução adequada para o paradoxo do mentiroso não pode bloquear irrestritamente o uso de expressões sobre verdade e falsidade na linguagem-objeto, pois isso

implicaria que não poderíamos afirmar muitas coisas que afirmamos cotidianamente sem a pretensão de sempre gerar frases paradoxais.

Como uma primeira construção da linguagem kripkeana, o autor foca no aspecto da arriscabilidade das frases da linguagem, que inclui tanto as frases paradoxais quanto as não paradoxais:²² Kripke exige a condição de que elas possam ter seu valor de verdade empiricamente determinado. Para isso, ele fornece a distinção entre frases *fundadas* e frases *infundadas*: de maneira resumida, frases *fundadas* são aquelas frases que, em última análise, podem remeter a uma frase que não contenha predicados de verdade. Frases *infundadas*, por outro lado, ou remetem a outras frases que contêm predicados de verdade ou elas próprias contêm seu próprio predicado de verdade. É possível esquematizar ainda mais a noção de *fundação*: por exemplo, seja uma sequência finita S de frases $\langle P_n, P_{n-1}, \dots, P_i, \dots, P_1 \rangle$, em que cada P_j ($1 < i \leq n$) é da forma $P_j: P_{j-1}$ é verdadeira. Queremos determinar se P_n é uma frase fundada ou infundada. Pela forma de P_n , a determinação de sua fundação depende de P_{n-1} , que, por sua vez, depende da determinação de P_{n-2} , e assim por diante. Por fim, a determinação de P_n depende de P_1 ser fundada ou não. Caso P_1 contenha a noção de verdade (por exemplo: P_1 afirma que ' P_1 é verdadeira'), então P_1 é infundada, caso no qual P_n é infundada. Caso P_1 não contenha a noção de verdade (por exemplo: P_1 afirma 'O cachorro late'), então P_1 é fundada, caso no qual P_n é fundada.

No caso em que uma frase é fundada, ela é verdadeira ou falsa. Quando é infundada, ela não é verdadeira nem falsa, ou seja, ela tem um valor indeterminado:

Em geral, se uma frase [...] afirma que (a maioria, algumas, todas, etc.) as frases de uma certa classe C são verdadeiras, seu valor de verdade pode ser determinado se os valores de verdade das frases da classe C são determinados. Se algumas dessas frases envolve a noção de verdade, seu valor de verdade deve, em troca, ser asserido ao remeter a *outras* frases, e assim por diante. Se, em última instância, esse processo termina em frases não mencionando o conceito de verdade, tais que o valor de verdade da frase original pode ser asserido, denominamos a frase original de *fundada*; de outro modo, *infundada*. (KRIPKE, 1975, p. 693, tradução minha)

Com base nessa ideia, Kripke esboça uma solução ao paradoxo do mentiroso alternativa à ortodoxa, cujo núcleo está contido na noção de *truth gaps*, ou seja, lacunas no valor de verdade de algumas frases da linguagem: no presente caso, uma frase que afirme a falsidade de si mesma não é nem verdadeira nem falsa, pois ela é infundada.

²² Isto é, neste primeiro momento, Kripke enfatiza as *semelhanças* entre frases do mentiroso e frases do contador honesto: "Há tempos, foi reconhecido que algo do problema intuitivo com frases do Mentiroso (*Liar sentences*) é compartilhado com frases como

(3) (3) é verdadeira." (KRIPKE, 1975, p. 693).

2.3 A solução kripkeana

Como visto na seção anterior, a solução de Kripke deve ser capaz de exibir uma linguagem com *truth gaps*. Para isso, é preciso fazer uma distinção entre uma frase e uma proposição: por um lado, frases são tipos de sequências de palavras que são tentativas de expressar algum conteúdo. Por outro lado, proposições são os conteúdos declarativos efetivo das frases, logo, proposições sempre serão ou verdadeiras ou falsas. Em outras palavras, podemos dizer que frases são um 'esqueleto' de proposições possíveis, pois para cada situação elas podem ser preenchidas – ou não – por alguma proposição.

Uma sentença [frase] é um conjunto bem formado de expressões que utilizamos (enunciamos) em determinados contextos a fim de expressar uma proposição. [...] A proposição é o que é expresso por uma sentença. [...] 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. (CARDOSO, 2010, p. 63)

Tendo feito esta distinção, Kripke pode sustentar a tese fraca segundo a qual (i) nem toda frase expressa uma proposição em todos os contextos, diferenciando-a da tese forte que afirma (ii) nem toda proposição é verdadeira ou falsa. Pela tese (ii), Kripke deveria explicar como é possível que um estado de coisas nem ocorra nem não ocorra. Por (i), a questão pode permanecer em um nível linguístico.

Essa tese de Kripke, segundo a qual frases podem falhar em expressar alguma proposição em determinada situação, mas manter seu significado, não é trivial. O autor remete a defesa dessa tese a Strawson²³. Não obstante, o debate ainda é disputado por pelo menos outra posição, que toma forma no artigo *On Denoting*, de Bertrand Russell.

Segundo Strawson, certas frases pressupõem a verdade de outras, no sentido de que o sucesso em expressar algum conteúdo depende de tais pressuposições serem verdadeiras. Por exemplo, a frase 'O rei da França é careca' pressupõe que haja um rei da França, isto é, que uma frase como 'A França possui um rei' seja verdadeira. Caso não haja um rei da França, a frase original, que pressupunha a existência de um rei, não é nem verdadeira nem falsa.

A frase, "O rei da França é sábio", é certamente significativa; contudo, isso não significa que qualquer uso particular dela é verdadeiro ou falso. [...] Então, quando nós afirmamos a frase sem, de fato, mencionar ninguém pelo uso da expressão "o rei da França", a frase não deixa de ser significativa: simplesmente *falhamos* em dizer algo verdadeiro ou falso [...] (STRAWSON, 1950, p. 330 – 1, tradução minha)

Russell, por outro lado, analisa a frase "O rei da França é G" de uma maneira diferente. A ideia de Russell é que a forma gramatical dessa expressão não reflete a forma lógica:

²³ Ver Kripke (1975, p. 699): "Seguindo a literatura mencionada acima, propomos investigar linguagens com lacunas no valor de verdade. Sob a influência de Strawson, caracterizamos uma frase como uma tentativa de elaborar um enunciado, expressar uma proposição, ou algo similar".

enquanto a primeira revela uma forma sujeito-predicado simples, a última é, de fato, uma frase da forma existencial. Sendo assim, quando se afirma "O rei da França é G", tal expressão equivale logicamente à expressão "Existe um e apenas um rei da França, e o rei da França é G", ou seja, no caso de não haver rei da França, a frase é *falsa* (e não sem valor de verdade), pois um dos conjuntos é falso:

Se "C" é uma expressão denotativa [...], então "C possui a propriedade Φ " significa "um e apenas um termo tem a propriedade F, e esse um tem a propriedade Φ ". Se, agora, a propriedade F pertence a nenhum termo, ou a vários, segue-se que "C possui a propriedade Φ " é falso para *todos* os valores de Φ . (RUSSELL, 1905, p. 490, tradução minha)

Embora tal discussão seja interessante por seus próprios méritos, apenas a menciono como um lembrete de que a teoria da verdade de Kripke não é, de início, imune a objeções: poderíamos desenvolver um argumento que defendesse a teoria das descrições de Russell em detrimento da teoria de Strawson. Nesse momento, no entanto, pretendo reconstruir a teoria kripkeana, portanto, tratamos frases como "O rei da França é calvo" como um caso de frases nem verdadeiras nem falsas, assim como adotamos a distinção entre frases significativas e proposições.

Tendo aceitado uma versão da teoria strawsoniana, Kripke esboça sua solução à construção do predicado de verdade. Sua ideia é construir uma hierarquia de linguagens em que, a cada elevação de nível, o predicado 'é verdadeiro' abarque em sua extensão todas as frases verdadeiras dos níveis precedentes. Até aqui, a solução soa tarskiana. No entanto, essa construção não se estende *ad infinitum*, pois há um ponto fixo na hierarquia, a partir do qual os níveis superiores não acrescentarão necessariamente outras frases à extensão do predicado 'é verdadeiro'. Nesse ponto fixo, a linguagem é semanticamente fechada, embora não seja clássica de maneira restrita.

A linguagem que adotaremos para ilustrar essa construção, conforme a sugestão de Michael Kremer (1988, p. 229), é a lógica de primeira ordem com a adição do predicado 'T', cuja extensão são as fórmulas verdadeiras, além da adição de aspas (''), que podem formar nomes de todas as fórmulas da linguagem (nome de A = 'A').

Nessa linguagem, os termos lógicos usuais ('~', ' \wedge ', ' \rightarrow ', ' \vee ') permanecem operando do mesmo modo da lógica clássica. Dado um domínio D de objetos, e uma função interpretação I dos termos de D , construímos um modelo $\langle D, I \rangle$, de modo que, para cada constante individual, podemos relacionar um elemento de D via função I .²⁴ A diferença do sistema de

²⁴ Ver Kremer (1988, p. 230, tradução minha): "Agora, suponhamos que seja dada uma interpretação clássica de todas as constantes não-lógicas de L , exceto T . Isto é, é dado um par $\langle D, I \rangle$, em que D é um conjunto não-vazio, e I relaciona uma interpretação clássica para cada constante não-lógica de L – se a é uma constante

Kripke em relação ao sistema clássico é que, no caso da lógica clássica, existe uma função que associa a cada fórmula um e apenas um elemento do conjunto $\{V, F\}$. No caso do sistema de Kripke, existe uma função que associa a cada fórmula um e apenas um elemento do conjunto $\{V, F, \#\}$, em que " $\#$ " é um valor de verdade indeterminado:

A ferramenta que Kripke utiliza [...] é a lógica trivalorativa: o predicado verdade possui uma interpretação com três valores (ou parcial). [...] Assim, um modelo de L será uma tripla $\langle D, I, X \rangle$, em que $\langle D, I \rangle$ é uma interpretação clássica de todas as constantes não lógicas de L exceto T [...] e X é uma função de D ao conjunto de três elementos $\{t, f, *\}$. (KREMER, 1988, p. 230, tradução minha)

Conforme Kripke explicara anteriormente, algumas frases são infundadas, ou seja, não possuem valor de verdade determinado (nem verdadeiro nem falso). Isso implica que seu sistema de lógica não pode ser estritamente clássico, no sentido de haver uma função Y *completa* – que associe a cada fórmula um elemento de $\{V, F\}$. O que a lógica trivalorativa captura é a ideia de uma função *parcial*: a divisão entre as fórmulas verdadeiras e as falsas não é exaustiva. Além disso, como estamos tratando da lógica de primeira ordem, se soubermos o valor de verdade das fórmulas elementares, é possível determinar o valor de verdade das outras fórmulas. Kripke sugere utilizar o sistema trivalorativo forte de Kleene (KRIPKE, 1975, p. 700), em que a lógica clássica é preservada nas situações em que é possível decidir o valor de verdade das fórmulas com base nos valores clássicos, e que o valor resultante se mantém indeterminado quando não é possível decidir o valor de verdade das fórmulas com base nos valores clássicos (KLEENE, 1980, p. 334).²⁵

Sejam α, β fórmulas de L. Nesse caso,

- (a) $X(\sim\alpha) = V$ sse $X(\alpha) = F$; $X(\sim\alpha) = \#$ em qualquer outro caso.
- (b) $X(\alpha \vee \beta) = V$ sse $X(\alpha) = V$ ou $X(\beta) = V$; $X(\alpha \vee \beta) = F$ sse $X(\alpha) = F$ e $X(\beta) = F$; $X(\alpha \vee \beta) = \#$ em qualquer outro caso.
- (c) $X(\alpha \wedge \beta) = V$ sse $X(\alpha) = V$ e $X(\beta) = V$; $X(\alpha \wedge \beta) = F$ sse $X(\alpha) = F$ ou $X(\beta) = F$; $X(\alpha \wedge \beta) = \#$ em qualquer outro caso.
- (d) $X(\alpha \rightarrow \beta) = V$ sse $X(\alpha) = F$ ou $X(\beta) = V$; $X(\alpha \rightarrow \beta) = F$ sse $X(\alpha) = V$ e $X(\beta) = F$; $X(\alpha \rightarrow \beta) = \#$ em qualquer outro caso.

Assim, resta saber como determinar o valor de verdade das fórmulas elementares. A estratégia kripkeana, nesse ponto, é investigar de que modo as pessoas aprendem a distinguir as fórmulas verdadeiras de uma linguagem. Segundo o autor, nós começamos aprendendo frases simples como 'A banana é amarela', 'O copo é cilíndrico', para, então, afirmar "A

individual de L, então $I(a) \in D$, e se P é um símbolo de um predicado n-ário de L, então $I(P)$ é um subconjunto de D^n ."

²⁵ Outro sistema sugerido por Kripke, sobre o qual não entraremos em detalhes, é o sistema de Vann McGee (MCGEE, 1991).

banana é amarela' é verdadeiro", etc. Isto é, o uso do predicado 'é verdadeiro' depende de sabermos quais frases estamos autorizados a afirmar independentemente desse predicado:

Suponhamos que estejamos explicando a palavra "verdadeiro" para alguém que ainda não a entenda. Dizemos que estamos autorizados a asserir (ou negar) que uma frase qualquer é verdadeira precisamente sob as mesmas circunstâncias em que podemos asserir (ou negar) a própria frase. (KRIPKE, 1975, p. 701, tradução minha)

Portanto, as frases *fundadas* que contém o predicado verdade devem ser tais que elas são equivalentes a frases *fundadas* que não contém o predicado verdade. Ao avaliar se a frase "Tudo que Maria disse é verdadeiro", a *fundação* de tal frase deve ser avaliada empiricamente, isto é, devemos verificar o que Maria afirmou. Caso Maria tenha afirmado frases como "João comeu doces à noite", "João dormiu bastante", podemos avaliar se a frase anterior é verdadeira ou falsa.²⁶ Ao explicar a determinação dos valores de verdade de fórmulas contendo o predicado verdade, é possível estabelecer uma hierarquia de fórmulas. Por exemplo, a determinação da frase (P₁) "O que Maria disse sobre a afirmação de João é verdadeiro" depende da afirmação de Maria. Suponhamos que Maria tenha dito (P₂) "O que João afirmou é verdadeiro". Nesse caso, a afirmação de Maria depende da afirmação de João . Suponhamos que João tenha dito (P₃) "O pão está velho". A afirmação de João, por sua vez, não contém o predicado verdade, portanto, ela é decidível em seus próprios termos. A partir dessas frases, podemos estabelecer uma hierarquia, em que

L₀: todas as frases que não contém o predicado verdade [P₁]

L₁: todas as frases que não contém o predicado verdade + todas as frases cujo predicado verdade se aplica sobre frases de L₀ [P₁ + P₂]

L₂: todas as frases que não contém o predicado verdade + todas as frases cujo predicado verdade se aplica sobre frases de L₀ + todas as frases cujo predicado verdade se aplica sobre frases de L₁, e assim por diante [P₁ + P₂ + P₃].

As linguagens L_n são extensões de L₀. O que diferencia, por exemplo, L₀ de L₂? Em L₀, o predicado verdade é indefinido para todas as frases da linguagem, ou seja, a extensão de Tx é vazia. Em L₂, a extensão de Tx é {P₁, P₂}. Suponhamos que em um determinado nível, uma frase receba o valor verdadeiro. Nas extensões subsequentes, o seu valor de verdade permanece inalterado, pois a única coisa que as extensões fazem é acrescentar frases que contém o predicado verdade sobre as frases *já* estabelecidas como verdadeiras (ou falsas). Essa característica é denominada de *monotonicidade*.

²⁶ Ver Cardoso (2010, p. 65): "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, o interlocutor não está em condições de asserir nem de rejeitar que tal sentença é verdadeira."

A pergunta que se pode formular em seguida é sobre a expansão da extensão da linguagem inicial:

O processo vai adiante para sempre com mais e mais frases declaradas verdadeiras ou falsas, ou ele finalmente para? [...] Se novas frases de L fossem sendo decididas em cada nível, nós eventualmente esgotaríamos L em algum nível e seríamos incapazes de decidir além disso. (KRIPKE, 1975, p. 704)

Pode-se mostrar que em hierarquias desse tipo existem *pontos fixos*: em algum L_i , a extensão de L_i (L_{i+1}) possui o predicado verdade com a mesma extensão de L_i , ou seja, nesse nível a linguagem é semanticamente fechada (KRIPKE, 1975, p. 703 – 5).

A importância de tais pontos fixos é que a partir deles podemos avaliar o valor de verdade de algumas frases que envolvem predicados de verdade. O primeiro ponto fixo definido por Kripke é o ponto fixo *minimal*, que é o nível em que a extensão do predicado de verdade da linguagem inicial não aumenta mais.²⁷

Nesse ponto fixo minimal, todas as frases que poderiam receber algum valor de verdade com base no critério inicial de Kripke, de fato, recebem-no. É a partir desse nível que o autor define as noções de frases *fundadas* e *infundadas*: as frases fundadas são aquelas que recebem algum valor de verdade no ponto fixo minimal, ao passo que as infundadas são as que não recebem nenhum valor de verdade no ponto fixo minimal (KRIPKE, 1975, p. 706). Nesse panorama, o paradoxo do mentiroso não pode surgir, pois a frase do Mentiroso (*the Liar sentence*) é infundada: o processo de atribuição de valores de verdade nunca determinará um valor à frase do Mentiroso, pois seu predicado de verdade não remete a outra frase que não contenha o predicado de verdade.²⁸

2.4 Objeções a Kripke

Depois de formular a lógica paracompleta necessária para resolver o paradoxo do mentiroso na linguagem-objeto em questão, um momento de honestidade intelectual se destaca:

[...] as linguagens da presente abordagem contêm seus próprios predicados de verdade e até mesmo de satisfação, e, desse modo, o esperado foi realizado. No

²⁷ Este não é o único ponto fixo, pois poderíamos designar valores de verdade arbitrários a frases que ainda não os tenham recebido. No entanto, este é o ponto fixo minimal, pois é o ponto fixo a que se pode chegar com base apenas no critério da referência a frases que não contenham o predicado verdade.

²⁸ O ponto fixo minimal determina as frases fundadas; não obstante, Kripke determina outros pontos fixos, que são extensões do minimal: essas extensões são modelos consistentes de atribuição de valores de verdade às demais frases indeterminadas no ponto fixo minimal. Assim, o contador de verdades e a história de João e Pedro, ao contrário da solução tarskiana, recebem valores de verdade – ainda que arbitrariamente determinados – em pontos fixos além do minimal.

entanto, a presente abordagem certamente não pretende ser uma linguagem universal, e eu duvido que tal objetivo possa ser atingido. (KRIPKE, 1975, p. 714)

Em outras palavras, o conceito semântico de *verdade* e de *satisfação* podem ser aplicados irrestritamente na linguagem construída por Kripke. Entretanto, como veremos adiante, nem todos conceitos semânticos estão livres de uma restrição metalinguística.

A origem de tal restrição se encontra em uma distinção entre dois tipos de negação: a negação de *exclusão* e a negação de *escolha*. De maneira resumida, negar exclusivamente uma frase é o mesmo que afirmar que ela possui qualquer valor *exceto* o verdadeiro. Negar por escolha, por outro lado, é o mesmo que afirmar que a frase *possui* o valor falso.²⁹ Na linguagem-objeto paracompleta, Kripke admite apenas a negação por escolha: não há lacunas na negação exclusiva, pois negar exclusivamente A é afirmar que A é falso ou outro valor distinto do verdadeiro. Entretanto, analisemos o seguinte caso: ao analisar a construção de uma linguagem com lacunas, um observador comenta que "A frase do mentiroso não é verdadeira". Como na linguagem natural possuímos tal artifício de negar exclusivamente, poder-se-ia dizer que o observador está certo, pois a frase do mentiroso não se aplica nem à extensão nem à anti-extensão do predicado da verdade, ou seja, ela não é verdadeira nem falsa.

Nesta nova linguagem, a negação exclusiva está à disposição, ou seja, a metalinguagem não é paracompleta, o que a torna suscetível à geração de paradoxos, à maneira das linguagens tarskianas. Desse modo, a única maneira de evitar o surgimento de contradições *na* metalinguagem é adotando uma saída tarskiana, a partir da construção de uma hierarquia de linguagens. Nas palavras de Kripke:

A necessidade de ascender a uma metalinguagem talvez seja uma das fraquezas da presente teoria. O fantasma de Tarski ainda está conosco. (KRIPKE, 1975, p. 714, tradução minha)

Para Kripke, todavia, isso não significa que a construção de sua linguagem tenha sido um fracasso. Sua saída é caracterizar a linguagem natural constituída por diferentes *etapas*: alguns conceitos são introduzidos apenas em estágios mais desenvolvidos da linguagem, ou seja, a teoria de Kripke seria uma boa caracterização de um estágio inicial da linguagem, em que estaria disponível apenas o conceito de negação por escolha, sendo a negação exclusiva introduzida posteriormente (KRIPKE, 1975, p. 714).

²⁹ "Se interpretarmos '~' como uma negação exclusiva, ~A é verdadeiro sse A é falso ou indefinido, e ~A é falso sse A é verdadeiro. A negação exclusiva é contrastada com a *negação por escolha*. Se '~' é interpretado como negação por escolha, ~A é verdadeiro sse A é falso, ~A é falso sse A é verdadeiro, e ~A é indefinido sse A é indefinido. (SIMMONS, 2008, p. 53, tradução minha)

Keith Simmons, ao objetar contra Kripke, acertadamente localiza o problema do paradoxo do mentiroso:

"É, em geral, um erro ver a emergência de um paradoxo ter a ver essencialmente com a negação exclusiva: o que é essencial é a construção de um contra-valor."
(SIMMONS, 2008, p. 54, tradução minha).

Essa noção de contra-valor pode ser caracterizada do seguinte modo: seja $T(x)$ o predicado aplicado às frases de L_k . Um contra-valor de $T(x)$ é um predicado complementar ($C(x)$, por exemplo) de $T(x)$ que define a quais outras frases se aplica $C(x)$. Sem precisarmos avançar nossa linguagem ao estágio em que Kripke conceda que precisemos e ascender a uma metalinguagem, a partir dos recursos básicos da linguagem pré-filosófica, conseguimos gerar um contra-valor que seja equivalente à negação exclusiva, com o uso das noções '*falso*' e '*nem falso nem verdadeiro*', construindo $C(x)$ como sendo aplicado apenas às frases falsas ou nem falsas nem verdadeiras.³⁰

Essa noção [o contra-valor] é composta de noções cotidianas de verdade e falsidade, e a construção '*nem-nem*'. E essa noção possui aplicações claras e intuitivas – por exemplo, para frases sem significado ou sem sentido. (SIMMONS, 2008, p. 54, tradução minha).

Em síntese, a teoria de Kripke, por mais que consiga construir um predicado $T(x)$ que se aplique a todas as frases verdadeiras da linguagem-objeto - e, nesse sentido, tal linguagem é semanticamente fechada -, está ameaçada em duas frentes: em primeiro lugar, a extensão da linguagem-objeto, para admitir o conceito de negação exclusiva, exige que ascendamos a uma metalinguagem, e, em última instância, a adoção de uma hierarquia tarskiana que evite o surgimento de paradoxos. Além disso, mesmo que Kripke aceite tal limitação, ainda existe a ameaça de que possamos construir conceitos na linguagem-objeto que originem paradoxos. Neste caso, o problema é que a construção de tais noções exige apenas noções básicas pré-filosóficas, ou seja, Kripke deveria admiti-las em uma linguagem que pretende refletir um estágio básico da linguagem natural.³¹

³⁰ Para gerar contradições na linguagem de Kripke, Simmons utiliza o predicado '*x não é verdadeiro de si mesmo*', ou seja, na verdade ele está utilizando uma versão do predicado heterológico (SIMMONS, 2008, p. 54). Por razões de elucidação, utilizei o predicado $C(x)$.

³¹ Ver o diagnóstico de Simmons (2008, p. 55).

3 A solução deflacionista

Por fim, abordaremos uma solução que, em um aspecto muito básico, destoa das tentativas anteriores. Até o momento, foi suposto que quando alguém afirma algo como " 'A grama é verde' é verdadeiro", tal pessoa está afirmando algo com a estrutura semelhante a, por exemplo, "O céu é azul". Isto é, supõe-se que "é verdadeiro" atua como um predicado lógico, que atribui uma propriedade determinada à frase em questão. Contudo, existem registros, na história da filosofia³², de posições que não tratam "ser verdadeiro" como um predicado substancial, ou seja, não se está atribuindo alguma qualidade ao afirmar a verdade (ou falsidade) de uma frase, ao contrário do que acontece quando predicamos a vermelhidão da maçã, por exemplo. No contexto do paradoxo do mentiroso, essa posição evitaria o surgimento de uma antinomia ao cortar o mal pela raiz: a frase que afirma a sua própria falsidade não seria significativa, ou seja, tal expressão não contaria como uma fórmula bem formada da linguagem, ou o esquema-T não seria aplicável a tal frase.

3.1 A concepção clássica de verdade³³

Os deflacionistas surgem como uma resposta a uma tradição que posiciona o conceito de verdade como uma noção central para a filosofia, que deve ser melhor explicada e que serviria como explicação de muitos problemas filosóficos. Denominarei tal posição como a *concepção clássica* da verdade. Nesse contexto, a concepção clássica não é meramente a concepção correspondencialista da verdade, segundo a qual a verdade consiste em uma determinada relação entre o que é dito e o que há no mundo. Quero dizer, de maneira mais geral, que a concepção clássica é aquela que vê a noção de verdade como problemática, ou seja, que a verdade seja um conceito que precisa de uma definição que permita distinguir os membros que pertencem ao grupo de frases verdadeiras daqueles que não pertencem. Sob essa perspectiva, outras teorias da verdade como a pragmática, a coerentista, e a da

³² A posição deflacionista possui diversas ramificações: o minimalismo (HORWICH, 2010), a teoria da redundância (STRAWSON, 1950), a teoria da citação (QUINE, 1994), dentre outras.

³³ Na exposição introdutória que se segue, seguirei de perto a análise de Paul Horwich em seu livro *Truth, Meaning, Reality* (HORWICH, 2010).

demonstrabilidade, também se encaixam na concepção clássica (assim como a teoria correspondencialista):³⁴

Várias tentativas de *analisar* a verdade, de mostrar como ela é definida em termos de noções mais fundamentais, foram realizadas nos últimos dois mil anos ou mais. (HORWICH, 2010, p. 2)

A teoria deflacionista, por sua vez, não embarca no seguinte *projeto* clássico: há um problema, a saber, como a verdade deve ser definida, que deve seguir os parâmetros de uma definição adequada.³⁵ Assim, o que o deflacionismo propõe é a dissolução de tal projeto.

3.2 A concepção deflacionista da verdade

A história do deflacionismo na lógica moderna começa com Frege, em uma série de apontamentos perspicazes em seu texto *O Pensamento* (FREGE, 2002). Depois de descartar a teoria da verdade como correspondência com base no fato de que nunca será possível corresponder completamente frases a objetos,³⁶ Frege se pergunta se é possível de alguma maneira oferecer uma definição da verdade. Imaginemos que a resposta seja positiva, e tenhamos uma definição que diga que ser verdadeiro é possuir as características A, B, e C. Ao estabelecer que tais são as notas do conceito, deve-se perguntar se é verdadeiro que tais são as notas desse conceito, ou seja, nunca conseguimos eliminar completamente a circularidade em uma definição da verdade. Logo, a verdade é uma noção indefinível, pois toda definição adequada não é circular.

[...] numa definição adequada cumpre indicar certas notas características da verdade; e ao aplicá-la a um caso particular surgiria novamente a questão de se é verdadeiro que tais notas são constatadas. E assim nos moveríamos em círculo. (FREGE, 2002, p. 13-14)

Ainda que a verdade não seja definível, pode ser que a sua atribuição em determinadas frases possua um papel essencial, ou seja, pode ser que a noção de verdade possua uma função exclusiva na lógica das linguagens naturais e formais. Frege, no entanto, esforça-se em

³⁴ A teoria pragmática procura equacionar o ser verdadeiro ao ser útil; a teoria da demonstrabilidade, o ser verdadeiro ao ser demonstrável (capaz de prova); a teoria coerentista, o ser verdadeiro ao ser coerente com as demais frases (HORWICH, 2010, p.2-3). Nessa medida, todas essas teorias vão além de uma caracterização operacional da verdade, que apenas nos mostraria como e em que contextos utilizamos a expressão "é verdadeiro", e procuram encontrar um elemento que unifique todas as instâncias de frases verdadeiras, ou seja, é buscada uma qualidade comum às frases verdadeiras que seja diferente do fato (trivial) de todas elas serem verdadeiras.

³⁵ Em linhas gerais, uma definição deve ser capaz de capturar todos os membros sendo definidos, e apenas esses membros; ela não pode supor o termo que está sendo definido, direta ou indiretamente; etc.

³⁶ A argumentação de Frege pode ser resumida do seguinte modo: a noção de verdade não admite "meias-verdades": para algo ser verdadeiro, é exigido que seja completamente verdadeiro. Se verdade for correspondência, é exigida completa correspondência, ou seja, o que é dito precisa *coincidir* com o conteúdo do que é dito, o que é trivialmente absurdo (FREGE, 2002, p. 13).

mostrar como o conceito de verdade pode ser eliminado das frases usuais como " "O céu é azul" é verdadeiro", ou " "A grama é verde" é verdadeiro". Em última instância, aquilo que pretendemos comunicar com a afirmação de " "p" é verdadeiro" pode ser comunicado de maneira igualmente eficaz apenas ao afirmamos "p".

Expressamos o reconhecimento da verdade sob a forma de uma sentença assertiva. Para isto, não precisamos da palavra "verdadeiro". E mesmo quando dela fazemos uso, a força assertiva não se encontra propriamente nela, mas na forma da sentença assertiva, e quando esta perde sua força assertiva, a palavra "verdadeiro" não poderá restituí-la. (FREGE, 2002, p. 17-18)

Em outras palavras, para tomar uma determinada frase como verdadeira, é preciso apenas que tal frase esteja *sendo asserida* – que eu a esteja afirmando. Acrescentar o predicado gramatical "verdadeiro" não irá alterar nada: se eu não tomo a frase como verdadeira (como, por exemplo, quando faço uma pergunta), afirmar sua verdade não irá torná-la verdadeira; se a frase já é tomada como verdadeira, afirmar sua verdade é trivial.

Frank Ramsey também identificou precocemente o *insight* deflacionista segundo o qual não há nada de misterioso na noção de verdade:

[...] antes de procedermos com a análise do julgar, é necessário dizer algo sobre a verdade e a falsidade, para mostrar que não há nenhum problema separado da verdade, mas apenas uma confusão linguística. (RAMSEY, 1931, p. 128, tradução minha)

A estratégia deflacionista de Ramsey para mostrar que não há problema separado da verdade é analisar proposições que contenham o predicado gramatical "é verdadeiro" em proposições em que tal expressão não apareça. Dessa maneira, pode-se mostrar que o predicado da verdade não é um predicado lógico genuíno, mas apenas uma maneira diferente de formular proposições que não contenham a noção de verdade (RAMSEY, 1931, p. 128). Para ser bem sucedido em tal tarefa, Ramsey deve dar conta de explicar dois tipos de proposições que contêm o predicado gramatical "é verdadeiro":

- (i) 'p' é verdadeiro, para alguma proposição definida p
- (ii) [Para cada p_i de Γ / Algum p_i de Γ] é verdadeiro, para alguma proposição p_i (i e N)

Começemos por (i). Tomemos como exemplo a frase (a) " "O prédio é azul" é verdadeiro". É possível eliminar o predicado gramatical ser verdadeiro e obter uma frase com o mesmo significado de (a)? Segundo Ramsey, "é verdadeiro" é um verbo que pretende dotar a frase como um todo de uma *força assertiva*, ou seja, o verbo é responsável por reiterar que as coisas são de tal modo. No entanto, a frase "O prédio é azul" já contém um verbo, ou seja, ela já possui a força assertiva que o predicado gramatical da verdade apenas reitera. Logo, " 'p' é verdadeiro" pode ser substituído, simplesmente, por 'p'. Assim, afirmar que " "O prédio é azul" é verdadeiro" significa o mesmo que afirmar "O prédio é azul".

Em resumo, o suposto problema da verdade é dissolvido ao notarmos que "temos que adicionar [...] o verbo 'é verdadeiro' para fornecer um verbo à frase, esquecendo que 'p' já contém [...]" (RAMSEY, 1931, p. 128, tradução minha).

Ao tratar de proposições quantificadas, a estratégia é basicamente a mesma. Consideremos a frase "Tudo que Platão diz é verdadeiro", enunciada por um jovem seguidor fanático de tal filósofo. Aparentemente, o predicado "é verdadeiro" é ineliminável da frase, pois não conseguimos identificar exatamente sobre qual frase se trata a descrição "Tudo que Platão diz". Esquematizando tal frase, no entanto, obtemos (b) $\forall p$ (Platão afirma $p \rightarrow p$ é verdadeiro). Conforme visto em (i), " 'p' é verdadeiro" pode ser analisado como 'p'. Logo, (b) pode ser analisado em (b*) $\forall p$ (Platão afirma $p \rightarrow p$). (b*), por sua vez, não contém o predicado gramatical "é verdadeiro", logo, a análise de Ramsey mostra que esse predicado não é um predicado lógico genuíno, mas apenas uma maneira redundante de assegurar força assertiva a 'p'. Conforme afirma Ramsey, as frases contendo esse predicado "são expressões que utilizamos por razões de ênfase ou estilísticas [...]" (RAMSEY, 1931, p. 128, tradução minha).

Em síntese, o deflacionismo parte de uma ideia geral de desmistificar o conceito da verdade. Logo, é natural que surjam diferentes estratégias que procurem estabelecer esse objetivo. As principais versões dessa posição filosófica emergente são:

- (i) o *minimalismo*, cuja tese principal é que o significado de "verdade" é explicado inteiramente pelo uso das instâncias do esquema-T.³⁷
- (ii) a teoria da *variável sentencial*, cuja tese principal é que frases contendo o predicado gramatical "é verdadeiro" podem ser analisadas em frases sem tal predicado.³⁸
- (iii) a teoria da citação, que enfatiza o fato de "p" e " 'p' é verdadeiro" diferirem apenas pelo fato de "p" ser asserido no primeiro caso e mencionado no último.³⁹

Horwich também cita a teoria da redundância e a teoria prosentencial (HORWICH, 2010, p. 19).

3.3 A solução de inspiração ramseyana

³⁷ HORWICH, 2010.

³⁸ RAMSEY, 1931.

³⁹ QUINE, 1994.

Por fim, desenvolverei uma resposta pouco explorada no contexto deflacionista, mas que possui bastante relevância. Comumente, uma das primeiras respostas oferecidas ao paradoxo do mentiroso é que a frase que nega sua própria verdade não tem sentido, ou seja, a expressão (a) "Esta frase é falsa" não possui significado. Caso sejamos bem sucedidos em defender que tal expressão não possui significado, não é possível derivar uma contradição da maneira como vínhamos fazendo, pois (i) não é possível supor que (a) seja verdadeira ou falsa; (ii) não é possível perguntar pelas condições de verdade de (a); (iii) não é possível aplicar o esquema-T a (a). Em síntese, todos os ingredientes para a formação do paradoxo são excluídos.

No entanto, essa solução é pouco atrativa para muitos filósofos pelo fato de que não parece haver uma razão forte para rejeitar a significatividade de (a). Afinal, raciocinamos de maneira pedestre para derivar o paradoxo. Além disso, a versão do mentiroso contingente mostra que em determinadas circunstâncias, uma mesma expressão pode gerar um paradoxo ou não, dependendo de circunstâncias empíricas. A significatividade de uma frase depende de circunstâncias empíricas? Se, por outro lado, argumentarmos que (a) não é significativa por conta de ela gerar um paradoxo, caímos em um círculo vicioso, pois buscávamos uma solução que evitasse o paradoxo, e a justificativa da solução é que nossa saída evita o paradoxo, ou seja, essa solução é *ad hoc* (BEALL, 2001, p. 126).

O deflacionismo pode oferecer uma luz ao final do túnel para tal estratégia. Em especial, procurarei mostrar que a teoria ramseyana pode oferecer uma razão satisfatória para defendermos que (a) não possui significado.

Conforme visto na seção 5.2, a teoria ramseyana defende que o predicado "é verdadeiro" não é um predicado lógico genuíno, pois todas as frases que o contém podem ser analisadas em frases que não contém tal predicado. Assim, por exemplo, a frase "A inteligência artificial dominará o mundo" é verdadeiro" é analisada como significando simplesmente "A inteligência artificial dominará o mundo". Generalizando tal resultado, toda frase contendo o predicado "é verdadeiro" deve ser capaz de ser analisada em uma frase que não contém tal predicado. Conforme afirma Beall:

O ponto básico é que o deflacionismo, mais que qualquer outra teoria da verdade, está em condição de – se não obrigado, talvez – defender que 'verdade' é *em princípio eliminável* de qualquer frase significativa. (BEALL, 2001, p. 127, tradução minha)

Utilizando a sugestão de Beall, o seguinte princípio norteia a estratégia em questão: uma frase *p* contendo o predicado "é verdadeiro" só é significativa caso o predicado "é verdadeiro" seja *eliminável* de *p*. Formulando de maneira mais clara: seja *p* uma expressão

contendo o predicado "é verdadeiro". Nesse caso, vale que (b) "é verdadeiro" é *eliminável* de p se, e somente se, p é uma frase significativa.

A partir de (b), podemos derivar (b') Se "é verdadeiro" não é eliminável de p , então p não é uma frase significativa. Logo, a partir desse princípio da eliminabilidade, existe uma razão independente para defender que a frase (a) não é significativa, pois as condições de verdade de (a) dependem apenas de (a), que é, por sua vez, uma frase contendo o predicado "é verdadeiro", ou seja, "é verdadeiro" não é eliminável de (a). Portanto, por (b'), (a) não é uma frase significativa.

3.4 Críticas ao deflacionismo

A resposta de inspiração ramseyana (denominêmo-la saída R) é uma novidade em relação às posições anteriores. Conforme visto no capítulo 2, as respostas ao paradoxo do mentiroso estão sujeitas a vinganças, que são reformulações do paradoxo formuladas para se esquivar da resposta oferecida. A saída R consiste em defender que frases contendo predicados da verdade que não podem ser eliminados não possuem significado, ou seja, a frase do mentiroso, (p) "p é falsa", não possui significado.

Para construir a vingança para essa resposta, basta construir uma disjunção em que um dos disjuntos seja "p é falsa" e o outro seja nossa resposta, ou seja, teremos a frase (k) "k é falsa ou k não possui significado", que também é paradoxal.

Apesar disso, creio que há uma resposta satisfatória a esse tipo de objeção. Primordialmente, observemos o seguinte: dada uma frase da forma α ou β , é razoável supor a seguinte regra:

(C) α ou β possui significado sse α possui significado e β possui significado

Isto é, uma disjunção só possui significado se ambos os disjuntos são significativos. Como o primeiro disjunto de (k) não possui significado, segue-se, por (C), que (k) não possui significado. Logo, não é possível perguntar nem pelas condições de verdade, nem pelas condições de significado de (k), ou seja, não é possível erigir o raciocínio que deriva uma contradição.⁴⁰

⁴⁰ Alguém poderia replicar: você diz que essa frase não possui significado, mas eu consigo raciocinar perfeitamente utilizando esta frase como base. Para essa pessoa, vale lembrar que há uma distinção entre a forma lógica *aparente* (a forma gramatical) das frases e a forma lógica genuína. Quando se afirma que uma expressão não possui significado, significa que a frase não possui uma forma lógica genuína (a análise de Ramsey das frases contendo o predicado "é verdade" mostra que a forma lógica genuína de tais frases não contém tal predicado).

A segunda objeção possível parte do mentiroso contingente. Seja a frase (k') "A última frase do caderno sobre a mesa é falsa". Nesse caso, (k') ser paradoxal depende de um fator empírico – a saber, qual é a última frase escrita no caderno sobre a mesa. Suponhamos que a frase em questão seja "Os pinguins podem voar". Como os pinguins não podem voar, (k') é verdadeira, ou seja, (k') possui um significado determinado. No entanto, suponhamos a última frase no caderno seja "Esta frase é falsa". Pelo princípio da eliminabilidade, a frase do mentiroso não possui significado, logo, (k') também não passa pelo teste da eliminabilidade, pois a frase a que (k') se refere não trata de fatos não contendo a noção de verdade, ou seja, (k') não possui significado.

Em resumo, em algumas situações, (k') possuirá significado, em outras, não. No entanto, se dissermos que (k') não possui significado, então isso significa que não é possível perguntar pelas condições de verdade e de significado de (k'). No entanto, se esse fosse o caso, como seria possível possuir uma regra que nos guiasse de (k') até a última frase contida no caderno sobre a mesa? Em outras palavras, se (k') é um mero balbuciar de letras, como esse balbuciar nos guiaria acuradamente para outra frase, como é característico de uma regra?

(k'), no entanto, ainda que não possua significado enquanto uma frase como um todo, possui um componente que é significativo: a descrição definida "a última frase do caderno sobre a mesa". A significatividade dessa descrição não é dependente da significatividade da frase, e a própria descrição que funciona como uma 'regra' para encontrar a frase do mentiroso contingente. Assim, não há problema em aceitar que (k') não possui significado e aceitar que a descrição "a última frase do caderno sobre a mesa" possui significado – ou seja, aceitar que ela faça referência a alguma frase.

Em síntese, ao utilizar uma estratégia de inspiração ramseyana, pretende-se que toda frase contendo "é verdadeiro" seja analisável em uma fórmula que torne tal expressão desnecessária:

Nós temos que adicionar 'é verdadeiro' para dar um verbo a uma frase, esquecendo que 'p' já contém um verbo (variável). [...] Quando todas as formas das fórmulas estão incluídas a análise é mais complicada, mas não essencialmente diferente; e fica claro que o problema não é a natureza do falso ou do verdadeiro [...] (RAMSEY, 1931, p. 128-9)

Em outras palavras, a forma lógica das frases contendo o predicado "é verdadeiro" não reflete sua forma de superfície, a mera forma gramatical. Portanto, uma frase como "A frase escrita no quadro negro da sala 101 é verdadeira" pode ser enganosa: talvez tal expressão, da maneira como *aparece* gramaticalmente, leve-nos a achar que tal expressão é uma fórmula completa. Conforme tentarei argumentar, tal frase exige uma análise peculiar.

A análise que se segue é uma tentativa de aplicar a teoria das descrições de Russell ao caso de uma frase do tipo mentiroso contingente.⁴¹ Pois bem, analisemos a frase (i) a frase escrita no quadro da sala 101 é verdadeira. Segundo a teoria russelliana, a forma lógica de tal expressão é complexa, pois ela assume a existência de uma frase com uma certa propriedade – a saber, estar escrita no quadro da sala 101 - e atribui a verdade a tal frase. Seja R^1 o predicado unário que varia sobre o domínio das *fórmulas* da linguagem, que equivale a 'x está escrito no quadro da sala 101', seja p um nome para a frase referida pela descrição definida, e seja E^*x um quantificador existencial que quantifica sobre o domínio das *fórmulas* da linguagem (não dos objetos). Nesse caso, (i) pode ser analisada como (i') $E^*!x (R^1x \wedge x = p \wedge "p" \text{ é verdadeira})$. Conforme a teoria ramseyana, porém, "p" é verdadeira significa o mesmo que "p" (RAMSEY, 1931, p. 128), ou seja, (i') pode ser analisada como (i'') $E^*!x (R^1x \wedge x = p \wedge p)$.

Através da análise lógica de (i), fica clara a razão do mentiroso contingente ser sem sentido *em alguns casos* e em outros não: a análise completa mostra que tal frase é complexa, afirmando que existe uma única frase x tal que x tem a propriedade F , sendo x igual a p , e p . Caso não haja uma tal frase (ou seja, o nome p não representa nada), a frase não tem sentido, pois uma conjunção só é significativa caso todos seus conjuntos sejam significativos. Como no caso de falhas de referência, quando a referência de uma descrição definida aponta para a frase do mentiroso, a falta de sentido da última torna a *fortiori* a primeira sem sentido.

Tal raciocínio aparenta ser circular. Aparentemente, eu queria mostrar que a frase do mentiroso era sem sentido, mas, no final das contas, eu assumi que tal frase do mentiroso era sem sentido. É assim que Armour-Garb argumentaria, ao menos.

No entanto, deixe-me esclarecer um ponto: não estou desprezando o fato de chamar o princípio da eliminabilidade de *princípio*. Enquanto pedra fundamental para erguer o deflacionismo (ao menos a concepção que estou tentando defender), o princípio da eliminabilidade não pode ser provado sem circularidade. Ao invés de tentarmos prová-lo, mais eficaz seria usarmos-lo como princípio norteador das investigações, buscando analisar as frases contendo "é verdadeiro". Se nos depararmos com alguma frase que se mostre anômala a tal tratamento, então teremos razão para duvidar da eliminabilidade.

Portanto, meu intuito ao responder a objeção de Armour-Garb foi mostrar que é possível mostrar que frases do tipo mentiroso contingente são sem sentido *em alguns casos*, sem que tal resposta recaia em armadilhas kripkeanas.

⁴¹ A análise a seguir é inspirada no *modo* de análise russelliana. Isso não significa que sua *conclusão* é equivalente à russelliana.

O teste final à teoria eliminativista diz respeito à ameaça de inexpressabilidade. Pelo princípio da eliminabilidade, uma frase contendo a noção de verdade só possui sentido se ela for analisável em termos que não envolvam essa noção. Digamos que tenhamos convencido alguém que a frase do mentiroso não possui significado. Procurando entender tal resultado, o sujeito realiza algumas inferências. Entre elas, ele afirma que "a frase do mentiroso não é verdadeira nem falsa – pois ela não possui significado". Como analisaremos tal frase?

Consideremos uma situação menos problemática primeiro. Frente à frase "Macacos são divertidos", uma frase significativa que denominaremos p , afirmamos que " 'p' não é verdadeira nem falsa". Como analisar *essa* frase? Não é difícil: " 'p' não é verdadeira nem falsa" equivale a " 'p' não é verdadeira e 'p' não é falsa".

Estipulemos que estejamos utilizando a negação por escolha. Nesse caso, podemos concluir a análise como ' $\sim p \wedge \sim\sim p$ ', que classicamente se configura como ' $\sim p \wedge p$ ', o que é falso, visto que apenas 'p' é verdadeiro. Seria errôneo, contudo, atribuir a negação por escolha a tal afirmação, pois isso é melhor interpretado como afirmando " 'p' possui outro valor de verdade além do verdadeiro e do falso". Chamemos de ' \sim^* ' a negação exclusiva. Nesse caso, a frase é analisada como ' $\sim^*p \wedge \sim^*\sim p$ '. Ela ainda continua falsa, pois 'p' é verdadeiro, ou seja, ' \sim^*p ' é falso.

Assim, o sujeito que antes inferiu "a frase do mentiroso não é verdadeira nem falsa" estaria afirmando ' $\sim^*s \wedge \sim^*\sim s$ ', sendo 's' uma abreviação da frase do mentiroso. Ainda que tal frase pareça significativa e verdadeira, como a frase do mentiroso não possui significado, não é possível realizar operações booleanas sobre s. Logo, a inferência feita não possui significado. O objetor à estratégia eliminativista tomaria esse raciocínio como exemplar da inadequação do eliminativismo à linguagem natural, afinal, estamos certos em inferir 'a frase do mentiroso não é verdade nem falsa' da teoria eliminativista.

Mostrarei, no entanto, que tal objeção não se sustenta, pois há um erro na análise da frase em termos de uma negação exclusiva. Primeiramente, analisemos com detalhe o comportamento da negação exclusiva. Seja p uma fórmula qualquer. A operação de ' \sim^* ' sobre p se comporta do seguinte modo: $v(\sim^*p) = 0$ se $v(p) = 1$, $v(\sim^*p) = 1$ se $v(p) \neq 1$. Esse esquema captura a ideia intuitiva que negar exclusivamente p é o mesmo que dizer que p possui outro valor de verdade além/excluído do verdadeiro.

O objetor diria que a frase do mentiroso não possui o valor falso nem o verdadeiro, e pensaria que a negação exclusiva estivesse afirmando *apenas* isso. No entanto, a negação exclusiva de uma frase p só é verdadeira quando o valor de verdade de p é *diferente* de 1, ou seja, quando existe *algum* valor de verdade associado a p , o que não é o caso na frase do

mentiroso, pois ela não possui significado. O que ocorreria de fato nessa situação é uma negação categorial, do tipo 'a árvore não é cega': uma árvore não é cega nem vidente, mas isso não significa que árvores violem o princípio do terceiro excluído, pois o tipo de negação envolvida em cada caso é diferente. A negação categorial, ao contrário da negação exclusiva e da negação por escolha, não precisa ser operada apenas sobre frases (ou seja, expressões com significado).

Assim, "a frase do mentiroso não é verdadeira nem falsa" recai sob o mesmo tipo de análise da frase "um balbuciar de latidos do meu cachorro não é verdadeiro nem falso". *Como* analisar tais frases, eis um problema. No entanto, ao mostrar que existem frases não problemáticas do mesmo tipo que a frase referente ao mentiroso, fica claro que não há nada de inerentemente errado com frases que fazem que referência a frases do mentiroso.⁴² Caso alguém queira levantar um caso contra afirmações teóricas sobre o paradoxo do mentiroso na teoria eliminativista, será necessário explicar como é possível falar veritativamente sobre qualquer expressão sem significado.⁴³

Nosso objetor, ainda insatisfeito com o eliminativismo, argumenta o seguinte: no caso da frase 'o balbuciar do meu cachorro não é verdadeiro', utilizando a negação categorial, não parece ser possível eliminar a expressão 'verdadeiro'. Conforme comenta Ramsey, a eliminabilidade depende da expressão *p* possuir um verbo prévio que dê força assertiva suficiente: "nós temos que adicionar 'é verdadeiro' para dar um verbo a uma frase, esquecendo que 'p' já contém um verbo (variável)" (RAMSEY, 1931, p. 128, tradução minha). No presente caso, porém a expressão 'p' não contém um verbo, pois 'p' é 'o balbuciar do meu cachorro não é verdadeiro'. Logo, conclui o objetor, o princípio da eliminabilidade possui vários contraexemplos – a saber, todas as frases que afirmam a verdade (ou falsidade) de frases sem sentido –, o que mina o projeto eliminativista.

O eliminativismo possui ao menos duas saídas de tal objeção, cada uma custando caro ao seu projeto: em primeiro lugar, o eliminativista pode sustentar que o princípio da eliminabilidade não vale no caso da negação categorial. Esse resultado seria desanimador, pois a motivação para o eliminativismo são os casos particulares em que a análise eliminativista funciona bem. A proliferação de contraexemplos seria um indício forte de que

⁴² Ainda utilizo a expressão "frase do mentiroso", mas nesse ponto fica claro, segundo o eliminativismo, que "esta frase é falsa" não pode ser chamada corretamente de frase.

⁴³ Ao utilizar a expressão "frase do mentiroso", não mais me refiro ao conteúdo expresso por frases do *tipo* do mentiroso, mas sim a alguma ocorrência específica de uma sequência de palavras.

tal princípio não possui a generalidade que pensávamos possuir.⁴⁴ Em segundo lugar, o eliminativista pode cortar o mal pela raiz, e sustentar que não há tal coisa como uma negação categorial em relação à verdade. Esse caminho, no entanto, nega a inteligibilidade de expressões cotidianas como 'o balbuciar do meu cachorro não é nem verdadeiro nem falso' e a inteligibilidade de frases da teoria eliminativista sobre o paradoxo do mentiroso que envolvam a noção de verdade.⁴⁵

O segundo caminho, apesar dos pesares, é o mais promissor: no germe do deflacionismo está a ideia que o predicado "é verdadeiro" não se comporta da mesma maneira que outros predicados empíricos (HORWICH, 2010, p. 4), como "é cego", "é vermelho", ou seja, a verdade não é uma noção substancial. Repare, no entanto, como se explica a negação categorial:

A negação de que se trata aqui não pode ser, portanto, predicativa, mas deve ser categorial, pois não se está dizendo que uma parte de um campo predicativo não se aplica a um objeto, mas sim a outra, complementar, como, por exemplo, quando dizemos que este número não é par, mas ímpar, mas se está excluindo a aplicação do campo predicativo inteiro ao objeto, como se declarássemos que este número não é pesado [...] (FLECK, 2003, p. 85)

Isto é, a negação categorial exclui o objeto de todo campo predicativo engendrado pelo predicado em questão por conta de suas qualidades inerentes, ou seja, a negação categorial envolve apelo a algum tipo de apelo à natureza dos objetos, que explicaria o fato de um objeto poder não fazer parte da categoria de objetos que possui ou não possui tal qualidade. Logo, a negação categorial pressupõe a noção de um predicado substancial, que vai contra um *insight* fundamental do deflacionismo em geral. Portanto, ainda que tomar o segundo caminho tenha a indesejada consequência de tornar o eliminativismo menos intuitivo, a negação da negação categorial à "verdade" possui uma razão específica que permite entender melhor o funcionamento do termo "verdade" de maneira correta.

Em síntese, pela ótica do eliminativismo, os paradoxos semânticos de autorreferência são dissolvidos como problemas inexistentes por conta de uma má concepção do modo como o predicado "é verdadeiro" funciona. Por outro lado, porém, o eliminativismo descarta casos não problemáticos de autoreferência, como o veraz – 'esta frase é verdadeira' –, e outras atribuições circulares que possuem modelos consistentes.

Ao ser confrontado com tal objeção, o eliminativista é forçado a expor quão radical – ainda que pretenda ser uma teoria descritivista em sua essência – sua posição é. Como o

⁴⁴ Enquanto *teoria da verdade*, o eliminativismo estaria gravemente comprometido, pois haveria um uso da noção de verdade que não estaria explicado pela análise eliminativista.

⁴⁵ Ainda poderíamos afirmar, por exemplo, que a frase do mentiroso não possui significado, pois é apenas o predicado de verdade que foi caracterizado como 'não-substancial', além do fato de que no deflacionismo, a noção de significado não pode ser construída sobre a noção de verdade (ver Armour-Garb (2003)).

princípio da eliminação é generalizado, ele deve ser aplicado às frases não problemáticas de autorreferência,⁴⁶ o que bloqueia o uso de alguns casos pedestres de "é verdadeiro". Apesar de não parecer intuitivo, essa é a consequência de caracterizar a verdade como não sendo um predicado lógico genuíno: qualquer caso em que a noção de verdade não esteja resguardada por algum outro verbo, a análise da frase se torna infactível.

Dessa maneira, o problema da frase do mentiroso está localizado no uso substancial da noção de verdade pressuposto na autorreferência. Ao se entender isso, é possível entender que as demais frases bloqueadas pelo eliminativismo, ainda que não derivem contradições, sofrem do mesmo problema: elas estendem o uso do predicado "é verdadeiro" para um uso substancial. Pelo véu do cotidiano, esquecemos a diferença fundamental entre os predicados substanciais como "é azul", "é pequeno", e o predicado "é verdadeiro", e montamos pequenas ilhas isoladas na linguagem em que se pode aplicar "é verdadeiro" como se fosse um predicado substancial – às vezes, isso não gera problemas, como no caso veraz. Outras vezes, no entanto, nossa leviandade pode causar consequências catastróficas, como no caso do paradoxo do mentiroso.

⁴⁶ Lembrando que a autorreferência só é bloqueada, pelo eliminativismo, quando também envolve a noção de verdade.

4 CONCLUSÃO

Façamos uma avaliação final das posições lógico-filosóficas apresentadas. Meu primeiro veredito diz respeito ao diagnóstico clássico de Tarski. É cabível distinguir entre (i) a resposta de Tarski ao paradoxo do mentiroso e (ii) respostas de inspiração tarskiana. Como visto no capítulo 1, a resposta de Tarski ao paradoxo do mentiroso se divide em um tratamento da linguagem natural e outro sobre linguagens formais. É apenas para as linguagens formais que Tarski constrói sua hierarquia de predicados, pois a linguagem natural não possui uma estrutura definida o suficiente para tal empresa. Nesse sentido, pode-se entender que Tarski esteja propondo uma lógica *revisionista*, ou seja, ele pretende *prescrever* a melhor maneira de usarmos o(s) predicado(s) "verdadeiro" (nesse caso, a caracterização é restrita a linguagens formais). Isto é, Tarski oferece uma maneira eficaz de evitar o surgimento de paradoxos em linguagens formais, sem que isso tenha comprometimentos em relação ao uso do predicado "verdadeiro" em linguagens naturais.

Uma resposta de tipo (ii) é uma extensão do projeto tarskiano de definir o conceito de verdade quando aplicado à linguagem natural, ou seja, (ii) busca reconstruir uma hierarquia de linguagens no português, no inglês, etc. É por conta dessa resposta que surgem suspeitas em relação à adequação da resposta de Tarski, pois, ao se tratar de linguagens naturais, estamos tratando de uma lógica *descritiva*, ou seja, uma lógica que pretenda ser fiel ao modo como nós efetivamente raciocinamos (corretamente) e utilizamos as palavras. Nesse caso, é contraintuitivo pensar que haja uma hierarquia de predicados de "verdade", pois seguidamente utilizamos cadeias de raciocínios circulares que não são paradoxais, além de nunca estarmos pensando no "nível" de uma determinada frase para anexar o predicado adequado a ela (KRIPKE, 1975, p. 691).

Kripke, contrariamente a Tarski, buscou respostas ao paradoxo do mentiroso conforme ele surge na linguagem natural, portanto, o veredito a seu respeito deve pesar se ele obteve sucesso em se adequar a tal tarefa.⁴⁷

Ele procurou enfatizar a *universalidade* da linguagem natural. Assim, seu desafio consiste em desenvolver uma linguagem semanticamente fechada que não gere resultados inaceitáveis. Apesar disso, Kripke está continuamente assombrado pelo "fantasma de Tarski" (KRIPKE, 1975, p. 714), pois alguns conceitos semânticos que não são expressáveis na

⁴⁷ Ver Kripke (1975, p. 695).

linguagem-objeto de Kripke se elevam a uma formulação metalinguística que gera uma hierarquia de linguagens tarskiana.

A solução eliminativista, por outro lado, mostrou-se resistente às objeções. Embora tal opção ainda seja em grande medida uma aposta (pois não há um argumento cabal que convença a todos do princípio da eliminabilidade), a análise das frases contendo o predicado de verdade a partir da análise ramseyana e da teoria das descrições de Russell promete ser um caminho frutífero para conter a ameaça de paradoxos semânticos na linguagem, pois permite que, em casos cotidianos, predicados semânticos continuem se comportando de maneira típica – desde que sua força assertiva provenha de outro verbo – (a frase do mentiroso contingente, *quando faz referência a frases fundadas*, possui um sentido bem determinado), ao passo que seu funcionamento atrofia quando são considerados fora desse contexto (pois, pelo princípio da eliminabilidade, a frase do mentiroso contingente – *no contexto do paradoxo* – não faz referência a uma frase significativa).

Em síntese, uma exposição sistemática das respostas contemporâneas mostrou que há elementos recorrentes que o paradoxo traz à tona. Há um padrão na insuficiência das teorias: sua falha acontece ora por conta da *artificialidade*, no caso de Tarski, em que se construíram linguagens desconexas à linguagem natural, ora por conta da *inexpressibilidade*, nos casos de Kripke e do eliminativismo, em que algumas intuições naturais se mostraram errôneas. Dessa maneira, o paradoxo do mentiroso nos ensina uma lição sobre a linguagem que pode ser resumida no seguinte lema: as coisas são como são, mas do jeito como estão, elas não podem ser.

O eliminativismo, por fim, se conforma exemplarmente ao lema do paradoxo: seu princípio fundamental se baseia em uma análise de como utilizamos a verdade nos casos mais simples e básicos da linguagem. A teoria resultante, porém, nos ensina a melhor utilizar o conceito de verdade, reconhecendo que uma análise detida de um paradoxo grandioso como o paradoxo do mentiroso nos fará olhar para a linguagem como nunca antes pensamos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ARMOUR-GARB, B. Deflationism and the meaningless strategy. **Analysis**, vol. 61, n. 4, p. 280-289, outubro, 2001.
- BEALL, J. C. A neglected deflationist approach to the liar. **Analysis**, vol. 61, n. 2, p. 126-129, abril, 2001.
- CARDOSO, G. **Repercussões Contemporâneas do Paradoxo do Mentiroso**: Tarski e Kripke. 2010. Dissertação (Mestrado em Filosofia) – Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais. Belo Horizonte, 2010.
- FLECK, F. P. Eternidade, negação e conhecimento. **Analytica**, vol. 7, n. 1, p. 83-96, 2003.
- FREGE, G. Begriffsschrift. In: HEIJENOORT, J. (Ed). **From Frege to Gödel**: a source book in Mathematical Logic. Londres: Harvard University Press, 1967, p. 1-82.
- FREGE, G. **Investigações Lógicas**. Tradução: Paulo Alcoforado. Porto Alegre: Edipucrs, 2002.
- GRIM, P. **The Incomplete Universe**: totality, knowledge and truth. Londres: The MIT Press, 1991.
- GUPTA, A. Truth and paradox. **Journal of Philosophical Logic**, vol. 11, p. 1-60, 1982.
- HORWICH, P. **Truth – Meaning – Reality**. Nova York: Oxford University Press, 2010.
- KLEENE, S. C. **Introduction to Metamathematics**. Oxford: Wolters-Noordhoff Publishing, 1971.
- KREMER, M. Kripke and the logic of truth. **Journal of Philosophical Logic**, vol. 17, p. 225-278, 1988.
- KRIPKE, S. Outline of a theory of truth. **The Journal of Philosophy**, vol. 72, n. 19, p. 690-716, novembro, 1975.
- MCGEE, V. **Truth, Vagueness, and Paradox**: an essay on the logic of truth. Londres: The MIT Press, 1991.
- QUINE, W. V. **Pursuit of Truth**. Cambridge: Harvard University Press, 1992.
- RAMSEY, F. Facts and Propositions. In: RAMSEY, F. **The Foundations of Mathematics and other logical essays**. Londres: Routledge, 1931, p. 125-37.
- RUSSELL, B. On Denoting. **Mind**, vol. 14, n. 56, p. 479 – 493, 1905.
- SMULLYAN, R. M. **Diagonalization and Self-Reference**. Oxford: Clarendon Press, 1994.
- SMULLYAN, R. M. **Lógica de Primeira Ordem**. Tradução: LOPARIC, A. M.; MAZAK, R. P.; VICENTE, L. 2ª ed. São Paulo: Editora UNESP, 2009.

SOAMES, S. **Understanding Truth**. Nova York: Oxford University Press, 1998.

STRAWSON, P. F. On Referring. **Mind**, vol. 59, n. 235, p. 320 – 344, 1950.

TARSKI, A. **The Concept of True Sentence in Formalized Languages**. In: **Logic, Semantics, Metamathematics**: papers from 1923 to 1938. Tradução: J. H. Woodger. Oxford: Clarendon Press, 1956.

TARSKI, A. **The Semantic Conception of Truth**: and the Foundations of Semantics. *Philosophy and Phenomenological Research*, vol. 4, n. 3, p. 341-376, março, 1944.