Paradoxos Semânticos

Edição de 2014 do

Compêndio em Linha de Problemas de Filosofia Analítica

2012-2015 FCT Project PTDC/FIL-FIL/121209/2010

Editado por
João Branquinho e Ricardo Santos

ISBN: 978-989-8553-22-5

Centro de Filosofia da Universidade de Lisboa
Alameda da Universidade, Campo Grande, 1600-214 Lisboa

DOI: https://doi.org/10.51427/cfi.2021.0059

Resumo
Os paradoxos semânticos são uma família de argumentos – onde se incluem o paradoxo do mentiroso, o paradoxo de Curry, o paradoxo da heterologicalidade (de Grelling), os paradoxos da definibilidade (de Richard e de Berry) e outros – que têm duas coisas em comum: primeiro, envolvem de maneira essencial conceitos semânticos como os de verdade, satisfação, referência, definição, etc.; e, segundo, parecem muito bons argumentos até observarmos que a sua conclusão é contraditória ou absurda. Estes argumentos levantam sérias suspeitas quanto à coerência dos conceitos envolvidos. Neste artigo, faremos uma apresentação introdutória de algumas das principais teorias que têm sido propostas para resolver os paradoxos e afastar essas suspeitas. Incluímos também uma breve história dos paradoxos semânticos de Eubúlides a Tarski e a Curry.

Palavras-chave
Verdade, paradoxo do mentiroso, auto-referência, transparência, lógicas não-clássicas

Abstract
The semantic paradoxes are a family of arguments – including the liar paradox, Curry's paradox, Grelling's paradox of heterologicality, Richard's and Berry's paradoxes of definability, and others – which have two things in common: first, they make an essential use of such semantic concepts as those of truth, satisfaction, reference, definition, etc.; second, they seem to be very good arguments until we see that their conclusions are contradictory or absurd. These arguments raise serious doubts concerning the coherence of the concepts involved. This article will offer an introduction to some of the main theories that have been proposed to solve the paradoxes and avert those doubts. Included is also a brief history of the semantic paradoxes from Eubulides to Tarski and Curry.

Keywords
Truth, liar paradox, self-reference, transparency, non-classical logics

Paradoxos Semânticos

DOI: https://doi.org/10.51427/cfi.2021.0059

1 Breve História dos Paradoxos Semânticos

Atribui-se geralmente a Eubúlides de Mileto, filósofo megárico do século IV a.C., a invenção do paradoxo do mentiroso: se um homem diz ‘Estou a mentir’, será o que ele diz verdadeiro ou falso? Se for verdadeiro, então ele está realmente a mentir e, por isso, o que diz é falso. Se for falso, isso concorda com o que ele afirma, parecendo então que o que diz é verdadeiro.

Há quem pense que Aristóteles teve conhecimento do paradoxo do mentiroso e que propôs uma solução para ele no capítulo 25 das *Refutações Sofísticas*, a qual consistiria essencialmente em defender que o homem que diz que está a mentir nem está a dizer algo verdadeiro nem está a dizer algo falso (veja-se Crivelli 2004). Mas esta interpretação é duvidosa (veja-se Santos 2009) e, a julgar pelos textos que nos chegaram, parece mais seguro afirmar que Aristóteles, se porventura conheceu alguma versão do paradoxo, não lhe terá atribuído importância filosófica.

O contrário terá acontecido com Crisipo, filósofo estóico do século III a.C. De acordo com o catálogo das suas obras preservado por Diógenes Laércio, terá escrito nada menos que catorze livros (!) sobre o paradoxo do mentiroso. Infelizmente, todos eles se perderam (à excepção de um pequeno e obscuro fragmento) e, por isso, a posição de Crisipo a respeito do paradoxo pode apenas ser conjecturada de maneira muito especulativa. Estudos recentes (Cavini 1993, Mignucci 1999, Papazian 2012) têm discutido se é plausível atribuir-lhe uma solução segundo a qual frases como ‘Estou agora a dizer uma falsidade’ não expressam nenhuma proposição.

O paradoxo do mentiroso é muitas vezes também designado por ‘paradoxo de Epiménides’. Epiménides terá nascido e vivido na ilha de Creta (no século VI a.C.) e, segundo parece, tinha em fraca consideração os seus concidadãos. O apóstolo Paulo, na Carta a Tito (incluída no *Novo Testamento*), relata um famoso dito deste cretense antigo: “Os cretenses são sempre mentirosos, bestas más e ventres preguiçosos”; e, na opinião de Paulo, “este testemunho é verdadeiro”. Focado sobretudo nas qualidades morais dos habitantes da ilha,

Publicado pela primeira vez em 2014

o apóstolo não se terá apercebido das consequências do seu relato: é que, se o dito de Epiménides é realmente verdadeiro, então tudo o que os cretenses dizem é falso e, como ele próprio é cretense, o seu dito é falso.

No final da antiguidade, o interesse pelo estudo do paradoxo desvanece-se, ganhando preponderância opiniões como esta de Séneca: "Para quê deixares-te torturar por um problema que é mais correcto ignorar do que tentar resolver?" (*Cartas a Lucílio,* 49, 6).

Nos séculos XII e XIII ocorre um ressurgimento do interesse (motivado, ao que parece, pelo desenvolvimento de um género específico da lógica medieval, que é a teoria das 'obrigações' [veja-se Martin 2001]), tanto no paradoxo do mentiroso como em outros semelhantes, genericamente designados como 'insolúveis'. Uma das primeiras abordagens foi a solução "cassatória" (de *cassatio*, que significa anular ou cancelar) segundo a qual quem profere uma frase como 'O que estou agora a dizer é falso' não está a dizer nada. Outra abordagem medieval que merece ser referida é a daqueles (conhecidos como *restringentes*) que alegavam que uma parte de uma elocução não pode referir a totalidade dessa mesma elocução – impedindo assim aquilo a que hoje chamamos a auto-referência (de uma frase ou de uma elocução). Uma terceira abordagem interessante que surge no século XIV tem como ideia central a de que todas as frases significam ou implicam a sua própria verdade. Partindo daqui seria possível argumentar que a frase 'Esta frase é falsa' é simplesmente falsa; e daí não se derivaria que ela seja também verdadeira, porque a frase significa mais do que a sua falsidade (e, para uma frase ser verdadeira, é preciso que tudo o que ela significa seja o caso). Thomas Bradwardine, Alberto da Saxónia e Jean Buridan são três dos principais autores que exploraram este género de abordagem (veja-se Spade 1975, Spade 1982 e Read 2002). As concepções medievais sobre os paradoxos semânticos têm sido objecto de bons estudos recentes (veja-se Rahman, Tulenheimo e Genot 2008 e Dutilh Novaes 2008, entre outros).

Após um novo período longo de desinteresse, o estudo dos paradoxos semânticos é retomado no início do século XX, de Bertrand Russell em diante, a par com o desenvolvimento da lógica matemática.

No artigo de 1908, em que apresentou a teoria lógica dos tipos, Russell começa por passar em revista um conjunto de paradoxos, que considera serem todos reveladores do mesmo fenómeno. Além do paradoxo do mentiroso e do paradoxo por ele próprio descoberto (o famoso paradoxo da classe de todas as classes que não são membros de si mesmas, publicado em 1903), há nessa lista dois outros paradoxos que merecem ser aqui referidos – o paradoxo de Richard e o paradoxo de Berry.

Jules Richard era professor de matemática no liceu de Dijon quando, em Junho de 1905, enviou uma pequena carta ao editor da *Revue génerale des sciences pures et appliquées*, comunicando a descoberta de uma certa "contradição". O que Richard aí apresenta é algo muito semelhante ao "argumento diagonal" usado por Cantor (em 1891) para demonstrar que existem conjuntos infinitos que não são contáveis, isto é, que são "maiores" do que o conjunto dos números naturais. Usando essencialmente a mesma técnica, Richard começa por definir um certo conjunto C de números, a saber, *o conjunto de todos os números reais que são definíveis com um número finito de palavras*; depois, Richard forma um número real N, define-o com um número finito de palavras e mostra que ele não pertence ao conjunto C. No ano seguinte, Peano observou com razão que o paradoxo de Richard "não pertence à matemática, mas sim à linguística". De facto, trata-se de um paradoxo *semântico*, pois diz respeito à relação de *definibilidade*, entendida como relação entre os números, por um lado, e as expressões da linguagem que estamos a usar (e que era a língua francesa, no caso de Richard), por outro. (Mais tarde, foi sobretudo com o artigo de Ramsey (1925) que a distinção entre paradoxos semânticos e paradoxos matemáticos ou da teoria dos conjuntos se estabeleceu.)

O paradoxo de Berry possui também o mesmo carácter semântico, relacionando os números com os seus nomes (ou descrições definidas) numa certa linguagem. G. G. Berry era bibliotecário na Bodleian Library em Oxford e comunicou pessoalmente este paradoxo a Russell. A expressão 'o menor número inteiro não nomeável em menos do que vinte e oito sílabas' denota um certo número específico, que podemos dar-nos ao trabalho de encontrar, tomando por referência a língua portuguesa. Esse número é o menor número inteiro não nomeável (em português) em menos do que vinte e oito sílabas. Acontece que a expressão que usámos para o nomear contém vinte e sete sílabas. Conclui-se, então, que o menor número inteiro não nomeável em menos do que vinte e oito sílabas é nomeável em menos do que vinte e oito sílabas.

Embora tenha como objectivo principal resolver o paradoxo das classes que não pertencem a si próprias, o artigo de Russell de 1908 inclui também um esboço de uma solução para o paradoxo do mentiroso. Segundo Russell, quando uma pessoa diz 'Estou agora a mentir', ela teria a pretensão de afirmar uma proposição que seria acerca de todas as proposições. Na sua perspectiva, aquela frase seria equivalente a 'Existe uma proposição *p* tal que eu estou a afirmar *p* e *p* não é verdadeira', o que é por sua vez equivalente a 'Não é o caso que, para toda a proposição *p*, se afirmo *p*, então *p* é verdadeira'. Mas Russell defende que a pretensão de falar acerca de todas as proposições é ilegítima, pois não existe tal totalidade. O seu argumento tem a forma de uma redução ao absurdo: se existisse a totalidade das proposições, haveria proposições acerca dessa totalidade, as quais não poderiam pertencer a essa totalidade, pois, se lhe pertencessem, seriam verdadeiras e falsas ao mesmo tempo.

Em 1907, Kurt Grelling, aluno de David Hilbert na Universidade de Göttingen, descobriu um novo paradoxo, que publicou no ano seguinte num artigo escrito em conjunto com o seu mentor Leonard Nelson (Grelling e Nelson 1908; sobre o contexto desta descoberta, leia-se o informativo ensaio de Peckhaus 1995). Um predicado como 'é verde' é *verdadeiro de* todas as coisas verdes e somente das coisas verdes. Um predicado como 'é uma expressão com doze sílabas' é verdadeiro das, e somente das, expressões com doze sílabas; por isso, em particular, ele é verdadeiro dele próprio. Há, então, predicados que são verdadeiros deles próprios (como 'é um predicado', 'é português', 'não é verde', etc.) e predicados que não são verdadeiros deles próprios (como 'é uma bicicleta', 'é chinês', 'não tem significado', etc.). Chamemos 'autológicos' aos primeiros e 'heterológicos' aos segundos. A qual destas duas classes pertence o predicado 'é heterológico'? Apesar de ter de pertencer a uma delas, um raciocínio simples mostra que não pode pertencer a nenhuma. Pois suponhamos, primeiro, que o predicado 'é heterológico' é autológico. Pela definição de 'autológico', segue-se daí que ele é verdadeiro dele próprio. Mas, se ele é verdadeiro dele próprio, então é heterológico e não autológico. Suponhamos, em vez disso, que o predicado 'é heterológico' é heterológico. Pela definição de 'heterológico', segue-se daí que ele não é verdadeiro dele próprio. Mas, então, ele não é heterológico.

O trabalho de Alfred Tarski sobre o conceito de verdade é justamente reconhecido como o ponto de partida obrigatório para qualquer estudo contemporâneo sério sobre os paradoxos semânticos. Num longo artigo (em polaco) publicado em 1933 (geralmente referido com o título da tradução alemã como "O Conceito de Verdade nas Linguagens Formalizadas"), Tarski propôs-se resolver "o problema da definição de verdade", o qual consiste em construir, para cada linguagem *L* de um certo tipo, uma definição materialmente adequada e formalmente correcta do predicado 'é uma frase verdadeira de *L*'. Na sua perspectiva, o paradoxo do mentiroso e os restantes paradoxos semânticos (a que Tarski costumava chamar 'antinomias') constituíam os principais obstáculos, não só para a construção de uma definição de verdade, mas até para a construção da semântica como disciplina científica. Noutro trabalho, escreve:

> Deve enfatizar-se que as antinomias desempenharam um papel proeminente no estabelecimento dos fundamentos das ciências dedutivas modernas. E tal como as antinomias da teoria dos conjuntos, e em particular a antinomia de Russell [...], foram o ponto de partida para as tentativas bem-sucedidas de uma formalização consistente da lógica e da matemática, assim a antinomia do mentiroso e outras antinomias semânticas originam a construção da semântica teórica. (Tarski 1944: 348)

Apresentaremos na próxima secção a solução hierárquica que Tarski propôs para os paradoxos semânticos. Por agora, limitamos a nossa atenção a um dos aspectos mais importantes da maneira como formula o paradoxo do mentiroso, que é o papel das chamadas 'frases-V' ou 'bicondicionais descitativas'. É curioso notar que, apesar de Tarski atribuir explicitamente essa formulação "extremamente simples" do paradoxo a Jan Łukasiewicz (cf. Tarski 1983: 157), tudo indica que o papel fundamental das frases-V lhe terá sido apontado pelo seu antigo professor em Varsóvia, Stanisław Lesniewski. (Sobre estas duas influências, veja-se Wolenski 1994, Betti 2004 e Patterson 2012.)

Consideremos uma frase portuguesa particular como, por exemplo, a frase 'A neve é branca'. O que significa dizermos que esta frase é verdadeira? Segundo Tarski, o conceito de verdade, na sua aplicação *a esta frase particular*, pode definir-se assim:

‘A neve é branca’ é verdadeira se e somente se a neve é branca.

Esta bicondicional dá condições necessárias e suficientes para a aplicação do conceito de verdade àquela frase particular – e condições que são simples e intuitivamente correctas. As bicondicionais deste tipo são chamadas ‘frases-V’, mas há quem prefira chamar-lhes ‘bicondicionais *descitativas*’, por considerar que elas afirmam que há equivalência entre dizer que uma frase é verdadeira e retirar-lhe as aspas (quer dizer, fazer o inverso da operação de *citar*). Segundo Tarski, elas são fundamentais para a construção de uma definição de verdade para uma linguagem. Pois, se uma frase-V diz tudo o que há a dizer sobre a verdade de uma frase particular, o problema resume-se ao de gerar a totalidade das frases-V (ou a sua “soma lógica”) para a totalidade das frases da linguagem em causa. A famosa “Convenção V” de Tarski diz então que uma definição de verdade para uma linguagem *L* só será adequada se tiver como consequência, para cada frase de *L*, a respectiva frase-V, com a forma:

x é verdadeira se e somente se *p*

em que o lugar de ‘*p*’ é ocupado por uma frase de *L* e no lugar de ‘*x*’ está um nome (ou termo singular descritivo) dessa frase.

Então e se, no lugar de ‘*p*’, estiver uma frase que nega a sua própria verdade, ou seja, uma frase c idêntica a ‘c não é verdadeira’? A frase-V para esta frase será:

‘c não é verdadeira’ é verdadeira se e somente se c não é verdadeira.

Mas como sabemos que c = ‘c não é verdadeira’, podemos substituir um nome pelo outro e obtemos:

c é verdadeira se e somente se c não é verdadeira.

E isto implica, na lógica clássica, a contradição ‘c é verdadeira e c não é verdadeira’.

Tarski retira desta formulação do paradoxo a seguinte *lição geral*: qualquer linguagem que seja capaz de nomear todas as suas frases, que contenha o seu próprio predicado de verdade, na qual todas as frases-V sejam verdadeiras e cuja lógica seja clássica é uma linguagem inconsistente. Em concordância com Lesniewski, Tarski afirma que a linguagem natural, porque satisfaz estas condições, é inconsistente.

Mais adiante no artigo, após dirigir a sua atenção para as linguagens formalizadas, Tarski transformará uma versão formal deste argumento numa demonstração da indefinibilidade numa linguagem (suficientemente expressiva) da verdade para essa mesma linguagem.

Haskell Curry foi um lógico americano, que estudou com Hilbert e Bernays em Göttingen, tendo-se doutorado em 1930 com uma tese sobre lógica combinatória. Já regressado aos EUA, como professor no Pennsylvania State College, Curry publicou em 1942 um artigo onde apresenta um novo paradoxo que também explora a auto-referência e a circularidade, mas que, diferentemente do paradoxo de Russell e do paradoxo do mentiroso, *não usa a negação*. O paradoxo de Curry permite provar, a partir de premissas simples e aparentemente inócuas, tudo o que se quiser (brincando com isso, num artigo influente, Prior (1955) exemplifica o paradoxo construindo uma demonstração formal de que a lua é feita de queijo verde). O paradoxo admite várias versões, mas aqui interessa-nos sobretudo a versão semântica, que usa uma frase **s** que é idêntica à condicional 'Se **s** é verdadeira, então 1 é igual a 2', que podemos reescrever como 'Vs→1=2'. A frase-V para esta frase diz que

$$Vs \leftrightarrow (Vs \rightarrow 1=2)$$

que podemos dividir em duas partes:

$$Vs \rightarrow (Vs \rightarrow 1=2) \quad \text{e} \quad (Vs \rightarrow 1=2) \rightarrow Vs$$

Além da frase-V assim dividida, o argumento usa apenas o *Modus Ponens* e uma regra clássica conhecida como 'princípio da contracção', que diz que

de $A \rightarrow (A \rightarrow B)$, podemos inferir que $A \rightarrow B$.

A demonstração é então a seguinte:

1.	$Vs \rightarrow (Vs \rightarrow 1=2)$	frase-V, parte i
2.	$(Vs \rightarrow 1=2) \rightarrow Vs$	frase-V, parte ii
3.	$Vs \rightarrow 1=2$	1, contracção
4.	Vs	2,3, MP
5.	$1=2$	3,4, MP

2 Alfred Tarski: a teoria da verdade, o teorema da indefinibilidade e a hierarquia de linguagens

A teoria da verdade de Tarski propõe, não uma definição geral única do conceito de verdade, mas um método para a construção de definições de verdade, cada uma das quais dizendo respeito apenas a uma linguagem particular. (Para uma apresentação mais pormenorizada da teoria, leia-se Soames 1999 e Santos 2003.) Mais exactamente, o método proposto permite, para cada linguagem formal *L* de um certo tipo, construir, numa metalinguagem *L** apropriada, uma definição do conceito de *frase-verdadeira-de-L*. Mais exactamente ainda, o que se define em *L** é o *predicado* de um lugar 'é verdadeira em *L*'. Para isso, requer-se que *L** contenha nomes para todas as expressões de *L*, traduções para todas as expressões de *L* e recursos elementares de lógica e teoria dos conjuntos.

Quanto ao âmbito de aplicação do método, Tarski concluiu que, através dele, o predicado de verdade-em-*L* pode ser definido para todas as linguagens *L*, desde que a definição seja construída numa metalinguagem "de ordem superior" (Tarski 1983: 273). Esta condição de que a metalinguagem seja num certo sentido mais forte do que a linguagem-objecto é requerida para evitar que a definição construída na metalinguagem pudesse ser de algum modo traduzida para a própria linguagem-objecto, o que dotaria esta última do seu próprio predicado de verdade. Pois nenhuma linguagem (ou, pelo menos, nenhuma linguagem "clássica", do tipo daquelas que Tarski tinha em vista) pode conter o seu próprio predicado de verdade – é esse o conteúdo do famoso *teorema da indefinibilidade* de Tarski.

O teorema da indefinibilidade demonstra-se facilmente a partir de um teorema da aritmética conhecido como *lema da diagonalização*. Todas as linguagens interessantes têm a capacidade de nomear as suas próprias expressões e fórmulas. Como mostrou Gödel, até a linguagem da aritmética tem esta capacidade, de maneira indirecta, usando um sistema de codificação numérica. Suponhamos então que, quando φ é uma expressão qualquer de uma linguagem *L*, $[\varphi]$ é o nome dessa expressão (ou do número-de-código que lhe foi atribuído) disponível em *L*. A existência destes nomes permite formar em *L*, não apenas frases que "afirmam coisas" acerca de diversas expressões ou fórmulas de *L*, mas também frases que "afirmam coisas" acerca de si mesmas. E que "coisas", ou melhor, que propriedades são essas que uma frase pode afirmar acerca de si mesma? *Toda e qualquer propriedade que possa ser expressa na linguagem em causa.* Mais exactamente:

> Lema da Diagonalização: Para qualquer fórmula $\varphi(\upsilon)$ da linguagem da aritmética, com υ como sua única variável livre, existe uma frase γ nessa linguagem que é tal que $\gamma \leftrightarrow \varphi([\gamma])$ é uma verdade aritmética (ou é demonstrável na aritmética).[1]

A partir deste lema[2], a prova da indefinibilidade procede por *reductio*, supondo que seria possível definir em L um predicado $V_L(x)$ de verdade-em-L. A Convenção-V de Tarski estabelece que, se $V_L(x)$ é um predicado de verdade-em-L, então, para toda a frase φ de L, a bicondicional

[1] Demonstração do lema da diagonalização: O objectivo da prova é, partindo de uma propriedade qualquer, expressa por uma fórmula $\varphi(x)$, mostrar que existe uma frase que atribui essa propriedade a si mesma. O que se faz construindo essa frase. Seja então $\varphi(x)$ uma fórmula qualquer (da linguagem da aritmética), com x como sua única variável livre. Como já sabemos, o nome de $\varphi(x)$ escreve-se $[\varphi(x)]$. A *diagonal* de $\varphi(x)$ é a fórmula que resulta de substituirmos a variável pelo nome da própria fórmula, ou seja, é $\varphi([\varphi(x)])$. A operação de diagonalização expressa-se na linguagem através de um predicado Dxy ('a fórmula x é a diagonal da fórmula y'). Então, dada a fórmula $\varphi(x)$, seja $\psi(x)$ a fórmula $\exists y(Dyx \wedge \varphi(y))$ (que "diz" que "alguma diagonal de x tem a propriedade φ"). A diagonal de $\psi(x)$, ou seja, de $\exists y(Dyx \wedge \varphi(y))$, é a fórmula

$$\exists y\ (Dy[\exists y\ (Dyx \wedge \varphi(y))] \wedge \varphi(y)),$$

que podemos abreviar como γ (e que "diz" que "alguma diagonal de $\exists y(Dyx \wedge \varphi(y))$, i.e. de $\psi(x)$, tem a propriedade φ"). Mas é óbvio (e facilmente demonstrável) que existe uma única diagonal de $\psi(x)$, que é a própria fórmula γ. Podemos então concluir que γ é verdadeira se e somente se γ tem a propriedade φ; ou, mais exactamente, podemos concluir que

$$\gamma \leftrightarrow \exists y\ (y = [\gamma] \wedge \varphi(y)),$$

o que é equivalente a

$$\gamma \leftrightarrow \varphi([\gamma]).$$

[2] O lema da diagonalização foi originalmente usado por Gödel (1931) para construir uma frase que atribui a si própria a propriedade sintáctica da *indemonstrabilidade*, quer dizer, para construir uma frase que é verdadeira se e somente se é indemonstrável, ou seja, para construir uma frase G equivalente a $\neg$*Demonstrável*($[G]$).

(1) $V_L([\varphi]) \leftrightarrow \varphi$

é verdadeira. Mas se $V_L(x)$ é uma fórmula de L (com uma única variável livre), então a sua negação, $\neg V_L(x)$, também o é. Dito de outro modo: se a propriedade de *ser-verdadeira-em-L* é expressável (em L), então a propriedade de *não-ser-verdadeira-em-L* também o é. Agora, o lema da diagonalização diz-nos que, se esta última propriedade é expressável, então existe na linguagem uma frase que a atribui a si própria, quer dizer, existe na linguagem uma Frase Mentirosa que nega a sua própria verdade. Mais exactamente, se $\neg V_L(x)$ é uma fórmula de L, com x como única variável livre, então existe uma frase φ de L tal que

(2) $\varphi \leftrightarrow \neg V_L([\varphi])$

é verdadeira.

Mas se (1) e (2) são ambas verdadeiras, então também é verdade que

(3) $V_L([\varphi]) \leftrightarrow \neg V_L([\varphi])$

Mas (3) é contraditória. Logo, a suposição inicial tem de ser negada, chegando-se assim ao

> Teorema da Indefinibilidade: Se L é uma linguagem capaz de expressar a aritmética, então não é possível definir em L um predicado de verdade-em-L. (Ou: a noção de verdade aritmética não é definível na aritmética.)

O teorema aplica-se à linguagem da aritmética e a qualquer linguagem mais rica, que tenha a capacidade de expressar a sua própria sintaxe[3]. Na realidade, se observarmos bem, podemos ver que o teorema de Tarski, mais do que a indefinibilidade, prova a *inexpressibilidade* da verdade. Pois uma *reductio* em tudo semelhante pode ser feita para a suposição de que *L contém* (mesmo sem o definir) um predicado de verdade-em-L. Daí que seja comum afirmar-se que Tarski provou que nenhuma linguagem "clássica" pode *conter* o seu próprio predicado de verdade.

[3] Quine (1946) mostrou que a linguagem da aritmética pode ser traduzida (de maneira codificada) para qualquer linguagem capaz de designar os seus símbolos e expressar a operação sintáctica de *concatenação*.

Este resultado, na concepção de Tarski, não constitui indicação de nenhum carácter misterioso ou inefável do conceito de verdade. Pois uma linguagem *L** mais rica do que *L* pode perfeitamente expressar, e até definir, o conceito de verdade para *L*. E se, depois, quisermos falar da verdade-em-*L**, também podemos fazê-lo sem problemas, numa outra linguagem mais rica *L***. E assim sucessivamente, ao longo de uma hierarquia de linguagens com poder expressivo crescente.

Com o trabalho de Tarski, o conhecimento dos paradoxos semânticos atingiu um nível que nunca tinha alcançado antes. Como vimos na secção anterior, o paradoxo do mentiroso é conhecido desde a antiguidade e atraiu a atenção de muitos pensadores (e de alguns dos mais capazes) ao longo da história. Outros não lhe atribuíram importância, ou viram-no como uma mera curiosidade, uma espécie de charada ou de quebra-cabeças sem grandes consequências. Tarski levou-o muito a sério. Em 1944 escreveu:

> Na minha opinião, seria bastante errado e perigoso do ponto de vista do progresso científico depreciarmos a importância desta e de outras antinomias, e tratarmo-las como brincadeiras ou jogos sofísticos. É um facto que estamos aqui na presença de uma absurdidade e que fomos compelidos a afirmar uma frase falsa ([...] uma equivalência entre duas frases contraditórias [...]). Se levamos o nosso trabalho a sério, não podemos reconciliar-nos com este facto. Temos de descobrir a sua causa [...]. (Tarski 1944: 348)

Na mesma linha, afirma em 1969 que o paradoxo deve ser encarado como "um sintoma de doença" (Tarski 1969: 110). Ora, o que Tarski por fim julgou ter descoberto foi que "a causa" do paradoxo reside na indefinibilidade da verdade. Na sua perspectiva, o paradoxo do mentiroso é um sintoma da impossibilidade de uma linguagem conter o seu próprio predicado de verdade. O paradoxo pressupõe que isso seria possível e, por isso, começa por propor à consideração uma frase que nega a sua própria verdade; o argumento termina numa contradição, porque aquela pressuposição é falsa. Analogamente, na demonstração do teorema, começa-se por supor que a linguagem contém o seu próprio predicado de verdade e procede-se a uma *reductio* para concluir que a suposição é falsa. O diagnóstico para a "doença" é, então, o de que não existe realmente, em nenhuma linguagem consistente, uma frase que negue a sua própria verdade. Para afirmarmos ou negarmos a verdade de uma frase de uma linguagem, temos de deslocar-nos para outra linguagem que inclua aquela (ou que possa traduzi-la), usando um predicado que existe na segunda, mas não na primeira.

No entanto, esta solução de Tarski é difícil de aceitar. É muito implausível que a linguagem que falamos, quando usamos uma língua natural como o português, não contenha o seu próprio predicado de verdade. Se podemos perfeitamente falar das frases da linguagem que estamos a falar, porque não poderíamos dizer que uma delas é verdadeira? Se um aluno me diz 'Tarski era um lógico brilhante' e eu reajo com 'Isso é bem verdade', será que não podemos, por princípio, estar a falar na mesma linguagem?

Num certo sentido, Tarski estava de acordo quanto a este ponto. Pois ele reconhecia que a linguagem natural contém o seu próprio predicado de verdade e que, por isso, o paradoxo do mentiroso é, na linguagem natural, um argumento válido com premissas verdadeiras. Mas, então, a consequência que daí se tem de retirar – e que Tarski de facto retirou – é que a linguagem natural é inconsistente, não apenas no sentido em que nela existem frases falsas (como a conclusão do argumento paradoxal) que são verdadeiras, mas, mais geralmente (dado o princípio clássico *ex contradictione quodlibet*), no sentido em que nessa linguagem tudo é verdadeiro (incluindo a afirmação de que a lua é feita de queijo verde). Perante isto, pode haver quem prefira retroceder e admitir, contra as aparências, que a linguagem natural não contém realmente o seu próprio predicado de verdade.

3 Saul Kripke: consistência sem bivalência

Num artigo muito influente (Kripke 1975), Kripke tomou como ponto de partida os resultados de Tarski e propôs-se mostrar que uma linguagem consistente pode conter o seu próprio predicado de verdade, *desde que a sua semântica não seja bivalente*. O teorema de Tarski aplica-se a linguagens auto-referenciais e com uma semântica clássica, na qual todas as frases são verdadeiras ou falsas (e nenhuma tem ambos os valores). Em alternativa, Kripke considera linguagens com falhas de valor de verdade (*truth-value gaps*), nas quais algumas frases não são verdadeiras nem falsas. Tal como Strawson (1950), Kripke distingue o facto de uma frase ter significado linguístico do facto de ela (proferida com força assertiva numa dada circunstância) expressar uma proposição. A tese filosófica subjacente é a de que as frases mentirosas são gramaticalmente correctas, mas não expressam nenhuma proposição.

Em particular, Kripke considera linguagens com predicados que não estão totalmente definidos, quer dizer, com predicados que têm uma extensão (que é o conjunto dos objectos a que o predicado se aplica) e uma anti-extensão (que é o conjunto dos objectos a que o predicado não se aplica), as quais são mutuamente exclusivas (não há nada que pertença a ambas), mas não são conjuntamente exaustivas (pois há objectos do domínio que não pertencem a nenhuma delas). A caracterização semântica destas linguagens requer um esquema não-clássico, que permita avaliar frases moleculares (negações, disjunções, quantificações existenciais, etc.) quando alguma das suas frases componentes não tem um valor clássico (i.e., não é verdadeira nem falsa). Para efeitos de exemplificação, Kripke usa as regras fortes de Kleene (1952) (uma lógica geralmente conhecida como K3). De acordo com estas, existem três valores semânticos: 1, ½ e 0; o valor de uma negação $\neg\varphi$ é igual à diferença $1 - val(\varphi)$; o valor de uma disjunção $\varphi \vee \psi$ é o valor que for mais alto entre $val(\varphi)$ e $val(\psi)$; e as quantificações existenciais são como disjunções generalizadas. (As restantes funções podem definir-se a partir destas da maneira habitual.)[4]

Suponhamos então que L é uma linguagem que contém o seu próprio predicado de verdade $V(x)$. Este predicado permite formar frases com a forma $V(t)$, em que t é um termo singular de L que nomeia uma frase φ de L. O valor semântico de $V(t)$ será 1, ½ ou 0, dependendo de, e em conformidade com, o valor semântico (1, ½ ou 0) da frase φ nomeada por t. Para simplificar, supomos que $V(x)$ é o único predicado parcialmente definido da linguagem; todos os outros são predicados totais. Aliás, toda a restante linguagem tem uma semântica perfeitamente clássica. Podemos, por isso, distinguir duas linguagens: uma linguagem-de-base L_B (que é apenas L sem o predicado de verdade), que tem um modelo clássico M_B; e a linguagem L, para a qual procuramos um modelo M que seja uma expansão de M_B. Para construir esse modelo para L, a questão crucial é encontrar uma interpretação adequada para o predicado $V(x)$. Como é um pre-

[4] Kripke menciona também, como alternativa às regras de Kleene, as regras da semântica supervalorativista proposta por van Fraassen (1966).

dicado parcial, $V(x)$ deverá ser interpretado por um par de conjuntos $<E, A>$, em que E é a sua extensão e A é a sua anti-extensão. Mas, como o predicado $V(x)$ se aplica também a frases que o contêm, a sua interpretação terá de ser construída por um processo sequencial infinito: por exemplo, na etapa inicial, I_0, nenhuma frase pertence à extensão do predicado $V(x)$; na etapa seguinte, I_1, frases atómicas e não-semânticas verdadeiras como '1=1' ou 'A neve é branca' pertencem à extensão de $V(x)$; na etapa seguinte, I_2, frases como '"1=1" é verdadeira' pertencem à extensão de $V(x)$; na etapa seguinte, I_3, frases como '"'1=1' é verdadeira" é verdadeira' pertencem à extensão de $V(x)$; e assim sucessivamente. O processo é monotónico, no sentido em que, a partir do momento em que uma nova frase entra na extensão de $V(x)$, já nunca mais de lá sai. As extensões (e as anti-extensões) do predicado que formam esta sequência são provisórias. No final, uma delas será escolhida como extensão definitiva.

A semântica de L é dada através de uma definição indutiva de valor semântico *relativamente a uma extensão-provisória* $<E, A>$ para as fórmulas φ de L. Essa definição incluirá cláusulas do seguinte tipo:

(*i*) Se Π é um predicado de *n* lugares de L, então $\Pi(t_1,\ldots,t_n)$ tem valor 1 relativamente a $<E, A>$ se existem objectos $o_1,\ldots,o_n$ referidos respectivamente por $t_1,\ldots,t_n$ tais que $<o_1,\ldots,o_n>$ pertence à extensão de Π; caso contrário, tem valor 0 relativamente a $<E, A>$.

(*ii*) $V(t)$ tem valor 1 relativamente a $<E, A>$ se existe um objecto o referido por t tal que $o \in E$; e tem valor 0 relativamente a $<E, A>$ se existe um objecto o referido por t tal que $o \in A$; nos outros casos, tem valor ½ relativamente a $<E, A>$.

(*iii*) $\neg\varphi$ tem valor 1 relativamente a $<E, A>$ se φ tem valor *o* relativamente a $<E, A>$; e tem valor 0 relativamente a $<E, A>$ se φ tem valor 1 relativamente a $<E, A>$; nos outros casos, tem valor ½ relativamente a $<E, A>$.

(*iv*) $\varphi \vee \psi$ tem valor 1 relativamente a $<E, A>$ se pelo menos uma das sub-frases, φ e ψ, tem valor 1 relativamente a $<E, A>$; e tem valor 0 relativamente a $<E, A>$ se φ e ψ têm ambas valor 0 rela tivamente a $<E, A>$; nos outros casos, tem valor ½ relativamente a $<E, A>$.

Com este género de definição, podemos construir sequências infinitas de extensões-provisórias, as quais exibem como característica mais interessante o facto de conterem *pontos fixos*. Suponhamos que a sequência começa com um par $<E_0, A_0>$ de conjuntos vazios. Na etapa seguinte, E_1 é o conjunto das frases que têm valor 1 relativamente a $<E_0, A_0>$ e A_1 é o conjunto dos objectos que não são frases e das frases que têm valor 0 relativamente a $<E_0, A_0>$. Em geral, para qualquer ordinal α, $E_\alpha+1$ é o conjunto das frases que têm valor 1 relativamente a $<E_\alpha, A_\alpha>$. E, para um ordinal limite λ, E_λ é o conjunto das frases que têm valor 1 relativamente a algum $<E_\beta, A_\beta>$ tal que $\beta<\lambda$. Os conjuntos E e A que formam cada par vão crescendo ao longo da sequência, mas não podem crescer indefinidamente (pois o conjunto das frases de L tem uma cardinalidade fixa, contável). Há um ordinal α para o qual $<E_\alpha+1, A_\alpha+1> = <E_\alpha, A_\alpha>$; e diz-se, nesse caso, que $<E_\alpha, A_\alpha>$ é um ponto fixo. Na realidade, existe uma grande multiplicidade de pontos fixos. Mas o mais interessante é o ponto fixo mínimo, que se obtém quando o processo começa, tal como indicámos, com dois conjuntos vazios $<E_0, A_0>$. Podemos seleccionar o ponto fixo mínimo como extensão definitiva para o predicado $V(x)$. No ponto fixo $<E_\alpha, A_\alpha>$, as frases que têm valor 1 relativamente a $<E_\alpha, A_\alpha>$ coincidem com as frases que pertencem a E_α e as frases que têm valor 0 relativamente a $<E_\alpha, A_\alpha>$ coincidem com as frases que pertencem a A_α. Mas as frases paradoxais não pertencem a E_α nem a A_α; e têm valor ½ relativamente a $<E_\alpha, A_\alpha>$.

Deste modo, Kripke mostrou que uma linguagem pode conter o seu próprio predicado de verdade e preservar a consistência, desde que a sua semântica tenha três valores – 1, 0 e ½ (ou verdadeiro, falso e indefinido) – e que as frases paradoxais recebam o valor ½ (ou indefinido). De facto, a análise do paradoxo do mentiroso mostrou que, no contexto de uma teoria bivalente clássica, é impossível atribuir ao predicado $V(x)$ uma extensão que esteja em conformidade com a Convenção V, quer dizer, uma extensão que seja tal que, para qualquer frase φ, φ é verdadeira (ou falsa) exactamente quando $V([\varphi])$ é verdadeira (ou falsa). Pois, se λ é uma frase mentirosa, verificamos que, quando atribuímos a λ um valor de verdade, somos forçados a atribuir a $V([\lambda])$ precisamente o valor oposto. Mas, numa teoria trivalente (com a lógica de Kleene), o problema parece solucionável. No ponto fixo mínimo, o predicado $V(x)$ recebe uma extensão e uma anti-extensão que são tais que, para qualquer frase φ, φ é verdadeira, falsa ou indefinida exactamente quando $V([\varphi])$ é verdadeira, falsa ou indefinida.

Porém, a construção de Kripke, sem mais elementos, está longe de proporcionar uma teoria da verdade que seja satisfatória. Podemos apontar-lhe duas deficiências principais.

Em primeiro lugar, a lógica de Kleene é demasiado fraca. É uma lógica que não aceita a lei do terceiro excluído, $\varphi \vee \neg\varphi$. Pois quando φ é indefinida (nomeadamente, quando se trata de uma frase paradoxal), a sua negação $\neg\varphi$ também é indefinida; e a disjunção de duas frases indefinidas é ela própria indefinida. Mas a fraqueza desta lógica não reside apenas nisto. Na realidade, trata-se de uma lógica que não reconhece a existência de *nenhuma verdade lógica*.[5] Por exemplo, nesta lógica, $\varphi \rightarrow \varphi$ e $\varphi \leftrightarrow \varphi$ não são verdades lógicas. No presente contexto, este facto é especialmente grave. Pois, embora tenhamos conseguido recuperar a equivalência entre φ e $V([\varphi])$ – mais exactamente: a sua intersubstituibilidade em todos os contextos não-opacos –, o esquema V de Tarski, $\varphi \leftrightarrow V([\varphi])$, deixou de ser uma verdade lógica. Na teoria de Kripke, algumas bicondicionais de Tarski não são verdadeiras.

A fraqueza da lógica de Kleene é ainda visível no facto de existirem trivialidades que gostaríamos de ver incluídas numa teoria da verdade, mas que, quando formuladas nesta lógica, não são aceites como verdadeiras. Eis dois exemplos dessas trivialidades: 'Todas as frases verdadeiras são verdadeiras'[6] e 'Uma conjunção é verdadeira se as frases que a compõem são ambas verdadeiras'.[7]

[5] Um modelo em que todas as frases atómicas da linguagem sejam indefinidas é, com esta lógica, um modelo em que todas as frases são indefinidas. Como esse modelo existe, não há nenhuma frase que seja verdadeira em todos os modelos.

[6] Se λ é indefinida, então $V([\lambda])$ também é indefinida. Então a condicional $V([\lambda]) \rightarrow V([\lambda])$ também é indefinida. Então a generalização $\forall x\ (V(x) \rightarrow V(x))$ também é indefinida.

[7] Se μ e ν são ambas indefinidas, então a sua conjunção também é indefinida. Então a condicional $(V([\mu]) \wedge V([\nu])) \rightarrow V([\mu \wedge \nu])$ também é indefinida. Então a

Em segundo lugar, a teoria da verdade de Kripke tem graves limitações expressivas. Em particular, a teoria parece ser incapaz de caracterizar semanticamente as frases paradoxais. Para apreciar este ponto, é importante recuperarmos a distinção entre linguagem-objecto e metalinguagem. Na metalinguagem, dizemos que as frases paradoxais são indefinidas ou que têm valor semântico ½ (no ponto fixo). Mas como é que isto poderia ser dito na linguagem-objecto? Na realidade, a linguagem-objecto, apesar de conter o seu próprio predicado de verdade – e de, por via dele, poder-se definir um predicado de falsidade,[8] bem como um predicado de nem-verdade-nem-falsidade e outros do mesmo género –, não contém nenhum predicado semântico com o qual possamos traduzir a afirmação metalinguística de que uma frase mentirosa é indefinida. É importante observar que, quando λ é uma frase mentirosa, $\neg V(\lambda) \wedge \neg F(\lambda)$ não pode ser a caracterização semântica correcta, na linguagem-objecto, de λ. Pois $\neg V(\lambda) \wedge \neg F(\lambda)$ implica $\neg V(\lambda)$; mas $\neg V(\lambda)$ é (ou é equivalente a) λ e, por isso, não pode ser aceite como verdadeira.

Este último problema também pode ser formulado usando as noções de aceitação e de rejeição. Se, como é natural, fizermos coincidir as frases que a teoria de Kripke aceita com as frases que estão no ponto fixo (quer dizer, as frases que pertencem à extensão do predicado $V(x)$ no ponto fixo), vemos que, nesta teoria, a rejeição de uma frase nem sempre coincide com a aceitação da sua negação. Por exemplo, a teoria rejeita a frase mentirosa λ e também rejeita a sua negação $\neg\lambda$; analogamente, a teoria rejeita a frase $V(\lambda)$ e também rejeita a sua negação $\neg V(\lambda)$. Em geral, uma teoria que rejeita a lei do terceiro excluído requer uma noção de rejeição que seja distinta da aceitação da negação. Na medida em que não a coloca no ponto fixo, é claro que a teoria de Kripke rejeita a frase mentirosa (assim como as restantes frases paradoxais). Mas como é que a teoria poderia *expressar* essa rejeição? Não parece haver nenhuma frase através de cuja aceitação a teoria pudesse expressar a sua rejeição da frase mentirosa. Se, como dissemos no início desta secção, a tese filosófica subjacente à teoria de Kripke é a de que as frases mentirosas, apesar de terem significado, não expressam nenhuma proposição, vemos que essa é

generalização $\forall x \, \forall y \, ((V(x) \wedge V(y)) \rightarrow V(\ulcorner x \wedge y \urcorner))$ também é indefinida.

[8] '*Falso*(x)' define-se, habitualmente, como '*Verdadeiro*($\ulcorner \neg x \urcorner$)'.

precisamente uma tese que a teoria não consegue expressar, na medida em que não consegue caracterizar a deficiência semântica que identifica nas frases mentirosas.

Poderia a teoria ser suplementada com a introdução de um operador primitivo N que expressasse a rejeição (e, mais do que isso, que a expressasse de uma maneira susceptível de ser encaixada numa parte de uma frase molecular, por exemplo na antecedente de uma condicional)? Suponhamos que, diferentemente da negação vulgar, N é um operador que, quando prefixado a uma frase φ com valor 0 ou ½, gera uma frase $N\varphi$ com valor 1 (e quando prefixado a uma frase com valor 1 gera uma frase com valor 0). Podemos chamar a N um operador de *negação forte*. Ora, verifica-se facilmente que, se a teoria fosse enriquecida com um tal operador, tornar-se-ia inconsistente. O operador N permitiria formar uma frase *fortemente mentirosa*, quer dizer, uma frase λ_F idêntica a $NV([\lambda_F])$, que nega-fortemente a sua própria verdade. Mas $NV([\lambda_F])$ teria valor semântico 1 se e somente se λ_F *não* tivesse valor semântico 1.

4 Salvando a lógica clássica

É razoável pensar-se que as insuficiências encontradas na teoria da verdade de Kripke se devem ao seu abandono da lógica clássica. (Aliás, o próprio Kripke manifestou insatisfação relativamente a esse abandono.) Nesta secção, iremos apresentar brevemente algumas outras teorias que tentam atingir o mesmo objectivo – a saber: o de mostrar que uma linguagem consistente pode conter o seu próprio predicado de verdade –, mas mantendo a lógica clássica.

4.1 *A teoria de Kripke-Feferman*

Um aspecto saliente da teoria de Kripke é o facto de se tratar de uma teoria puramente semântica. Kripke descreveu uma linguagem auto-referencial, dotou-a de um predicado de verdade aplicável às suas próprias frases e mostrou como se podem construir modelos (trivalentes) para essa linguagem. Mas nada disse sobre o sistema dedutivo que lhe estaria associado. Posteriormente, Solomon Feferman formulou uma teoria axiomática da verdade, cuja lógica é clássica e que pode ser vista como uma axiomatização da teoria semântica de

Kripke. Esta teoria é geralmente conhecida como KF (de 'Kripke--Feferman') e foi publicada pela primeira vez em Reinhardt 1986.

Por limitações de espaço, não apresentaremos aqui os pormenores da teoria KF (que podem ser consultados em Halbach 2011 e Horsten 2011). Do ponto de vista formal, é uma teoria com propriedades que a recomendam: é ómega-consistente, tem modelos clássicos baseados nos números naturais, é aritmeticamente correcta (*sound*) e é relativamente forte quando comparada com outras teorias axiomáticas. Porém, enquanto teoria da verdade, parece ser bastante insatisfatória.

Na teoria KF, é válida uma regra de inferência – a que podemos chamar regra da Eliminação da Verdade – que permite, para qualquer frase φ, inferir φ a partir de $V([\varphi])$. Por isso, em geral, se é possível demonstrar que φ é verdadeira, então φ também é um teorema. Mas KF consegue ser consistente à custa de não validar a regra de inferência inversa, da Introdução da Verdade. Ou seja, existem na teoria frases φ que exibem a seguinte característica: φ é demonstrável, mas $V([\varphi])$ não é demonstrável. Consequentemente, em KF, φ e $V([\varphi])$ não são intersubstituíveis em todos os contextos (não-opacos) e a condicional $\varphi \rightarrow V([\varphi])$ não é um teorema.

Como a lógica de KF é clássica, todas as instâncias do terceiro excluído são nela demonstráveis, incluindo, quando λ é uma frase mentirosa, a disjunção $\lambda \vee \neg\lambda$ e a disjunção $V([\lambda]) \vee \neg V([\lambda])$. Mas estas duas instâncias do terceiro excluído são demonstráveis por razões assimétricas: a primeira é-o porque λ é demonstrável e a segunda é-o porque $\neg V([\lambda])$ é demonstrável. De facto, em KF, a partir do lema da diagonalização, pode facilmente demonstrar-se que $\lambda \wedge \neg V([\lambda])$.

Este é um ponto importante de diferença: enquanto na teoria semântica de Kripke, a frase $\neg V([\lambda])$, que afirma que *a frase mentirosa não é verdadeira*, não podia ser aceite como verdadeira, na teoria KF, essa frase é aceite, pois é demonstrável. Isso poderia parecer uma vantagem de KF, até observarmos que, ao aceitar λ e também $\neg V([\lambda])$, a teoria demonstra coisas que ela própria declara não serem verdadeiras. Ora, uma teoria que demonstra que alguns dos seus teoremas não são verdadeiros, embora seja consistente e aritmeticamente correcta, não é certamente correcta enquanto teoria da verdade. (No entanto, Maudlin 2004 faz uma defesa de uma teoria exactamente com estes contornos.)

4.2 *Teorias da revisão*

Como alternativa à teoria de Kripke, Herzberger (1982a, 1982b), Gupta (1982) e Belnap (1982) apresentaram a chamada 'teoria da revisão' (*revision theory of truth*) (veja-se também Gupta e Belnap 1993), que regressa a uma abordagem dos paradoxos num contexto bivalente, em que a anti-extensão de cada predicado é o complemento da sua extensão. Tal como Kripke, os teóricos da revisão enfrentam o problema de saber como expandir um modelo clássico para uma linguagem-de-base, de modo a construir um modelo para essa linguagem enriquecida com o seu próprio predicado de verdade. Enquanto para Kripke a ideia central era a de interpretar o predicado $V(x)$ de maneira não-bivalente através de um par de conjuntos $<E, A>$ sem sobreposição e que não esgotam o domínio, para os teóricos da revisão a ideia central é a de interpretar o predicado $V(x)$ através de uma *regra de revisão*, em vez de uma regra de aplicação.

Quando dizemos, por exemplo, que uma frase como '*Azul*(x)' é verdadeira num modelo se e somente se o objecto que o modelo atribui a 'x' é azul, estamos a formular uma regra de aplicação, que atribui ao predicado '*azul*' um conjunto de coisas de que ele é verdadeiro. A teoria da revisão considera que esse tipo de interpretação não serve para o predicado de verdade (devido à circularidade que exibe quando permitimos que ele se aplique a frases que o contêm).[9] Para este, devemos antes usar uma regra de revisão, que é uma regra que diz, dada uma qualquer extensão hipotética para o predicado, como corrigi-la, gerando uma extensão melhor ou mais adequada. Gupta sugeriu que:

> Quando aprendemos o significado de 'verdadeiro', o que aprendemos é uma regra que nos permite *melhorar* a partir de um candidato proposto para a extensão da verdade. Pretendo defender que é a existência de uma regra assim que explica os traços característicos do conceito de verdade. (Gupta 1982: 212)

Para aplicarmos a regra de revisão, precisamos de ter uma extensão hipotética como ponto de partida. A teoria é completamente liberal a respeito da escolha da hipótese inicial: qualquer conjunto (de objectos do domínio) serve. O ponto importante é que, por muito má

[9] Nalguns casos, precisaríamos de conhecer já a extensão do predicado para decidirmos se uma frase pertence ou não pertence a essa extensão.

que seja a hipótese inicial (por exemplo, se for a hipótese de que tudo é verdadeiro), o processo de revisão vai gerar uma hipótese melhor, que por sua vez será revista, e assim sucessivamente, ao longo de uma sequência infinita. A regra de revisão usada é muito simples. Trata-se, na realidade, de usar as frases-V de Tarski para, dado um modelo de base M_B e uma extensão hipotética H_0 para o predicado $V(x)$, gerar uma extensão revista H_1 que é o conjunto das frases que são verdadeiras no modelo M_B+H_0. Por exemplo, se '1=1' não estava em H_0, ela está em H_1, pois é uma frase verdadeira no modelo de base; e se '2+2=3' estava em H_0, já não está em H_1, pois é uma frase falsa no modelo de base. Todas as frases não-semânticas são corrigidas logo na primeira etapa (i.e., todas as verdades não-semânticas estão em H_1 e nenhuma falsidade não-semântica está em H_1). Uma vez obtida a extensão H_1, ela é revista, construindo-se o conjunto H_2 das frases que são verdadeiras no modelo M_B+H_1. Por exemplo, uma frase como 'V('1=1')' entra seguramente em H_2 (se não estava já nas extensões anteriores).

Suponhamos, por exemplo, que a linguagem contém as seguintes frases:

A = '$V([B]) \vee V([C])$'
B = '$V([A])$'
C = '$\neg V([A])$'
D = '$\neg V([D])$'

Suponhamos ainda que, de acordo com a hipótese inicial, H_0, o predicado V(x) tem extensão nula (nada é verdadeiro). Então, no modelo M_B+H_0, as frases 'V([A])', 'V([B])', 'V([C])' e 'V([D])' são todas avaliadas como falsas; e, consequentemente:

Val_0(A) = falsa
Val_0(B) = falsa
Val_0(C) = verdadeira
Val_0(D) = verdadeira

A hipótese revista, H_1, dirá então que as frases C e D pertencem à extensão do predicado V(x), pelo que, no modelo M_B+H_1, as frases 'V([A])' e 'V([B])' são falsas, mas 'V([C])' e 'V([D])' são verdadeiras.

Consequentemente:

Val_1(A) = verdadeira
Val_1(B) = falsa
Val_1(C) = verdadeira
Val_1(D) = falsa

Se continuarmos o processo de revisão, obtemos como resultado a seguinte tabela:

	Val_0	Val_1	Val_2	Val_3	Val_4	...
A	f	v	v	v	v	...
B	f	f	v	v	v	...
C	v	v	f	f	f	...
D	v	f	v	f	v	...

Podemos observar que, após uma primeira avaliação como falsa, a frase A torna-se verdadeira na 1ª etapa e mantém esse valor daí em diante. As frases B e C estabilizam na 2ª etapa. Diz-se então que, nesta sequência, A e B são *estavelmente verdadeiras e C é estavelmente falsa*. Mas a frase paradoxal D nunca estabiliza. Por mais que a sequência se prolongue, ela continuará a oscilar de um valor para o outro. Mais do que isso, é fácil ver que D será *instável* em todas as sequências de revisão geradas por esta regra.

Quão longas deverão ser as sequências de revisão? Para encontrar uma resposta, consideremos um exemplo um pouco mais complexo.[10] Suponhamos que a linguagem contém a seguinte sequência de frases:

A_0 = 'V('1 = 1')'
A_1 = 'V([A_0])'
A_2 = 'V([A_1])'
A_3 = 'V([A_2])'
...

[10] Tanto este como o exemplo anterior são adaptados de Kremer 2006.

e, em geral, para cada número natural *n*,

A_n = ‘V([A_{n-1}])’

Além disso, há um predicado G(x) que, no modelo de base, tem como extensão o conjunto {A_0, A_1, A_2, A_3, ...}; e há uma frase B que é idêntica a ‘∀x (G(x) → V(x))’. Se supusermos que a hipótese inicial deixa vazia a extensão de V(x), a sequência de revisão começa com um modelo M_B+H_0 no qual todas as frases do conjunto {A_0, A_1, A_2, A_3, ..., B} são falsas. Mas, depois, a frase A_0 estabiliza como verdadeira na etapa 1; a frase A_1 estabiliza como verdadeira na etapa 2; a frase A_2 estabiliza como verdadeira na etapa 3; e, em geral, para cada *n*, a frase A_n estabiliza como verdadeira na etapa *n*+1. No entanto, em todas as etapas *finitas*, a frase B mantém-se falsa. Para que B possa ser avaliada como verdadeira (como intuitivamente reconhecemos que deve), o processo de revisão tem de continuar para lá das etapas finitas, gerando uma sequência transfinita. A tabela seguinte representa a sequência que então se obtém:

	Val_0	Val_1	Val_2	Val_3	Val_4	...	Val_ω	$Val_{\omega+1}$	$Val_{\omega+2}$	...
A_0	f	v	v	v	v	...	v	v	v	...
A_1	f	f	v	v	v	...	v	v	v	...
A_2	f	f	f	v	v	...	v	v	v	...
A_3	f	f	f	f	v	...	v	v	v	...
A_4	f	f	f	f	f	...	v	v	v	...
...	...	...	...	...	...	...	...	...	...	...
B	f	f	f	f	f	...	f	v	v	...

A consideração de sequências de revisão transfinitas levanta um problema novo, que é o de saber o que devemos fazer numa etapa limite (que é maior do que todas as etapas anteriores, mas não é o sucessor imediato de nenhuma delas). Se uma frase já estabilizou nas etapas anteriores, então deve receber na etapa limite o valor (v ou f) no qual estabilizou. E se não estabilizou? Não há concordância a este respeito entre os teóricos da revisão. Por exemplo, Herzberger propôs que, nesses casos, a frase seja considerada falsa na etapa limite; enquanto, segundo Gupta, ela deveria receber o mesmo valor que lhe foi atribuído pela hipótese inicial. A teoria mais simples foi proposta por Belnap e diz que, se a frase não estabilizou em nenhum dos dois valores nas etapas anteriores, então a etapa limite pode atribuir-lhe qualquer um deles. Além destas, existem outras versões possíveis da teoria da revisão (veja-se, sobre isso, Field 2008: 187).

Tendo por base um modelo M_B para a linguagem de base e uma hipótese inicial H_0 a respeito da extensão do predicado $V(x)$, uma sequência de revisão é então uma sequência transfinita de modelos bivalentes $<M_B+H_0, M_B+H_1, M_B+H_2, \ldots, M_B+H_\omega, M_B+H_{\omega+1}, \ldots, MB+H_{\omega\times2}, M_B+H_{(\omega\times2)}+1, \ldots>$ que interpretam a linguagem enriquecida com o seu próprio predicado de verdade. A parte M_B é comum a todos os modelos e, por isso, podemos simplificar e definir uma sequência de revisão como uma sequência transfinita de hipóteses ou de conjuntos H_α, cada um dos quais é um subconjunto do domínio de M_B e serve como extensão do predicado $V(x)$ na α-ésima etapa do processo de revisão. Se uma frase é verdadeira num modelo M_B+H_α, podemos simplificadamente dizer que ela é verdadeira relativamente a H_α. Nas etapas sucessoras, o conjunto $H_{\alpha+1}$ tem como membros as frases verdadeiras relativamente a H_α. A noção de *verdade estável* ou, mais exactamente, de uma frase ser estavelmente verdadeira (ou falsa) numa sequência de revisão é aqui uma noção central. Dada uma sequência *S* de tamanho γ, em que γ é um ordinal limite, dizemos que uma frase *A* é *estavelmente verdadeira* (ou *estavelmente falsa*) em *S* se e somente se, para algum ordinal β menor do que γ, *A* é verdadeira (ou falsa) relativamente a H_α, para todo o ordinal $\alpha \geq \beta$.

A teoria da revisão simples é uma teoria que aceita todas as frases que sejam estavelmente verdadeiras em todas as sequências de revisão e não aceita nada mais. É então evidente que, se λ é uma frase mentirosa, a teoria não aceita λ nem $\neg\lambda$, tal como não aceita $V([\lambda])$ nem $\neg V([\lambda])$. No entanto, como é uma teoria clássica, ela aceita as disjunções $\lambda \vee \neg\lambda$ e $V([\lambda]) \vee \neg V([\lambda])$. Quer dizer que, para esta teoria, há disjunções 'infundadas', que são verdadeiras sem que nenhum dos disjuntos seja verdadeiro. Outro exemplo importante é o da disjunção

$$(\lambda \wedge \neg V([\lambda])) \vee (\neg\lambda \wedge V([\lambda]))$$

A teoria da revisão aceita esta disjunção, pois ela é equivalente à bicondicional $(\lambda \leftrightarrow \neg V([\lambda]))$, classicamente demonstrável a partir do lema da diagonalização. No entanto, ao contrário da teoria de Kripke-Feferman, ela não aceita nenhum dos seus disjuntos. Esta ligeira vantagem não pode esconder um ponto fundamental em comum: a rejeição das frases-V (convertidas em regras de revisão, mas não aceites enquanto bicondicionais) e da intersubstituibilidade de φ e $V([\varphi])$ em todos os contextos. Para alguns autores, esta intersubstituibilidade é a *razão de ser* do predicado de verdade.

4.3 *Teorias contextualistas*

As teorias contextualistas propostas por autores como Parsons (1974), Burge (1979), Barwise e Etchemendy (1987), Simmons (1993), Goldstein (1992 e 2000), Glanzberg (2001 e 2004) têm como traço comum a ideia de que na origem dos paradoxos semânticos está um problema de dependência contextual, um problema que envolve a relação entre a verdade e o contexto. Tal como a frase 'Eu vou para Lisboa' exibe comportamentos semânticos diferentes em contextos diferentes, estes autores defendem que se dá o mesmo género de fenómeno com as frases mentirosas e outras do mesmo tipo.

Perseguindo o mesmo objectivo de Kripke (a saber: o de mostrar que uma linguagem consistente pode conter o seu próprio predicado de verdade) e procurando fazê-lo sem abandonar a lógica clássica, estas teorias são também favoráveis à abordagem que atribui às frases paradoxais uma falha de valor de verdade. No entanto, elas consideram que a principal dificuldade que essa abordagem tem de enfrentar é o chamado problema do 'mentiroso reforçado'. A diferença entre o mentiroso simples e o mentiroso reforçado seria a seguinte: no primeiro, o paradoxo é gerado por uma frase L idêntica a 'L é falsa', enquanto, no segundo, ele é gerado por uma frase L_R idêntica a 'L_R não é verdadeira'; enquanto o primeiro caso se pode resolver declarando que L não é verdadeira nem falsa, o segundo não se deixa imediatamente vencer por essa estratégia – pois, se L_R não é verdadeira nem falsa, então L_R não é verdadeira; mas, como isso é precisamente o que L_R diz, chega-se assim à conclusão de que L_R é verdadeira, ao mesmo tempo que não é verdadeira. Os contextualistas julgam que a solução para este problema envolve o reconhecimento da indexicalidade presente na frase L_R.

Diferentes teorias contextualistas identificam de maneiras diferentes aquilo que há de supostamente indexical nas frases mentirosas. Outro traço que diferencia as diversas teorias contextualistas é a selecção dos portadores de verdade (*truth bearers*) apropriados. Enquanto umas consideram que a verdade e a falsidade são primariamente propriedades de proposições, outras afirmam que elas são propriedades de espécimes (*tokens*) de frases ou, ainda, de actos de asserção (*statements*). Para efeitos de exposição, iremos aqui supor que a verdade é uma relação entre uma frase e um contexto. Diremos que, em geral, uma frase *A não tem um valor de verdade absoluto; ao invés, A* poderá ser verdadeira nalguns contextos, e falsa noutros contextos, e nem verdadeira nem falsa noutros contextos ainda. Alegadamente, esta relativização da verdade a um contexto proporcionaria uma saída satisfatória, e com valor explicativo, para o problema do mentiroso reforçado.

Se examinarmos mais atentamente o raciocínio envolvido no mentiroso reforçado, vemos que ele tem duas partes principais. Na primeira parte, partimos de um facto simples a respeito da identidade da frase mentirosa e concluímos que ela não é verdadeira nem falsa; depois, na segunda parte, partimos da conclusão a que chegámos na primeira parte e concluímos que, afinal, a frase mentirosa é verdadeira e não é verdadeira. O diagnóstico aqui proposto é o de que, entre estas duas partes do raciocínio, existe uma *mudança de contexto* – e que, se a tomarmos devidamente em conta, podemos ver que a conclusão a que se chegou não é realmente contraditória. Simplesmente, a frase mentirosa não é verdadeira no primeiro contexto e é verdadeira no segundo contexto. O primeiro contexto é o contexto no qual a frase mentirosa foi originalmente proferida. (Simmons, por exemplo, concebe uma situação em que Aristóteles, convencido de que aquilo que Platão escreveu no quadro na sala ao lado é falso, escreve no quadro 'A frase escrita no quadro na sala 101 não é verdadeira'. Mas Aristóteles está enganado a respeito das salas, pois ele é que está na sala 101.) É fácil observar que, *nesse contexto*, a frase mentirosa não é verdadeira (nem falsa). Mas, quando expressamos este facto, usamos a própria frase mentirosa (pois dizemos 'a frase mentirosa não é verdadeira'), quer dizer, proferimo-la assertivamente, manifestando assim que a aceitamos como verdadeira. Simplesmente, afirmam os contextualistas, este segundo uso da frase ocorre já *noutro contexto* (fora da sala 101, provavelmente), num contexto em que estamos a reflectir sobre o primeiro uso. Ora, não é contraditório afirmar (provavelmente num terceiro contexto, E) que a frase L_R não é verdadeira no contexto C (por exemplo, escrita por Aristóteles na sala 101) e é verdadeira no contexto D (por exemplo, escrita por Kripke num artigo em que defende a existência de falhas de valor de verdade).

Embora seja interessante, a proposta de soluções deste género para o paradoxo tem sido recebida com justificado cepticismo. As razões de insatisfação são várias. Em primeiro lugar, não basta invocar uma mudança de contexto para justificar a variação no valor semântico de diferentes elocuções da mesma frase. Precisamos também de identificar, na estrutura da frase, quais são os elementos que, por serem sensíveis a essa mudança, são responsáveis pela variação. Ora, neste caso, se compararmos a frase 'L_R não é verdadeira' proferida no contexto C com a frase 'L_R não é verdadeira' proferida no contexto D, observamos que, em ambos os contextos, a frase referida é a mesma e a propriedade que lhe é atribuída é também aparentemente a mesma. É então um mistério porque é que as duas elocuções teriam valores semânticos diferentes.

Além disso, mesmo que admitamos que *pode* haver uma mudança de contexto, não é seguro que *tenha de haver* uma mudança de contexto. Em princípio, não parece haver nada que impeça que a reflexão sobre uma elocução da frase mentirosa se faça *no mesmo contexto* em que ocorre a elocução, ou que se construa uma frase mentirosa cuja elocução seja auto-reflexiva (quer dizer, uma frase que diga de si mesma que não é verdadeira no contexto C, sendo C exactamente o contexto no qual ela é proferida). Mais do que isso, se os recursos expressivos à disposição do teórico contextualista – nomeadamente, a sua capacidade de falar dos diversos contextos possíveis de elocução da frase mentirosa – estiverem acessíveis ao utilizador da linguagem (e se não exigirmos, como fez Tarski, que a metalinguagem seja essencialmente mais rica do que a linguagem-objecto), poderemos construir uma frase super-mentirosa que diga de si mesma que não é verdadeira *em absolutamente nenhum contexto*. E, assim, teremos o paradoxo de volta.

5 Salvando a transparência

Muitas coisas que dizemos usando o predicado de verdade podem igualmente ser ditas sem ele. Por exemplo, em vez de dizermos que 'Aristóteles morreu com 73 anos' é uma frase verdadeira, podemos simplesmente dizer que Aristóteles morreu com 73 anos. Mas há algumas coisas que, aparentemente, só podemos dizer usando o predicado de verdade, como generalizações do género de 'Todas as condicionais materiais com consequente verdadeira são verdadeiras' ou afirmações do tipo de 'Se tudo o que está dito no relatório da comissão de inquérito é verdadeiro, então não houve nenhum atentado em Camarate'. Neste último exemplo, a alternativa seria reproduzir todo o conteúdo do relatório na antecedente da condicional – o que, além de pouco prático, podemos por razões contingentes (destruição do relatório num incêndio, limitações de memória, etc.) ser incapazes de fazer. Para que o predicado de verdade possa desempenhar esta função, é necessário que a atribuição de verdade a uma frase seja sempre equivalente a essa mesma frase, quer dizer, é necessário que, para toda a frase φ, $V([\varphi])$ e φ sejam intersubstituíveis em todos os contextos extensionais. Quando isto acontece, dizemos que o predicado de verdade é completamente transparente.

A transparência é um requisito que naturalmente se estende a todas as noções semânticas e não apenas à verdade. As noções de *ser verdadeiro de*, de *nomear* e de *definir*, que ocorrem nos paradoxos de Grelling, de Berry e de Richard, funcionam de maneira análoga. É com base na transparência que dizemos, por exemplo, que se o predicado 'azul' é verdadeiro do céu, então o céu é azul; e que, se o predicado 'heterológico' é verdadeiro dele próprio, então é heterológico.

Porém, em linguagens que contenham o seu próprio predicado de verdade, a transparência entra em conflito com a lógica clássica. É impossível respeitar ambas. Essas linguagens possuem recursos para construir frases mentirosas, quer dizer, frases λ que são equivalentes a $\neg V([\lambda])$. A transparência, por seu lado, requer que λ, como qualquer outra frase, seja equivalente a $V([\lambda])$. Então $V([\lambda])$ tem de ser equivalente a $\neg V([\lambda])$, o que é impossível na lógica clássica.

As teorias que pretendem admitir linguagens com o seu próprio predicado de verdade e manter a lógica clássica vêem-se, por isso, forçadas a limitar a transparência da verdade. A opção alternativa consiste em manter a transparência completa da verdade e restringir a lógica clássica. Nesta secção, iremos caracterizar brevemente os dois principais tipos de teoria que tomam esta opção.

5.1 *Teorias paracompletas*

Chamam-se 'paracompletas' as teorias que não aceitam a lei do terceiro excluído. A teoria de Kripke, pelo menos da maneira como a apresentámos, adoptando a lógica de Kleene, é uma teoria paracompleta que preserva a transparência da verdade. Vimos, no entanto, que ela tem uma lógica demasiado fraca, que não reconhece nenhuma verdade lógica, nem mesmo $\varphi \rightarrow \varphi$; e que, em consequência disso, há condicionais com a forma $V([\varphi]) \rightarrow \varphi$ ou com a forma $\varphi \rightarrow V([\varphi])$ que não são verdadeiras. Vimos, além disso, que a teoria de Kripke é também fraca quanto à sua capacidade expressiva, mostrando-se incapaz de caracterizar semanticamente as frases paradoxais.

Hartry Field propôs uma teoria paracompleta que tem como objectivo superar estas duas deficiências principais da teoria de Kripke (veja-se Field 2003 e 2008). A teoria de Field adopta como base a lógica trivalente de Kleene, mas depois acrescenta-lhe uma nova condicional $\Rightarrow$ que obedece à lei $\varphi \Rightarrow \varphi$. A linguagem fica assim com duas condicionais: por um lado, a condicional material $\varphi \rightarrow \psi$, que se define da maneira habitual pela negação e pela disjunção (como $\neg\varphi \vee \psi$) e que, como vimos, não obedece à lei $\varphi \rightarrow \varphi$ (que é equivalente à lei do terceiro excluído $\neg\varphi \vee \varphi$); por outro lado, uma nova condicional $\varphi \Rightarrow \psi$, que é uma conectiva primitiva, que não é verofuncional e que obedece à lei $\varphi \Rightarrow \varphi$. Uma vez que a transparência da verdade é preservada, este último aspecto tem como consequência que todas as bicondicionais de Tarski são verdadeiras na teoria de Field.

Para interpretar a linguagem enriquecida com esta nova condicional (além do seu próprio predicado de verdade), Field combina o método de Kripke com uma "regra de revisão" usada agora, não para o predicado de verdade, mas para a condicional. Em vez de uma sequência transfinita de modelos clássicos bivalentes (como acontecia nas teorias de Herzberger, Gupta e Belnap), a teoria é agora interpretada por meio de uma sequência transfinita de pontos fixos de Kripke. Recordemos que, na construção original de Kripke, o ponto fixo mínimo era gerado a partir de um modelo clássico para a linguagem de base e da atribuição do conjunto vazio como extensão e anti-extensão iniciais para o predicado de verdade. Agora, na adaptação de Field, cada ponto fixo tem também como ponto de partida uma atribuição de um dos valores 1, ½ ou 0 a todas as frases (simples ou complexas) cuja conectiva principal é $\Rightarrow$. Diremos que cada ponto fixo P^k é construído a partir de uma valoração inicial das condicionais C^k. Tal como antes, no ponto fixo, todas as frases têm um valor (1, ½ ou 0), que está em concordância com as regras composicionais de Kleene (para as conectivas extensionais) e que obedece ao princípio de transparência segundo o qual $V([\varphi])$ tem sempre o mesmo valor que φ. Os diversos pontos fixos que se obtêm diferenciam-se entre si porque partem de diferentes valorações iniciais das condicionais. Estas valorações são determinadas da seguinte maneira:

> *Regra de Base*: Na etapa de partida, a valoração inicial C^0 atribui a todas as condicionais o valor ½.
>
> *Regra para Sucessores*: Em cada etapa sucessora, C^{k+1} dá valor 1 a $\varphi \Rightarrow \psi$ se no ponto fixo da etapa anterior, P^k, o valor de φ for menor do que, ou igual a, o valor de ψ; caso contrário, C^{k+1} dá valor 0 a $\varphi \Rightarrow \psi$.
>
> *Regra para Limites*: Em cada etapa limite, C^l dá valor 1 a $\varphi \Rightarrow \psi$ se existir um ponto antes do limite tal que, depois desse ponto e até ao limite, o valor de φ é sempre menor do que, ou igual a, o valor de ψ; e C^l dá valor 0 a $\varphi \Rightarrow \psi$ se existir um ponto antes do limite tal que, depois desse ponto e até ao limite, o valor de φ é sempre maior do que o valor de ψ; e C^l dá valor ½ a $\varphi \Rightarrow \psi$ se nenhuma das condições anteriores se verificar.

Por fim, Field define o "valor último" de uma frase como sendo 1 se existir um ponto a partir do qual ela tenha sempre valor 1; 0 se existir um ponto a partir do qual ela tenha sempre valor 0; e ½ se nenhuma das condições anteriores se verificar. Field consegue provar que existem ordinais i tais que, para qualquer j diferente de zero, o valor de cada frase na etapa $i \times j$ é o mesmo que o seu valor último. E, o que é o mais importante, Field apresenta uma demonstração da consistência da teoria resultante.

A nova condicional de Field obedece a várias leis que conhecemos da lógica clássica, como, por exemplo, as seguintes:

$\models \varphi \Rightarrow \varphi$
$\models \varphi \Rightarrow \neg\neg\varphi$ e a sua conversa
$\models (\varphi \wedge \psi) \Rightarrow \varphi$
$\models \varphi \Rightarrow (\varphi \vee \psi)$
$\models \neg(\varphi \wedge \psi) \Rightarrow (\neg\varphi \vee \neg\psi)$
$\models (\varphi \wedge (\psi \vee \chi)) \Rightarrow ((\varphi \wedge \psi) \vee (\varphi \wedge \chi))$
$\models \forall x\varphi \Rightarrow \varphi(x/t)$ (para substituições apropriadas da variável)
$\varphi \Rightarrow \psi,\ \psi \Rightarrow \chi, \models \varphi \Rightarrow \chi$ (Transitividade)
$\varphi, \varphi \Rightarrow \psi \models \psi$ (Modus Ponens)
$\varphi \Rightarrow \psi \models \neg\psi \Rightarrow \neg\varphi$ (Contraposição)
$\forall x(\varphi \rightarrow \psi) \models \forall x\varphi \Rightarrow \forall x\psi$
$\varphi \Rightarrow \psi, \varphi \Rightarrow \chi \models \varphi \Rightarrow (\psi \wedge \chi)$
$\varphi \Rightarrow \chi, \psi \Rightarrow \chi \models (\varphi \vee \psi) \Rightarrow \chi$

Apesar de não ser uma condicional material (pois nem sequer é verofuncional), a nova condicional de Field está bastante próxima da condicional material. Por exemplo, ela obedece ao seguinte:

$\neg\varphi \vee \psi \models \varphi \Rightarrow \psi$, que pode ser reescrito como $\varphi \rightarrow \psi \models \varphi \Rightarrow \psi$

$\varphi, \neg\psi \models \neg(\varphi \Rightarrow \psi)$, que pode ser reescrito como $\neg(\varphi \rightarrow \psi) \models \neg(\varphi \Rightarrow \psi)$

Mais do que isso, verifica-se que, quando a antecedente e a consequente obedecem ambas à lei do terceiro excluído, as duas condicionais são equivalentes. Ou seja:

$$\varphi \vee \neg\varphi, \psi \vee \neg\psi \models (\varphi \rightarrow \psi) \leftrightarrow (\varphi \Rightarrow \psi)$$

Mas há também algumas leis clássicas a que a nova condicional de Field não obedece. Em particular, e felizmente, ela não obedece ao princípio da contracção:

$$\varphi \Rightarrow (\varphi \Rightarrow \psi) \not\models \varphi \Rightarrow \psi$$

Porquê "felizmente"? Porque, dado que as bicondicionais de Tarski são todas verdadeiras na teoria de Field e que a nova condicional obedece ao Modus Ponens, se ela obedecesse também ao princípio da contracção, cair-se-ia no paradoxo de Curry. Recordemos que uma frase de Curry é uma frase γ equivalente a $V([\gamma]) \Rightarrow 1=2$. Pela transparência, segue-se que $V([\gamma])$ também é equivalente a $V([\gamma]) \Rightarrow 1=2$; e, devido a essa equivalência, a condicional $V([\gamma]) \Rightarrow (V([\gamma]) \Rightarrow 1=2)$ tem valor último 1. Mas a consequente $V([\gamma]) \Rightarrow 1=2$, porque é uma frase paradoxal, tem valor último ½. Por isso, a contracção falha.

Para superar a segunda deficiência da teoria de Kripke (o silêncio a respeito do estatuto semântico das frases paradoxais), Field usa a nova condicional para definir um operador frásico D que permite expressar uma noção de *ser determinadamente* (tal ou tal), da seguinte maneira:

$$D\varphi =_{df} \varphi \wedge \neg(\varphi \Rightarrow \neg\varphi)$$

Definido assim, o operador de determinação obedece às seguintes leis:

$D\varphi \models \varphi$
$\varphi \models D\varphi$
$\models D\varphi \Rightarrow \varphi$
Se $\models \varphi \Rightarrow \neg\varphi$, então $\models \neg D\varphi$
Se $\models \varphi \Rightarrow \psi$, então $\models D\varphi \Rightarrow D\psi$

embora não obedeça a $\varphi \Rightarrow D\varphi$.

Com este operador, a teoria de Field dispõe de uma noção mais forte de verdade, com a qual já pode caracterizar a deficiência semântica das frases paradoxais e expressar a sua rejeição dessas frases. De facto, dada uma frase mentirosa λ equivalente a $\neg V([\lambda])$, a teoria aceita como verdade que λ *não é determinadamente verdadeira nem determinadamente falsa*, ou seja, que $\neg DV([\lambda]) \wedge \neg DF([\lambda])$, o que, dada a transparência da verdade e da falsidade, é equivalente a $\neg D\lambda \wedge \neg D\neg\lambda$.

Além disso, a teoria não está sujeita a fenómenos de "vingança", criados por frases super-mentirosas que dizem de si mesmas que não são determinadamente verdadeiras. Pois, se ξ é uma frase equivalente a $\neg D\xi$, a teoria rejeita essa frase e expressa essa rejeição afirmando que $\neg DD\xi \wedge \neg DD\neg\xi$. Mais do que um operador de determinação, a teoria dispõe de uma hierarquia infinita de operadores com força crescente (o que implica que $\neg D\xi$ não seja derivável de $\neg DD\xi$). Para cada frase mentirosa ξ_σ equivalente a $\neg D^\sigma \xi_\sigma$, em que D^σ é o operador D iterado σ vezes, existe um operador mais forte $D^{\sigma+1}$ com o qual podemos classificar, no nível apropriado, a deficiência semântica da frase. A paracompletude está presente em todos os níveis desta hierarquia, pois nenhuma instância do terceiro excluído $D^\sigma\xi_\sigma \vee \neg D^\sigma\xi_\sigma$ é aceite. Mas, enquanto na teoria de Kripke o abandono do terceiro excluído levava a uma pobreza expressiva, Field afirma que isso já não acontece na teoria que propõe.

5.2 *Teorias paraconsistentes*

Há uma via alternativa para a elaboração de uma teoria da verdade com uma lógica não-clássica, que mantém a lei do terceiro excluído, mas abandona antes a lei da não-contradição. Chamam-se 'paraconsistentes' (ou 'dialeteístas') as teorias que aceitam a existência de algumas contradições verdadeiras, rejeitando ao mesmo tempo o princípio de que tudo se segue de uma contradição (geralmente conhecido com o nome clássico de '*ex contradictione quodlibet*' ou com o nome moderno de 'princípio da explosão'). Graham Priest é um dos mais conhecidos defensores deste género de teoria (em Priest 1998 e 2006a). Nesta exposição, iremos no entanto focar-nos principalmente na teoria proposta por Beall (2005 e 2009), porque Beall, ao contrário de Priest, atribui grande importância à manutenção da transparência da verdade.

A principal razão para se pensar que há algumas contradições que são verdadeiras (ou seja, que há algumas frases verdadeiras cuja negação é também verdadeira, ou seja ainda, que há algumas frases verdadeiras que também são falsas) é, de acordo com muitos dialeteístas, fornecida pelo próprio paradoxo do mentiroso. Afinal, o paradoxo mostra que, se a frase mentirosa é verdadeira, então é falsa; e que, se é falsa, então é verdadeira. Se aceitarmos a bivalência (que é uma consequência do terceiro excluído e da transparência da verdade), temos no paradoxo uma prova de que a frase mentirosa é verdadeira e falsa. Isso foi precisamente o que Tarski viu quando concluiu que a linguagem natural é inconsistente. Mas Tarski considerava que a *inconsistência* implica a *trivialidade* – quer dizer: a aceitação de que tudo é verdadeiro –, e esse é precisamente o passo que os dialeteístas rejeitam.

Uma das críticas frequentemente feitas aos que advogam o abandono do terceiro excluído é que eles não respeitam o carácter *exaustivo* da negação, quer dizer, o princípio de que, para qualquer predicado Π, retiradas as coisas que o satisfazem, *tudo o resto* são coisas que não o satisfazem, ou, o que é equivalente, que satisfazem a sua negação 'não-Π'. Também no caso da verdade se deveria então dizer que todas as coisas são verdadeiras ou não são verdadeiras. Mas, na presença de frases mentirosas, da verdade transparente e da bivalência, uma negação exaustiva não pode comportar-se classicamente; pois, nalguns casos, uma frase verdadeira terá uma negação falsa e, nalguns casos (não necessariamente distintos), uma frase verdadeira terá uma negação verdadeira. A primeira tarefa de uma teoria paraconsistente será, por isso, a de caracterizar o funcionamento da negação.

Beall segue Routley e Routley 1972 para dar uma semântica não-extensional à negação. A semântica (geralmente conhecida como 'semântica de Routley-Meyer') é de tipo modal, em que as frases recebem valores relativamente a vários "mundos" ou "pontos". Alguns mundos são normais (nomeadamente, o mundo actual ou "ponto de base" é sempre normal), mas também pode haver mundos não-normais. Existe uma operação sobre o conjunto de mundos – a chamada *operação estrela* – que dota cada mundo m de um mundo-companheiro m^* e que é tal que $m^{**} = m$. As interpretações da linguagem obedecem, entre outras, às seguintes condições:

S1. Para toda a frase φ e todo o mundo normal m, φ é verdadeira em m ou φ não é verdadeira em m^*.

S2. Para toda a frase φ e todo o mundo m, $\ulcorner\neg\varphi\urcorner$ é verdadeira em m se e somente se φ não é verdadeira em m^*.

A noção de *consequência lógica* é definida dizendo que uma frase φ é consequência de um conjunto de frases Σ se e somente se, para todos os modelos, se as frases de Σ são verdadeiras no mundo actual do modelo, então φ é verdadeira no mundo actual do modelo.

Dado este género de semântica, podemos ver como é que pode haver falsidades verdadeiras, quer dizer, frases φ e mundos normais m tais que φ e $\ulcorner\neg\varphi\urcorner$ são ambas verdadeiras em m. O que isso requer (dada a condição S2 e a equivalência entre $\ulcorner\neg\neg\varphi\urcorner$ e φ resultante da condição $m^{**} = m$) é que, no mundo-companheiro de m, que é m^*, nem φ nem $\ulcorner\neg\varphi\urcorner$ sejam verdadeiras – o que, por sua vez, obriga a que *m** seja um mundo não-normal (dada a condição S1). A exaustividade da negação nos mundos normais é assim obtida à custa da sua não--exaustividade nos mundos não-normais.

Tal como era desejável, na lógica resultante, o princípio da explosão falha, o que se pode facilmente observar considerando um modelo no qual φ e $\ulcorner\neg\varphi\urcorner$ são verdadeiras, mas ψ não é verdadeira. Mas este mesmo modelo também permite verificar algo que, à partida, não parece tão desejável, a saber, que as regras do Silogismo Disjuntivo e do Modus Ponens (para a condicional material) também falham:

$$\varphi \vee \psi, \neg\varphi \not\models \psi$$

$$\varphi, \varphi \rightarrow \psi \not\models \psi$$

Beall reage de maneiras diferentes a estas duas perdas de regras lógicas clássicas. No primeiro caso, Beall alega que o Silogismo Disjuntivo é válido para a linguagem de base, quer dizer, para o fragmento da linguagem que está livre do predicado de verdade e por isso de frases paradoxais; é só quando usamos o predicado de verdade e, nomeadamente, quando o aplicamos a frases que já o contêm, que nos arriscamos a encontrar contra-exemplos ao Silogismo Disjuntivo. No segundo caso, porém, Beall parece julgar que a perda é mais grave, e que o Modus Ponens é uma forma de inferência que devemos ter como válida em toda a linguagem. Isso condu-lo a procurar, tal como fez Field, maneiras de acrescentar à linguagem uma nova condicional, agora com o objectivo de que ela obedeça ao Modus Ponens.

Para conseguir uma nova condicional (primitiva) adequada, Beall socorre-se da semântica de mundos ou pontos já usada para a negação. A ideia mais natural seria considerar que $\varphi \Rightarrow \psi$ é verdadeira num mundo quando se verifica que ψ é verdadeira em todos os mundos nos quais φ é verdadeira (fazendo de $\Rightarrow$ uma espécie de condicional estrita, modalizada). No entanto, sem condições adicionais, esta ideia não funciona, porque a condicional que dela resultaria obedece efectivamente ao Modus Ponens (como se desejava), mas também obedece ao princípio da contracção – e com isso, como já vimos, caímos no paradoxo de Curry. O desafio é, então, o de caracterizar uma condicional não-extensional $\Rightarrow$, que obedeça ao Modus Ponens mas que não contraia. Para isso, Beall caracteriza a condicional por meio de duas condições, uma relativa aos mundos normais e outra relativa a mundos não-normais:

> *Mundos normais*. Para todo o mundo normal m, $\varphi \Rightarrow \psi$ é verdadeira em m se e somente se, para todo o mundo n, se φ é verdadeira em n, então ψ é verdadeira em n.
>
> *Mundos não-normais*. Para todo o mundo não-normal m, $\varphi \Rightarrow \psi$ é verdadeira em m se e somente se, para todos os mundos n e o tais que $\mathbf{R}mno$, se φ é verdadeira em n, então ψ é verdadeira em o. (Em que **R** é uma relação ternária sobre o conjunto dos mundos, que se verifica quando o par $\langle n, o \rangle$ é acessível a partir do mundo m, quer dizer, quando aquilo que acontece em n e aquilo que acontece em o são "compossíveis" em m.)

É em virtude dos mundos não-normais, e da cláusula para os mundos não-normais, que existem contra-exemplos ao princípio da contracção. Pois não há nenhum mundo *normal* no qual φ e $\varphi \Rightarrow \psi$ sejam verdadeiras, mas ψ não seja verdadeira (o que garante a validade do Modus Ponens, uma vez que a relação de consequência é definida sobre os mundos actuais, e estes são normais); no entanto, há mundos *não-normais* onde isso acontece (basta, por exemplo, que seja um mundo que não "tem acesso" a nenhuns mundos) e, num modelo com mundos assim, $\varphi \Rightarrow \psi$ não é verdadeira no mundo actual do modelo, mas $\varphi \Rightarrow (\varphi \Rightarrow \psi)$ pode sê-lo (se $\varphi \Rightarrow \psi$ for verdadeira em todos os mundos em que φ seja verdadeira).

A lógica que resulta desta semântica pode ser axiomatizada e é uma lógica completa, já razoavelmente conhecida, chamada BX, que acrescenta o terceiro excluído (o X) à lógica relevante básica B (veja-se a sua apresentação em Dunn e Restall 2002). É uma lógica não-trivial, como foi provado por Brady (1989).

Uma característica saliente da teoria de Beall (uma teoria paraconsistente da verdade transparente) reside no facto de ela propor soluções diferentes e assimétricas para o paradoxo do mentiroso e para o paradoxo de Curry. Podemos apreciar este ponto, começando por observar que a teoria aceita como verdadeiras todas as instâncias do terceiro excluído, $\varphi \vee \neg\varphi$, mesmo quando φ é uma frase paradoxal. Além disso, uma vez que a disjunção é extensional, em cada instância do terceiro excluído, a teoria aceita pelo menos um dos disjuntos. No caso do mentiroso, quando λ é algo como 'Esta frase não é verdadeira', a teoria aceita $\lambda \vee \neg\lambda$, porque *aceita ambos os disjuntos*, considerando que *ambos são verdadeiros e falsos*. No caso de Curry, quando ξ é algo como 'Se esta frase é verdadeira, então tudo é verdadeiro' (em que 'se... então' é a condicional mais forte $\Rightarrow$), a teoria aceita $\xi \vee \neg\xi$, porque *aceita apenas* $\neg\xi$, considerando que ξ *é uma frase falsa* (quer dizer: *apenas falsa*). Para Beall, nem todas as frases paradoxais são falsidades verdadeiras; algumas, como as frases de Curry, são simplesmente falsas.

Uma frase de Curry é uma frase ξ equivalente a $V([\xi]) \Rightarrow \bot$, em que $\bot$ é uma frase explosiva, como 'tudo é verdadeiro'. Uma vez que a verdade é transparente, podemos simplificar e dizer que ξ é equivalente a $\xi \Rightarrow \bot$, ou seja, que a condicional tem sempre o mesmo valor que a sua antecedente. Como $\Rightarrow$ obedece ao Modus Ponens, vê-se facilmente que ξ não pode ser verdadeira. Pois, se o fosse, seria verdade que $\bot$. Nesta perspectiva, tem então de haver modelos nos quais (quer dizer: em cujo mundo actual) a condicional $\xi \Rightarrow \bot$ é falsa, apesar de a sua antecedente e a sua consequente serem ambas falsas. Mais uma vez, é isso que os mundos não-normais permitem que aconteça: num mundo não-normal, é possível que ξ e $\xi \Rightarrow \bot$ sejam verdadeiras, mas $\bot$ seja falsa; e, havendo um mundo assim, então ξ e $\xi \Rightarrow \bot$ são falsas em todos os mundos normais – logo, também no mundo actual.

Esta solução para o paradoxo de Curry é bem distinta da solução proposta para o mentiroso – onde, como vimos, o dialeteísta simplesmente aceita aquilo que o paradoxo parece querer mostrar, viz. que a contradição é verdadeira. Um dialeteísta como Beall aceita as frases mentirosas, mas rejeita as frases de Curry. E isto apesar de tanto umas como outras serem falsas. A diferença relevante está em que as segundas, mas não as primeiras, são *somente falsas*.

É importante notar que, apesar de esta ser uma diferença fundamental para a maneira como se propõe resolver os paradoxos semânticos, Beall tem dificuldade em expressá-la na sua teoria. Quando pretendemos dizer, numa teoria paraconsistente, que uma frase φ é *somente falsa*, a maneira natural de o fazer seria afirmando 'φ é falsa e não é verdadeira', que se traduz como $F([\varphi]) \wedge \neg V([\varphi])$. Mas, quando a verdade é completamente transparente, não há distinção possível entre *ser falso* e *não ser verdadeiro*: $F([\varphi]) \wedge \neg V([\varphi])$ é equivalente a $V([\neg\varphi]) \wedge \neg V([\varphi])$, que é equivalente a $\neg\varphi \wedge \neg\varphi$, que é equivalente a $\neg\varphi$. O resultado é o seguinte: quando a teoria tenta dizer que φ é somente falsa, o que ela diz é simplesmente que φ é falsa, sendo essa uma característica que não a distingue das outras falsidades, como a do mentiroso, que também são verdadeiras. Em suma, a teoria não consegue caracterizar semanticamente, de uma maneira distintiva, as frases que rejeita. Os dialeteístas têm procurado responder a este problema (veja-se Priest 2006a: 290-295, Priest 2006b: 103-115 e Beall 2009: 48-64).

Ricardo Santos
Universidade de Lisboa
LanCog Group CFUL

Referências

Aristotle. *On Sophistical Refutations*. Translated by E. S. Forster, The Loeb Classical Library, Cambridge, MA: Harvard University Press, 1955.

Barwise, Jon e John Etchemendy. 1987. *The Liar: An Essay on Truth and Circularity.* New York: Oxford University Press.

Beall, JC. 2005. Transparent Disquotationalism. In *Deflationism and Paradox*, ed. by JC Beall and Bradley Armour-Garb. Oxford: Clarendon Press.

Beall, JC. 2009. *Spandrels of Truth*. Oxford: Clarendon Press.

Belnap, Nuel. 1982. Gupta's Rule of Revision Theory of Truth. *Journal of Philosophical Logic* 11: 103-116.

Betti, Arianna. 2004. Lesniewski's Early Liar, Tarski and Natural Language. *Annals of Pure and Applied Logic* 127(1-3): 267-287.

Brady, Ross. 1989. The Non-Triviality of Dialectical Set Theory. In *Paraconsistent Logic: Essays on the Inconsistent*, ed. por G. Priest, R. Routley e J. Norman. Munich: Philosophia Verlag.

Burge, Tyler. 1979. Semantical Paradox. *Journal of Philosophy* 76: 169-198. Reimpresso (com um *post-scriptum*) em Martin 1984: 83-117.

Cavini, Walter. 1993. Chrysippus on Speaking Truly and the Liar. In *Dialektiker und Stoiker. Zur Logik der Stoiker und ihrer Vorläufer*, ed. by K. Döring and T. Ebert. Stuttgart: Franz Steiner.

Crivelli, Paolo. 2004. Aristotle on the Liar. *Topoi* 23(1): 61-70.

Curry, Haskell B. 1942. The Inconsistency of Certain Formal Logics. *Journal of Symbolic Logic* 7(3): 115-117.

Dunn, J. Michael e Greg Restall. 2002. Relevance Logic. In *Handbook of Philosophical Logic*, ed. por Dov M. Gabbay e F. Guenthner, 2ª edição, volume 6. Dordrecht: Kluwer.

Dutilh Novaes, Catarina. 2008. A Comparative Taxonomy of Medieval and Modern Approaches to Liar Sentences. *History and Philosophy of Logic* 29(3): 227-261.

Field, Hartry. 2003. The Semantic Paradoxes and the Paradoxes of Vagueness. In *Liars and Heaps: New Essays on Paradox*, ed. by JC Beall. Oxford: Clarendon Press.

Field, Hartry. 2008. *Saving Truth from Paradox*. Oxford: Oxford University Press.

Glanzberg, Michael. 2001. The Liar in Context. *Philosophical Studies* 103: 217-251.
Glanzberg, Michael. 2004. A Contextual-Hierarchical Approach to Truth and the Liar Paradox. *Journal of Philosophical Logic* 33: 27-88.
Gödel, Kurt. 1931. Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I. *Monatshefte für Mathematik und Physik* 38: 173-198. Text and translation by Jean van Heijenoort as: On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems I, in *Kurt Gödel: Collected Works*, Volume I, ed. by S. Feferman et al., New York and Oxford: Oxford University Press, 1986.
Goldstein, Laurence. 1992. 'This Statement Is Not True' Is Not True. *Analysis* 52: 1-5.
Goldstein, Laurence. 2000. A Unified Solution to Some Paradoxes. *Proceedings of the Aristotelian Society* 100: 53-74.
Grelling, Kurt, and Nelson, Leonard. 1908. Bemerkungen zu den Paradoxien von Russell und Burali-Forti. *Abhandlungen der Fries'schen Schule*, n.s. 2, no. 3: 301-334.
Gupta, Anil. 1982. Truth and Paradox. *Journal of Philosophical Logic* 11: 1-60. Reimpresso em Martin 1984: 175-235.
Gupta, Anil e Nuel Belnap. 1993. *The Revision Theory of Truth*. Cambridge, Mass.: MIT Press.
Halbach, Volker. 2011. *Axiomatic Theories of Truth.* Cambridge: Cambridge University Press.
Herzberger, Hans G. 1982a. Notes on Naive Semantics. *Journal of Philosophical Logic* 11: 61-102. Reimpresso em Martin 1984: 133-174.
Herzberger, Hans G. 1982b. Naive Semantics and the Liar Paradox. *Journal of Philosophy* 79: 479-497.
Horsten, Leon. 2011. *The Tarskian Turn*. Cambridge, Mass.: MIT Press.
Kleene, Stephen C. 1952. *Introduction to Metamathematics*. Amsterdam: North-Holland.
Kremer, Philip. 2006. The Revision Theory of Truth. In *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta.
Kripke, Saul. 1975. Outline of a Theory of Truth. *Journal of Philosophy* 72: 690-716. Reimpresso em Martin 1984: 53-81.
Martin, Christopher J. 2001. Obligations and Liars. In *Medieval Formal Logic*, ed. by M. Yrjönsuuri. Dordrecht: Kluwer.
Martin, Robert L. (ed). 1984. *Recent Essays on Truth and the Liar Paradox*. Oxford: Clarendon Press.
Maudlin, Tim. 2004. *Truth and Paradox: Solving the Riddles*. Oxford: Clarendon Press.
Mignucci, Mario. 1999. The Liar Paradox and the Stoics. In *Topics in Stoic Philosophy*. Ed. by K. Ierodiakonou. Oxford: Clarendon Press.
Papazian, Michael. 2012. Chrysippus Confronts the Liar: The Case for Stoic Cassationism. *History and Philosophy of Logic* 33(3): 197-214.
Parsons, Charles. 1974. The Liar Paradox. *Journal of Philosophical Logic* 3: 381-412. Reimpresso (com um *post-scriptum*) em Martin 1984: 9-45.
Patterson, Douglas. 2012. *Alfred Tarski: Philosophy of Language and Logic*. Palgrave Macmillan.
Paulo. Carta a Tito. In *Bíblia Sagrada*, coord. Herculano Alves, Centro Bíblico dos Capuchinhos, 3ª ed. rev., Lisboa e Fátima: Difusora Bíblica, 2001.
Peckhaus, Volker. 1995. The Genesis of Grelling's Paradox. In *Logik und Mathematik: Frege-Kolloquium, Jena 1993*. Ed. by Ingolf Max and Werner Stelzner. Berlin/New York: Walter de Gruyter.

Priest, Graham. 1998. What is So Bad About Contradictions? *Journal of Philosophy* 95: 410-426.
Priest, Graham. 2006a. *In Contradiction.* 2ª edição (aumentada). Oxford: Clarendon Press.
Priest, Graham. 2006b. *Doubt Truth to be a Liar*. Oxford: Clarendon Press.
Prior, Arthur N. 1955. Curry's Paradox and 3-valued Logic. *Australasian Journal of Philosophy* 33(3): 177-182.
Quine, W. V. 1946. Concatenation as a Basis for Arithmetic. *Journal of Symbolic Logic* 11: 105-114.
Rahman, Shahid, Tulenheimo, Tero and Genot, Emmanuel (eds.). 2008. *Unity, Truth and the Liar: The Modern Relevance of Medieval Solutions to the Liar Paradox*. Springer.
Ramsey, Frank P. 1925. The Foundations of Mathematics. *Proceedings of the London Mathematical Society* 25: 338-384. Reimpresso em Ramsey, Frank P. 1931. *The Foundations of Mathematics and Other Logical Essays*. Ed. por R. B. Braithwaite. London: Routledge & Kegan Paul.
Read, Stephen. 2002. The Liar Paradox from John Buridan back to Thomas Bradwardine. *Vivarium* 40(2): 189-218.
Reinhardt, William N. 1986. Some Remarks on Extending and Interpreting Theories with a Partial Predicate for Truth. *Journal of Philosophical Logic* 15: 219-251.
Richard, Jules. 1905. Les Principes des Mathématiques et le Problème des Ensembles. *Revue Générale des Sciences Pures et Appliquées*. Translated by Jean van Heijenoort as: The Principles of Mathematics and the Problem of Sets. In *From Frege to Gödel.* Ed. by J. van Heijenoort. Cambridge, MA: Harvard University Press, 1967.
Routley, Richard e Valerie Routley. 1972. Semantics of First-Degree Entailment. *Noûs* 3: 335-359.
Russell, Bertrand. 1908. Mathematical Logic as Based on the Theory of Types. *American Journal of Mathematics* 30: 222-262.
Santos, Ricardo. 2003. *A Verdade de um Ponto de Vista Lógico-Semântico*. Lisboa: Fundação Calouste Gulbenkian.
Santos, Ricardo. 2009. Aristóteles e o Paradoxo do Mentiroso. In *Razão e Liberdade: Homenagem a Manuel José do Carmo Ferreira*. Lisboa: CFUL.
Séneca, Lúcio Aneu. *Cartas a Lucílio.* Trad. de J. A. Segurado e Campos. Lisboa: Fundação Calouste Gulbenkian, 1991.
Simmons, Keith. 1993. *Universality and the Liar: An Essay on Truth and the Diagonal Argument*. Cambridge: Cambridge University Press.
Soames, Scott. 1999. *Understanding Truth*. New York: Oxford University Press.
Spade, Paul Vincent. 1975. *The Mediaeval Liar: A Catalogue of the Insolubilia-Literature*. Pontifical Institute of Mediaeval Studies.
Spade, Paul Vincent. 1982. Insolubilia. In *The Cambridge History of Later Medieval Philosophy.* Ed. by N. Kretzmann, A. Kenny and J. Pinborg. Cambridge: Cambridge University Press.
Strawson, P. F. 1950. On Referring. *Mind* 59: 320-344.
Tarski, Alfred. 1933. Pojecie prawdy w jezykach nauk dedukcyjnych. *Prace Towarzystwa Naukowego Warszawskiego, Wydział III Nauk Matematyczno-Fizycznych*, nr 34, Warszawa. Translated by J. H. Woodger as: The Concept of Truth in Formalized Languages. In Tarski 1983: 152-278. Traduzido por Cezar A. Mortari como: O Conceito de Verdade nas Linguagens Formalizadas. In Tarski 2007: 19-148.
Tarski, Alfred. 1944. The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research* 4 (3): 341-376.

Tarski, Alfred. 1969. Truth and Proof. *Scientific American*. Reprinted in *A Philosophical Companion to First-Order Logic*. Ed. by R.I.G. Hughes. Indianapolis: Hackett, 1993.

Tarski, Alfred. 1983. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. 2nd ed., ed. by John Corcoran. Indianapolis: Hackett.

Tarski, Alfred. 2007. *A Concepção Semântica da Verdade.* Tradução de Celso R. Braida, Cezar A. Mortari, Jesus P. Assis e Luiz Henrique A. Dutra. São Paulo: Editora UNESP.

van Fraassen, Bas C. 1966. Singular Terms, Truth-Value Gaps, and Free Logic. *Journal of Philosophy* 63: 481-495.

Wolenski, Jan. 1994. Jan Łukasiewicz on the Liar Paradox, Logical Consequence, Truth, and Induction. *Modern Logic*, vol. 4, no. 4: 392-400.