





**Revista de Pesquisa em Filosofia**

# **FUNDAMENTO**

Ouro Preto - Minas Gerais - Brasil



**UFOP**

**Universidade Federal  
de Ouro Preto**

Fundamento Revista de Pesquisa em Filosofia / Universidade Federal de Ouro Preto / Grupo de Pesquisa em Filosofia Contemporânea. n. 12, (Jan–jun. 2015) – Ouro Preto:

Ed. UFOP, 2015

Semestral – Tiragem: 200 exemplares.

ISSN: 2177–6563

1. Filosofia - Periódicos. I. Universidade Federal de Ouro Preto. II. Instituto de Filosofia, Artes e Cultura. III. Grupo de Pesquisa em Filosofia Contemporânea.

Endereço para correspondência:

**Revista Fundamento**  
**Instituto de Filosofia, Artes e Cultura – IFAC**  
**Rua Coronel Alves, 55, Centro, Ouro Preto, MG, Brasil**  
**CEP: 35400-000**

**Endereço eletrônico: [revistafundamento@ufop.br](mailto:revistafundamento@ufop.br)**

## **Revista Fundamento**

UNIVERSIDADE FEDERAL DE OURO PRETO

**Reitor**

Prof. Dr. Marcone Jamilson Freitas Souza

INSTITUTO DE FILOSOFIA, ARTES E CULTURA/IFAC

**Diretor**

Prof. Dr. Guilherme Paoliello

DEPARTAMENTO DE FILOSOFIA/DEFIL

**Chefe de Departamento**

Prof. Dr. Olímpio José Pimenta

**Bibliotecária**

Neide Nativa

FUNDAMENTO

REVISTA DE PESQUISA EM FILOSOFIA

**Diretores Gerais**

Prof. Dr. Mário Nogueira (UFOP)

Prof. Dr. Sérgio Miranda (UFOP)

**Diretores Executivos**

Aline Monteiro Homssi (UFOP)

**Editores**

Prof. Dr. Mário Nogueira (UFOP)

Prof. Dr. Sérgio Miranda (UFOP)

Prof. Me. Luiz Helvécio Marques Segundo (UFSC / UFOP)

**Diagramação**

Aline Monteiro Homssi (UFOP)

**Assistente de Revisão**

Aline Monteiro Homssi (UFOP)

**Comitê Editorial**

Prof. Dr. Alexandre Noronha Machado (UFPR)

Prof. Dr. Hélio Lopes da Silva (UFOP)

Prof. Dr. Danilo Marcondes (PUC-Rio)

Prof. Dr. Frank Thomas Sautter (UFSM)

Prof. Dr. Guido Imaguire (UFRJ)

Prof. Dr. Marco C. Ruffino (Unicamp)

Prof. Dr. Nelson Gonçalves Gomes (UNB)

Prof. Dr. Túlio Xavier de Aguiar (UFMG)

## SUMÁRIO

### I - ARTIGOS INÉDITOS

- 11 O QUE É UMA LÓGICA?  
*WHAT IS A LOGIC?*  
Newton G. A da Costa / Décio Krause - Universidade Federal de Santa Catarina
- 21 *IDENTIDADE, INDISCERNIBILIDADE E LÓGICA*  
*IDENTITY, INDISCERNIBILITY AND LOGIC*  
Kherian Gracher - Universidade Federal de Santa Catarina
- 41 UMA EXPLICAÇÃO PRÉ-SEMÂNTICA DAS DESCRIÇÕES  
REFERENCIAIS  
*A PRE-SEMANTIC ACCOUNT OF REFERENTIAL DESCRIPTIONS*  
Fernando Furtado - Universidade de Lisboa
- 65 VAGUENESS: TOLERANCE AND INCOHERENCE  
*VAGUEZA: TOLERÂNCIA E INCOERÊNCIA*  
Sagid Salles - Universidade Federal do Rio de Janeiro

### II - TRADUÇÃO

- 87 O DESVIO METAFÓRICO  
Monroe C. Beardsley  
Tradução: Carlos André Fernandes  
Revisão: L. H. Marques Segundo



# **Artigos Inéditos**



# O que é uma lógica?

## *What is a logic?*

*Newton C. A. da Costa  
Décio Krause  
Universidade Federal de Santa Catarina<sup>1</sup>*

### **Resumo**

Neste artigo pretendemos oferecer uma noção minimamente rigorosa do que seja uma lógica. A ideia central é que uma lógica, estritamente falando é um sistema lógico. Entenderemos um sistema lógico como formado, grosso modo, por um conjunto de fórmulas e um conjunto de regras de inferência. A partir daí veremos duas noções fundamentais: dedutibilidade e consequência lógica.

### **Palavras-chave**

Lógica. Sistema lógico. Dedutibilidade. Consequência lógica.

---

*1 - Grupo de Estudos em Lógica e Fundamentos da Ciência. Departamento de Filosofia Universidade Federal de Santa Catarina (Julho 2013).*

## Abstract

*Our aim in this paper is to provide a minimal rigorous understanding of a logic. The main idea is that a logic is, strickly speaking, a system of logic. Roughly, a system of logic will be taken as a set of formulas and a set of rules of inference. Then, we will look at two fundamental notions: deductibility and logical consequence.*

## Keywords

*Logic. System of logic. Deductibility. Logical consequence.*

## Lógica (disciplina) e “lógica” (sistema lógico)

A palavra “lógica” tem diversos significados. Comumente, ouve-se falar da ‘lógica do mercado’, ou da ‘lógica da criança’ (em distinção à ‘lógica do adulto’), etc. Todas essas denominações são imprecisas e necessitam ser vistas com cautela. Mas *Lógica* é também uma disciplina que é parte tanto da filosofia quanto da matemática, e tem se infiltrado em praticamente todas as áreas da investigação humana. Neste artigo, não falaremos da história da lógica, nem de suas aplicações, que hoje em dia vão da ciência da computação à medicina, passando pela matemática, pelas ciências empíricas (física, biologia), pela filosofia, pela economia, filosofia do direito e da psicanálise. Ficaremos restritos a desvendar o que se entende por *lógica* não como disciplina, mas como sinônimo de *sistema lógico*. Porém, algumas palavras devem ser ditas sobre esta distinção. Vários textos que se encontram na praça (livros introdutórios, principalmente) classificam a disciplina Lógica como o estudo de inferências válidas, e que seus “princípios fundamentais” seriam os princípios da identidade, da contradição e do terceiro excluído. (Quadro 1) Nada mais falso e desprovido de sentido do que isto. Basta que o leitor consulte o item 03 da *Mathematics Subject Classification*, editada pela *American Mathematical Society* em parceria com sua irmã alemã *Zentralbat für Mathematik*, (Quadro 2) (para constatar que a Lógica vai muito além disso, envolvendo a teoria da recursão, a teoria dos modelos, os fundamentos da teoria de conjuntos, dentre outras variadas áreas, que nem de perto se assemelham a estudo de raciocínios (aliás, o que seria um “raciocínio”, precisamente?). Aquela parte da lógica que

poderíamos chamar de teoria da argumentação, com certa condescendência, talvez pudesse se aproximar do que apregoam esses autores. Ademais, os três mencionados princípios não são os únicos que vigoram na lógica *clássica*, havendo muitos outros (dupla negação, lei de Peirce, lei de Scotus, etc); por fim, basta notar que não se pode fundar a parte mais básica da lógica (dita) clássica nesses três princípios. Eles não bastam para axiomatizar o cálculo proposicional clássico, por exemplo. E depois, porque identificar “lógica” com lógica clássica? Na lógica intuicionista de Brouwer e Heyting, o terceiro excluído não vale; nas lógicas paraconsistentes, o princípio da contradição não vale em geral, e nas lógicas não-reflexivas o princípio da identidade é que é restringido. (Quadro 3) Assim, como sustentar a afirmação acima? Afirmá-la constitui pura e simplesmente um erro conceitual. Mais à frente, alguns dos termos deste parágrafo ficarão mais claros.

## Lógica não é apenas linguagem

Outro vício que encontramos nos livros e textos é a afirmativa de que lógica (no sentido de sistema lógico) identifica-se com linguagem. Talvez isso se deva a uma tradição, que remonta a Frege e Russell, de apresentar uma lógica (no caso deles, a que chamamos de ‘clássica’) iniciando pela descrição de uma linguagem (símbolos primitivos, regras gramaticais de formação das expressões bem formadas, e depois o seu aparato dedutivo, composto por axiomas e regras de inferência). Mas este é apenas um dos lados da história. Com efeito, como mostraram os poloneses no início do século XX, uma lógica pode ser identificada com uma álgebra, e uma linguagem formal também é um tipo de álgebra ---trata-se de uma álgebra livre com um determinado conjunto de geradores, esses conceitos devendo ser entendidos no sentido de que se lhes dá em álgebra.

## Lógica como uma álgebra

Do ponto de vista algébrico, podemos ver uma lógica como um par ordenado  $L = \langle F, T \rangle$ , onde  $F$  é um conjunto não vazio (a natureza dos elementos não importa em uma primeira instância) e  $T$  é uma coleção de subconjuntos de  $F$ . Os elementos de  $F$  são denominados de *fórmulas*, e os de  $T$ , de *teorias*. Os seguintes axiomas devem ser satisfeitos: (1)  $F$  pertence a  $T$ , ou seja, o conjunto das fórmulas é uma teoria; (2) se  $T_1, T_2, \dots$  são teorias (o conjunto pode ser qualquer, não necessariamente enumerável, ou seja, não precisa admitir uma

correspondência um a um com o conjunto dos números naturais), então a sua interseção, que consta daquelas fórmulas que pertencem a todas essas teorias, também é uma teoria. Isso pode parecer difícil, mas não é; tudo o que se pede é concentração e capacidade de raciocínio abstrato. Vamos em frente.

Dado um conjunto de fórmulas  $\Gamma$  e uma fórmula  $A$ , escrevemos  $\Gamma \vdash A$  para indicar que toda teoria que contenha as fórmulas de  $\Gamma$  contém  $A$ . Neste caso, dizemos que  $A$  é *dedutível*, ou que é *consequência sintática*, das fórmulas de  $\Gamma$ . As fórmulas de  $\Gamma$  denominam-se de *premissas* da dedução e  $A$  é a *conclusão*. Escrevemos  $\Gamma \not\vdash A$  em caso contrário. É simples mostrar que o operador  $\vdash$ , que chamamos de *operador de dedução*, tem as seguintes propriedades: (a)  $\{A\} \vdash A$  (autodedutibilidade), (b) se  $\Gamma \vdash A$ , então  $\Gamma \cup \Delta \vdash A$ , sendo  $\Delta$  um conjunto qualquer de fórmulas (monotonicidade), e (c) se  $\Gamma \vdash A$  e se  $\Delta \vdash B$  para toda  $B$  em  $\Gamma$ , então  $\Delta \vdash A$  (lei do corte). A autodedutibilidade é óbvia: toda fórmula é dedutível dela mesma (ou de seu conjunto unitário). A monotonicidade significa que se deduzimos  $A$  de  $\Gamma$ , continuaremos deduzindo  $A$  não importa quantas premissas adicionemos às que já tínhamos. A lei (ou regra) do corte também é bastante intuitiva. Repare que na maioria dos ‘raciocínios’ que fazemos, ferimos a monotonicidade. Por exemplo, se sempre que uma criança houve seu pai buzinar, sabe que precisa ir correndo abrir o portão da garagem para ele entrar, não implica que hoje, ouvindo a buzina, deva fazer isso, pois sabe uma coisa a mais, a saber, que seu pai não esqueceu o controle do portão. Uma premissa a mais *derrota* (*defeat*) a derivação. Tais formas de inferência são extremamente importantes, mas não trataremos delas aqui.

A partir dessas definições, podemos introduzir todo o aparato sintático usual, bem como provar seus resultados correspondentes, mas também não é disso que queremos falar. Um outro conceito importante é o de *consequência lógica*; sendo  $\Gamma$  um conjunto de fórmulas, escrevemos  $Cn(\Gamma)$  para representar o conjunto de todas as fórmulas de  $F$  que são dedutíveis das fórmulas de  $\Gamma$ . Em outras palavras,  $A$  estará em toda teoria que contenha as fórmulas de  $\Gamma$ , e somente nessas teorias. Se dispusermos do símbolo de dedução, podemos definir  $Cn(\Gamma) = \{A : \Gamma \vdash A\}$ . O operador  $Cn$  é chamado de *operador de consequência de Tarski*. É imediato provar que ele satisfaz as seguintes condições: (i)  $\Gamma$  está contido em  $Cn(\Gamma)$ ; (ii) se  $\Gamma$  está contido em  $\Delta$ , então  $Cn(\Gamma)$  está contido em  $Cn(\Delta)$ , e (iii)  $Cn(Cn(\Gamma))$  está contido em  $Cn(\Gamma)$ , ou seja, aquilo que se deduz do que já foi deduzido, pode ser obtido apenas das premissas originais.

Vamos ver algumas formas alternativas de dizer a mesma coisa, talvez mais intuitivas à maioria dos leitores.

A primeira inverte a situação acima. Dizemos que uma lógica é um par  $L = \langle F, \vdash \rangle$

composto por um conjunto  $F$  como acima e por uma relação entre conjuntos de fórmulas e fórmulas, satisfazendo a autodedutibilidade, a monotonicidade e a lei do corte. Com isso, podemos agora definir *teoria*; uma teoria é um conjunto de fórmulas fechado por deduções, ou seja, é um conjunto  $T$  tal que tudo o que for deduzido de fórmulas de  $T$  ainda pertence a  $T$ . Podemos agora mostrar que as condições (1) e (2) impostas acima podem ser obtidas como teoremas. As duas abordagens são equivalentes. Mas há uma terceira que vale a pena mencionar. Pomos agora que uma lógica é um par  $L = \langle F, Cn \rangle$ , sendo  $F$  como acima e  $Cn$  uma relação entre conjuntos de fórmulas e fórmulas, satisfazendo (i)-(iii) acima. Podemos agora definir  $\vdash$  a partir de  $Cn$  do seguinte modo:  $\Gamma \vdash A$  se e somente se  $A$  pertence a  $Cn(\Gamma)$  e provar que valem (a)-(c) acima. A partir disso, voltamos a definir teoria e provar que as condições (1) e (2) são satisfeitas. Isso mostra que as três formas de abordagem são equivalentes. Uma quarta, também equivalente às anteriores é a seguinte. Dizemos que uma lógica é uma tripla ordenada  $L = \langle F, A, R \rangle$ , onde  $F$  é como antes,  $A$  é um subconjunto de  $F$ , dito conjunto dos *axiomas* de  $L$  e  $R$  é uma coleção de relações entre conjuntos de fórmulas (as premissas da regra) e fórmulas, ditas *regras de inferência* de  $L$ . Suporemos que as regras são *finitárias*, ou seja, há sempre um número finito de premissas, mas isso poderia ser generalizado. Se  $A_1, A_2, \dots, A_n$  forem as premissas de uma regra  $R$  e  $A$  for a conclusão, diremos que  $A$  é *consequência imediata* das  $A_j$  pela regra  $R$ .

Agora podemos introduzir a noção de dedução,  $\Gamma \vdash A$ , do seguinte modo. Dizemos que  $A$  é dedutível de  $\Gamma$  se existe uma sequência (no nosso caso, finita, mas isso pode ser estendido para casos mais gerais) de fórmulas  $B_1, B_2, \dots, B_n$  tal que (I)  $B_n$  é  $A$ ; (II) cada  $B_j$ , para  $j < n$ , é um axioma (elemento de  $A$ ), ou pertence a  $\Gamma$ , ou é consequência imediata de fórmulas precedentes da sequência por uma das regras de inferência. Claro que (a)-(c) acima valem e que podemos definir facilmente  $Cn(\Gamma)$  para qualquer  $\Gamma$ , de forma a valerem (i)-(iii). Teorias também podem ser definidas sem dificuldade (o leitor deveria tentar), de modo a cumprir (1) e (2).

Vamos ver um exemplo simples de uma lógica. Tomemos  $F = \{0, 1, 2, 3, 4, \dots\}$  (o conjunto dos números naturais) e seja  $A = \{0, 2, 4, 6, 8, \dots\}$  (o conjunto dos números pares). As regras são duas, definidas fazendo-se uso das operações conhecidas com números: ( $R_1$ ) de  $A$  e  $B$ , inferimos  $A+B$  e ( $R_2$ ) de  $A$  e  $B$ , inferimos  $A.B$ . É imediato constatar que  $\{1, 2\} \vdash 1$ ,  $\{1, 3\} \vdash 4$ ,  $\{2, 3\} \vdash 6$ , bem como que  $Cn(A) = A \neq F$ , etc.

Esta lógica, apesar de interessante para exercícios, não tem muita utilidade. Uma mais poderosa é obtida acrescentando-se à estrutura  $L = \langle F, T \rangle$  um operador binário definido sobre  $F$ , denotado  $\rightarrow$ . Escreveremos  $A \rightarrow B$  para indicar que  $A$  e  $B$  estão relacionados por  $\rightarrow$ . Pomos agora a seguinte definição:  $L$  é uma *lógica implicativa intuicionista* se forem satisfeitas

as seguintes condições, para toda teoria  $T$  de  $L$ , e todas  $A, B, C$  em  $T$ : ( $I_1$ )  $A \rightarrow (B \rightarrow A)$  pertence a  $T$ ; ( $I_2$ )  $(A \rightarrow B) \rightarrow ((A \rightarrow (B \rightarrow C)) \rightarrow (A \rightarrow C))$  pertence a  $T$ , e ( $I_3$ ) se  $A$  e se  $A \rightarrow B$  pertencem a  $T$ , então  $B$  pertence a  $T$ . (Quadro 4) Obtém-se uma *lógica implicativa clássica* se a essas condições impusermos uma a mais, conhecida como Lei de Peirce: ( $P$ )  $((A \rightarrow B) \rightarrow A) \rightarrow A$  pertence a  $T$ . O caminho está traçado. Acrescentando-se à estrutura  $L = \langle F, T \rangle$  outros operadores, como  $\wedge$  (conjunção),  $\vee$  (disjunção) e  $\neg$  (negação) obedecendo axiomas pertinentes, obtemos as lógicas *conjuntiva, disjuntiva, da negação* (minimal, por exemplo) ou então, colocando-se esses conectivos todos, obtendo  $L = \langle F, T, \rightarrow, \wedge, \vee, \neg \rangle$  com axiomas adequados, obtemos as lógicas usuais *positiva intuicionista, positiva clássica* (essas duas sem o operador unário de negação), *de Johansson-Kolmorov, de Brouwer-Heyting*, e a *clássica*. Claro que podemos estender essas estruturas para obter as correspondentes lógicas quantificacionais (isso daria um pouco mais de trabalho, e não dispomos deste espaço aqui). Operadores convenientes como  $\diamond$  (possibilidade),  $\partial$  (obrigatoriedade),  $\circ$  (bom comportamento) nos permite obter com facilidade os sistemas modais, deónticos e muitos cálculos paraconsistentes, seja de nível proposicional ou quantificacional. Adicionando-se um operador binário  $\in$  (pertinência), podemos obter as teorias de conjuntos usuais, como o sistema Zermelo-Fraenkel. As propriedades sintáticas e semânticas de todos esses sistemas podem ser estudadas como se faz usualmente. Saliente-se que mesmo *lógicas indutivas* podem ser estudadas via esta abordagem.

Como se vê, não se pode antecipar uma definição de lógica sem que se conheça a fundo os sistemas que são trabalhados pelos lógicos. Assim, ao se dizer que a lógica trata de raciocínios, do que se está falando? Um brouweriano não aceitaria uma redução ao absurdo, como faz um 'clássico'. Um lógico paraconsistente não vê problema em derrogar o "mais certo de todos os princípios", como se referiu Aristóteles ao princípio da contradição, e assim por diante. Da mesma forma, a lógica (qual lógica?) não pode ser vista como algo *a priori*. Temos mais familiaridade com a lógica (dita) clássica por questões culturais, mas nada impede que, em princípio, nossas categorias do entendimento nos tivessem levado a uma lógica distinta da clássica.

Ademais, o que é a lógica *clássica*? Para alguns, como Quine, ela deve se restringir à lógica quantificacional elementar (de primeira ordem), mas porque isso tem que ser assim? Vimos acima que, procedendo adequadamente, podemos obter sistemas tão fortes como as teorias de conjuntos. Qual o critério para excluí-las daquilo que denominamos 'lógica', relegando-as a serem 'matemática', como apregoava Quine?

Qualquer resposta não passará de mero palpite, opinião. Lógica é aquilo que faz parte da atividade do lógico, e como se constata verificando-se a *Mathematical Subject Classification*,

este campo do conhecimento, como qualquer outro, está em constante alteração (itens entram e saem dessa classificação quando ela é reavaliada, a cada 10 anos), o que atesta ser uma área viva e saudável de nossa atividade científica.

## Quadro 1

Os três mais célebres princípios da lógica tradicional (mas não os únicos) admitem cada um uma variedade de formulações não equivalentes. Por exemplo, o princípio da identidade pode ser  $A \rightarrow A$  (lógica proposicional), ou  $\forall x(x=x)$  (lógica de primeira ordem), “todo objeto é idêntico a si mesmo” (formulação semântica). Versões do princípio da contradição são:  $\neg(A \wedge \neg A)$ , “dentre duas proposições contraditórias (uma sendo a negação da outra), uma delas é falsa”. Quanto ao terceiro excluído, temos  $A \vee \neg A$ , e “dentre duas proposições contraditórias (uma sendo a negação da outra), uma delas é verdadeira”. Outros princípios são: a dupla negação (clássica) diz que  $\neg\neg A \leftrightarrow A$ , a redução ao absurdo clássico é  $(\neg A \rightarrow \neg B) \rightarrow ((\neg A \rightarrow B) \rightarrow A)$ , e assim por diante.

## Quadro 2

No site da Mathematics Subject Classification <http://www.ams.org/mathscinet/msc/msc2010.html>, o leitor encontrará as subdivisões datadas de 2010 da matemática presente. A seção 03 intitula-se Mathematical Logic and Foundations, e engloba, dentre várias outras, a teoria dos modelos, a teoria da prova, a teoria dos conjuntos, a teoria da computabilidade, a lógica algébrica. Como dizer que a lógica ocupa-se meramente do estudo de raciocínios válidos?

## Quadro 3

Lógicas que derogam o princípio da identidade (em alguma de suas formas) ou que restringem o conceito de identidade da lógica clássica são denominadas de não-reflexivas. Por exemplo, as *lógica da implicação causal*, que interpretam  $A \rightarrow B$  como “A causa B”, em geral não aceitam que uma coisa possa ser a causa dela mesma. O princípio  $\forall x(x=x)$  é

enfraquecido nas *Lógicas de Schrödinger* e nas teorias de *quase-conjuntos*, que operam sobre domínios nos quais se admite que a relação de identidade não se aplica para determinados objetos. As lógicas que derrogam o princípio do terceiro excluído são chamadas de *paracompletas*, estando entre elas a lógica intuicionista e as lógicas polivalentes.

## Quadro 4

Um exemplo de uma derivação na lógica positiva intuicionista. Provaremos que  $A \rightarrow A$  é uma tese desta lógica, em estrita conformidade com a definição de dedução dada no texto (neste caso, não há premissas além dos axiomas da lógica em questão). Temos a seguinte derivação, na qual indicamos à direita o que se passa com cada fórmula da sequência:

- |  |                        |
|--|------------------------|
| 1. $A \rightarrow (A \rightarrow A)$   | instancia do axioma I1 |
| 2. $(A \rightarrow (A \rightarrow A)) \rightarrow (A \rightarrow ((A \rightarrow A) \rightarrow A) \rightarrow (A \rightarrow A))$ | axioma I2              |
| 3. $(A \rightarrow ((A \rightarrow A) \rightarrow A)) \rightarrow (A \rightarrow A)$   | 1,2, axioma I3         |
| 4. $A \rightarrow ((A \rightarrow A) \rightarrow A)$   | axioma I1              |
| 5. $A \rightarrow A$   | 3,4, axioma I3         |

## Referências

Da Costa, N. C. A. [1980], *Ensaio sobre os Fundamentos da Lógica*. São Paulo, Hucitec-EdUSP.

### Newton C. A. da Costa

Aposentado dos departamentos de matemática e de filosofia da USP, é um dos criadores das lógicas paraconsistentes e tem dado colaboração em diversas áreas do conhecimento, tendo sido agraciado e reconhecido no Brasil e no exterior. Atualmente é professor do programa de pós-graduação do departamento de filosofia da UFSC. É pesquisador do CNPq.

**Décio Krause**

Aposentado do departamento de matemática da UFPR, é atualmente professor do departamento de filosofia da UFSC. É pesquisador do CNPq.



# Identidade, Indiscernibilidade e Lógica

## *Identity, Indiscernibility, and Logic*

*Kherian Gracher<sup>1</sup>*

*Universidade Federal de Santa Catarina*

### **Resumo**

É a identidade fundamental para todo sistema formal? Ainda que um certo sistema não contenha a identidade, será que esse conceito não é pressuposto de algum modo – seja em um nível metalinguístico ou mesmo intuitivo? Nesse artigo iremos discutir esse problema. Otávio Bueno (2014, 2016) argumenta contra a eliminação da identidade, afirmando que esse conceito é fundamental e não-eliminável (mesmo nos sistemas que alegam fazê-lo). Décio Krause e Jonas Arenhart (2015), por outro lado, apresentam uma série de objeções à tese de Bueno. Em um primeiro momento analisaremos como o conceito de identidade foi compreendido na tradição filosófica, como também parte de seu tratamento formal. Examinaremos também a noção de indiscernibilidade, que pela abordagem tradicional da identidade, é equivalente à noção de identidade, mas que (argumentavelmente) não lhe é logicamente equivalente. Apresentaremos assim a estratégia proposta por Krause e Arenhart, na qual um sistema formal pode eliminar a identidade em prol da noção de indiscernibilidade. E, por fim, iremos expor as críticas de Bueno a tal estratégia, oferecendo algumas objeções contra ele.

---

*1* - Programa de Pós-Graduação em Filosofia. Grupo de Pesquisa em Lógica e Fundamentos da Ciência (CNPq)  
O autor agradece a Décio Krause, Jonas Arenhart, Otávio Bueno e Luiz Helvécio pelas discussões frutíferas e pela paciência.

## **Palavras-chave**

Identidade; Indiscernibilidade; Filosofia da Lógica; Metafísica.

## **Abstract**

Is identity fundamental to formal systems? Even if a system have no the identity relation, is that concept is not assumed in any way – whether in a metalinguistic or intuitive level? In this paper we shall discuss this issue. Otávio Bueno (2014, 2016) argues against the elimination of identity, holding that this concept is fundamental and non-eliminable (even in does systems that claim to do so). Décio Arenhart Krause and Jonas (2015), by the other hand, have a number of objections to Bueno's thesis. Firstly, we look at how the concept of identity was accounted in the philosophical tradition, as well as part of its formal account. We will also examine the notion of indiscernibility that, according to the traditional approach to identity, is equivalent to the notion of identity, but (arguably) is not it logically equivalent. Thus, we shall state Krause and Arenhart's strategy, according to, in a formal system we can eliminate identity in favor of the indiscernibility notion. Finally, we shall expose Bueno's criticism to such strategy, and then to put forward some objections against it.

## **Keywords**

Identity; Indiscernibility; Philosophy of Logic; Metaphysics. Metafísica.

Otávio Bueno, no artigo *Why Identity is Fundamental* (2014), formulou quatro razões que tornariam a identidade fundamental: (1) a identidade é pressuposta em todo sistema conceitual; (2) é requerida para uma caracterização mínima de indivíduo; (3) não pode ser definida mesmo nos sistemas que, alegadamente, têm poder expressivo para tal; (4) é necessária para a inteligibilidade dos quantificadores. Por outro lado, Décio Krause e Jonas Arenhart (2015) ofereceram diversos desafios para a tese de Bueno, apresentando

objeções contra cada uma das razões anteriores. Segundo eles, isso nos permite recusar a aparente fundamentalidade da identidade – argumentando a favor da possível eliminação desta noção em prol de uma noção que seria mais *fraca* (*viz.*, indiscernibilidade). O debate seguiu com a posterior publicação de Bueno (2016), respondendo às críticas de Krause e Arenhart e oferecendo objeções contra eles.

A discussão se tornou ampla, com argumentos cada vez mais sofisticados e desafios intelectuais que põem em causa crenças e princípios sedimentados na tradição filosófica – quiçá em nossas intuições mais básicas. Uma abordagem completa dessa discussão, com todas as suas implicações, não pode ser feita de maneira apressada e em um espaço tão curto quanto este que temos. Portanto, farei duas restrições. Primeiro, iremos discutir a primeira razão defendida por Bueno, *i.e.*, a identidade é fundamental porque é pressuposta em todo sistema conceitual. A segunda restrição nos deixará com o seguinte problema, que levaremos para o âmbito de sistemas formais: é a identidade pressuposta em todo sistema formal?

Em um primeiro momento veremos como a identidade é compreendida pela tradição filosófica – e, *en passant*, como esse conceito é entendido formalmente. Tão logo falamos sobre a identidade, outro conceito estritamente relacionado aparece: a indiscernibilidade. Devemos, pois, diferenciar essas duas noções. Após esse preâmbulo, onde as noções gerais serão introduzidas, apresentaremos uma estratégia proposta por Krause e Arenhart (2015, p.3), que defendem a possibilidade de um sistema formal assumir a indiscernibilidade (e não a identidade) como conceito primitivo – o que tornaria a identidade, de acordo com eles, pelo menos em princípio, eliminável. Por fim, apresentaremos as críticas oferecidas por Bueno (2016) e algumas objeções a elas.

## 1 A Teoria Tradicional da Identidade

O que é a identidade?<sup>2</sup> Ainda que este artigo não se centre especificamente nesse problema, enfrentá-lo é inevitável. Chamarei de “Teoria Tradicional da Identidade” (TTI) a concepção de identidade que foi sedimentada na tradição filosófica.<sup>3</sup> Como veremos, parece razoável afirmar que a TTI influenciou o modo pelo qual a identidade foi considerada nos sistemas formais usuais.

2 - Há diversos usos e noções para o conceito de identidade. Discutiremos aqui a noção *numérica* da identidade, ou identidade *out court*. De modo geral, esta noção é empregada em expressões como “dois mais dois é *igual* a quatro” ou “Batman é *idêntico* a Bruce Wayne”. A noção numérica da identidade se distingue da noção *qualitativa*. Caso o leitor não esteja familiarizado com essa distinção, *ver* (LOCKE, 1996, XXVII), (SIDER; CONEE, 2005, p. 7-8), (NOONAN; CURTIS, 2014), (KORFMACHER, 2006) e (MCGINN, 2000, p. 2-3).

3 - Vamos seguir a terminologia empregada por (BRAIDA; KRAUSE, 2013, p. 177-1784).

## 1.1 Tratamento Filosófico da Identidade

A tradição filosófica associa ao conceito de identidade certas características que esse conceito preservaria, cujo tratamento é geralmente oferecido pela concepção metafísica que se assume. Algumas das principais características, que encontramos na literatura filosófica, compreendem a identidade como:

(1) Sendo uma relação universal. Ou seja, todas as entidades mantêm a relação de identidade consigo mesmo (por vezes tal característica é chamada de “auto-identidade”).

(2) Sua natureza é descrita em termos da indiscernibilidade das entidades. Em linhas gerais, a indiscernibilidade é a relação mantida quando entidades têm *as mesmas* características distinguidoras.<sup>4</sup>

(3) Preserva a substituição *salva veritate*. Tal característica garante que a substituição de termos co-referenciais (que têm *a mesma* referência) em uma proposição manterá o valor de verdade desta. Por exemplo, se a proposição *x é azul* é verdadeira (ou falsa), sendo *x igual a y*, a proposição *y é azul* será também verdadeira (ou falsa).

De modo geral, a tradição filosófica assumiu a chamada “Lei de Leibniz” como o modo de se caracterizar a identidade. Esse princípio pode ser formulado como: (LL) entidades são idênticas se, e somente se, são indiscerníveis. Nesta acepção a identidade é definida em termos da indiscernibilidade.<sup>5</sup> No entanto, a Lei de Leibniz pode ser analisada como a conjunção de duas diferentes teses:

### Identidade dos Indiscerníveis:

Se as entidades são indiscerníveis, então elas são idênticas.

---

<sup>4</sup> - Tais características dependem da metafísica subjacente que se assume. Se temos uma metafísica de entidades e propriedades, então a indiscernibilidade relaciona propriedades; se temos uma metafísica de substâncias, então consiste em terem ou não a mesma substância; etc..

<sup>5</sup> - Devemos notar que o próprio Leibniz não ofereceu esta formulação. Para mais, ver (RODRIGUEZ-PEREYRA, *No prelo*).

## Indiscernibilidade dos Idênticos:

Se as entidades são idênticas, então elas são indiscerníveis.

Dito em outros termos, a *Indiscernibilidade dos Idênticos* é a tese que: se  $x$  e  $y$  são numericamente idênticos (ou seja, são *o mesmo*; ou denotam o mesmo objeto), então toda característica distinguidora ou proposição atribuída a  $x$  será também característica distinguidora ou proposição atribuída a  $y$  (e vice-versa). Já a *Identidade dos Indiscerníveis* é a tese que: se  $x$  e  $y$  são indiscerníveis (ou seja, toda característica distinguidora ou proposição atribuída a  $x$  é também característica distinguidora ou proposição atribuída a  $y$ ), então  $x$  e  $y$  são, também, numericamente idênticos (ou seja, são *o mesmo* objeto).

## 1.2 Tratamento Formal da Identidade

É argumentável que a compreensão filosófica do conceito de identidade influenciou o modo como esse conceito é analisado em sistemas formais usuais, em particular da lógica clássica. De algum modo os sistemas formais clássicos assumem a TTI, no sentido de assumir ou incorporar algum aspecto da compreensão filosófica da identidade (e.g., a Lei de Leibniz) como sendo o modo de caracterizar a relação de identidade em seu sistema. Segue-se, portanto, que nas teorias que assumem a *Teoria Tradicional da Identidade*, nenhuma entidade é indiscernível de outra sem que isso resulte em serem as *mesmas* entidades (ou seja, sem que se assumam que as entidades são idênticas).<sup>6</sup>

A identidade nos sistemas clássicos é compreendida como uma relação binária (i.e., relaciona dois elementos) usualmente designada pelo símbolo “=”, garantindo as seguintes propriedades formais:

**Reflexividade:**  $\alpha = \alpha$

Todo objeto é idêntico a si mesmo.

6 - Não há um consenso sobre o que são sistemas formais clássicos, de modo que iremos compreender aqui a lógica clássica como sendo a lógica proposicional clássica, lógica clássica de predicados de *primeira-ordem* (adiante apenas lógica de *primeira-ordem* ou *elementar*) com ou sem identidade (além de alguns de seus subsistemas), as extensões conservativas da lógica de primeira-ordem, como a lógica clássica de predicados de ordem superior (teoria dos tipos), além de teorias clássicas de conjuntos, como *Zermelo-Fraenkel* (ZF), *von Neumann-Bernays-Gödel* (NBG), entre outras. Nota-se, portanto, que quando falo de *lógica* ou *lógica clássica*, assumo o que se conhece como *grande lógica*, envolvendo todos os sistemas de modo geral; se necessário, especificarei explicitamente a qual sistema estou me referindo. Assumirei que o leitor está familiarizado com a linguagem proposicional e de predicados. Para mais, ver (MENDELSON, 2010) e (KLEENE, 2002).

**Simetria:** Se  $\alpha = \beta$ , então  $\beta = \alpha$

Dado os objetos  $\alpha$  e  $\beta$ , se  $\alpha$  é idêntico a  $\beta$ , então  $\beta$  é idêntico a  $\alpha$ .

**Transitividade:** Se  $\alpha = \beta$  e  $\beta = \gamma$ , então  $\alpha = \gamma$

Dado os objetos  $\alpha$ ,  $\beta$  e  $\gamma$ , se  $\alpha$  é idêntico a  $\beta$ , e  $\beta$  é idêntico a  $\gamma$ , então  $\alpha$  é idêntico a  $\gamma$ .

**Substitutividade:** Se  $\alpha = \beta$ , então  $(\varphi(\alpha) \rightarrow \varphi(\beta))$

Dado os objetos  $\alpha$  e  $\beta$ , se  $\alpha$  é idêntico a  $\beta$ , então se uma fórmula é satisfeita por  $\alpha$ , ela será satisfeita por  $\beta$  (ou, se  $\alpha$  satisfaz um certo predicado, então  $\beta$  também satisfaz este predicado).

Estas propriedades que a identidade deve preservar garantem que ela seja uma *relação de congruência*, i.e., uma relação de equivalência que *preserva as características dos objetos* relacionados. Ao manter estas quatro características (que caracterizam uma *relação de congruência*), a identidade preserva as outras relações do sistema. Por exemplo, digamos que há em nossa linguagem uma relação binária  $R$ , que compreenderemos intuitivamente como sendo a relação de *ser pai de*. Portanto, interpretaremos  $xRz$  como sendo *x é pai de z*. Digamos agora que  $x$  é idêntico a  $y$ . Se é o caso que  $xRz$  e  $x = y$ , então se segue que  $yRz$ .  $\equiv$

Compreendemos, portanto, que a relação de identidade deve ser uma congruência. No entanto, cada sistema formal específico oferecerá algum modo para caracterizar ou definir a identidade – preservando as características apresentadas acima. Assim, há dois modos de se introduzir o conceito de identidade em um sistema formal elementar: como (1) conceito primitivo; ou como (2) conceito definido. Ao adotar um conceito como primitivo, o sistema formal deve *definir implicitamente* esse conceito através dos axiomas oferecidos. Por exemplo, nas formulações usuais de ZF a relação de *pertinência* (denotada pelo símbolo “ $\in$ ”) é tomado como primitiva no sistema, e seu funcionamento fica *implícito* através de seu uso nos axiomas oferecidos. Já ao definir um conceito, o modo comum é encontrar uma fórmula da linguagem da teoria que expresse o que se quer com o conceito. Por exemplo, na Axiomática de Peano para a aritmética, axiomatizada sobre uma linguagem de primeira-ordem, um dos conceitos usualmente adotados como primitivos é o conceito de *sucessor*, sendo algumas outras operações definidas em termos deste conceito. Como nota Krause (2002):

Não há, em princípio, qualquer conceito matemático que não possa ser definido. (...) Na verdade, tudo depende da axiomática particular que se está adotando; um conceito pode ser primitivo (logo, não definido) numa

axiomatização, mas definido em outra. (KRAUSE, 2002, p. 14-5)

A identidade é geralmente tomada como primitiva em uma lógica de predicados de primeira-ordem,<sup>7</sup> sendo os axiomas referentes à identidade:

**(Reflexividade)**  $\forall x (x = x)$

Para todo  $x$ ,  $x$  é idêntico a  $x$

**(Substitutividade)**  $\forall x \forall y (x = y \rightarrow (P(x) \leftrightarrow P(y)))$

Para todo  $x$  e  $y$ , se  $x$  é idêntico a  $y$ , e  $P(x)$  é uma fórmula qualquer que contém  $x$  livre, então  $P(y)$  resulta de  $P(x)$  pela substituição de  $x$  por  $y$  em algumas das ocorrências livres de  $x$ , desde que  $y$  seja livre para  $x$  em  $P(x)$ .

Através destes dois axiomas (e dos demais da lógica elementar) podemos obter como teorema que a identidade é simétrica e transitiva.<sup>8</sup> Devemos salientar alguns aspectos do Axioma da Substitutividade. O Axioma da Substitutividade é, na verdade, um *esquema de axiomas*. Em uma linguagem de primeira-ordem não quantificamos sobre fórmulas, de modo que o termo “ $P$ ” é uma metavaríavel que poderá ser substituído por qualquer fórmula que consta na linguagem. Por outro lado, compreende-se que linguagens de ordem-superior têm poder expressivo para que a identidade seja definida (e não tomada como primitiva) através de uma fórmula da linguagem, tal como se segue em uma lógica de segunda-ordem:

$$x = y =_{def} \forall P [P(x) \leftrightarrow P(y)]$$

onde  $x$  e  $y$  percorrem os termos individuais da linguagem e  $P$  é uma variável da própria linguagem para predicados de indivíduos.<sup>9</sup> Esta definição diz que se dois indivíduos são

7 - Quine (1986, p.63-4) ofereceu uma tentativa de definir a identidade em linguagens de primeira-ordem, mas essa alternativa enfrenta problemas. Cf. (SAVELLOS, 1990, p. 477) e (BÉZIAU, 2003). Devemos notar que podemos definir a identidade dentro de teorias específicas, erigidas sobre uma linguagem de primeira-ordem *sem* identidade – e.g., em certas formulações de ZF. Cf. (FRAENKEL; BAR-HILLEL; LEVY, 1973, p. 27).

8 - Cf. (GRACHER, 2016, p. 107-8).

9 - Devemos notar que “=def” é uma abreviação para “é igual por definição”, sendo este símbolo metateórico, não pertencendo assim à linguagem objeto em análise.

numericamente idênticos, então eles compartilham todas as mesmas propriedades. Como podemos ver, esta parece ser uma interpretação mais caridosa da *Lei de Leibniz*. Devemos notar que podemos tomar a identidade como um conceito primitivo em uma linguagem de ordem-superior com o seguinte axioma:

$$\forall x \forall y [(x = y) \leftrightarrow \forall P (P(x) \leftrightarrow P(y))]$$

Esta fórmula pode ser introduzida na linguagem como um axioma, compreendendo assim a identidade como um conceito primitivo. No entanto, é argumentável que, uma vez que uma teoria tem capacidade expressiva para definir um conceito, é formalmente – e, para um realista quanto a natureza da lógica, também seria ontologicamente – mais econômico definir tal conceito em vez de tomá-lo como primitivo. Esta ideia poderia ser expressa como uma espécie de *navalha de Ockham* aplicado a sistemas formais: Quanto menor o conjunto de conceitos primitivos adotados por uma teoria, melhor.<sup>10</sup>

## 2 Identidade e Indiscernibilidade

Podemos dizer que a *Teoria Tradicional da Identidade* “colapsa” os conceitos de identidade e indiscernibilidade. De acordo com a Lei de Leibniz, entidades são idênticas se, e somente se, são também indiscerníveis. Como vimos, as abordagens formais (usuais) do conceito de identidade são influenciadas por essa concepção, de modo que a análise da identidade é feita através da noção de indiscernibilidade. Mas será que as coisas precisam ser assim? Será que a equivalência entre identidade e indiscernibilidade precisa sempre ser assumida?

Lembrando, a Lei de Leibniz (sendo ela uma bicondicional) é equivalente à conjunção de duas teses, a Indiscernibilidade dos Idênticos (se  $x$  e  $y$  são idênticos, então são indiscerníveis) e a Identidade dos Indiscerníveis (se  $x$  e  $y$  são indiscerníveis, então são idênticos). Se negarmos uma das duas teses teremos, portanto, a falha da equivalência entre identidade e indiscernibilidade.

É comum aceitarmos a Indiscernibilidade dos Idênticos, pois é razoável a ideia que se  $x$  e  $y$  são idênticos, isso implica que eles têm as mesmas características. Pensemos na versão contrapositiva da Indiscernibilidade dos Idênticos: se há alguma característica que  $x$  tenha,

**10** - Os modos usuais de se introduzir a identidade em linguagens de predicados (seja de primeira ou de ordem-superior) enfrentam problemas. Para uma análise desses problemas, ver (GRACHER, 2016, p. 71-85).

mas não  $y$ , então  $x$  e  $y$  não são idênticos. Esta tese parece razoável. Por outro lado, será que a Identidade dos Indiscerníveis é uma tese inquestionável? Será o caso que se  $x$  e  $y$  são indiscerníveis isso implicará que eles são idênticos?

A intuição geral que temos é de preservarmos a Identidade dos Indiscerníveis. Contudo, suas principais dificuldades aparecem com o desenvolvimento das ciências modernas. Consideremos os compostos químicos. Dois átomos de hidrogênio e um de oxigênio, ligados de um certo modo, produzirão o composto químico que chamamos de “água”. Pense que um químico está no laboratório trabalhando com uma molécula de água. Ele sai do laboratório para almoçar e um amigo resolve lhe *pregar uma peça*, tirando um átomo de hidrogênio da molécula de água e colocando *outro* no lugar. Fará diferença a troca? O químico extrairá outras conclusões de seus experimentos por conta disso? A resposta é: não. Alguém poderá argumentar que houve uma modificação metafísica. No início tínhamos uma molécula de água, e agora temos “outra”. No entanto, se houve esta alteração (*strictu sensu*) significa que há alguma característica que o primeiro átomo de hidrogênio tem, que o segundo não tem. Mas há diferença entre átomos da mesma espécie? Se houver, a química encontrará grandes problemas, pois se assume que todos os átomos de uma mesma espécie são indiscerníveis entre si (*i.e.*, têm as mesmas características).

O exemplo anterior pode não convencer meu leitor mais exigente. Vejamos outro caso. Na mecânica quântica os estados de uma partícula são descritos em termos de *vetores* (em um dado espaço vetorial). Pensemos agora que temos um fóton que se localiza no polo norte, cujo estado é descrito pelo vetor  $|\psi_1\rangle$ , e um fóton no polo sul, cujos estados são descritos pelo vetor  $|\psi_2\rangle$ . Tudo ocorrerá normalmente quando tratamos cada um dos fótons separadamente. Contudo, o que ocorrerá se quisermos tratar do sistema com esses dois fótons? Teremos de fazer uma combinação linear entre esses dois vetores (que é chamado de “superposição”), que simplificarmente pode ser:

$$|\psi_1, \psi_2\rangle = a. |\psi_1\rangle + b. |\psi_2\rangle^{11}$$

A função acima é tal que qualquer medida realizada sobre o sistema conjunto, neste estado, é tal que o fóton no polo sul e o fóton no polo norte não poderão ser discernidos. Ou seja, nunca se poderá determinar qual é qual. Deste modo, temos *dois* fótons que não

**11** - Os coeficientes  $a$  e  $b$ , quando elevados ao quadrado, representarão a probabilidade do sistema  $|\psi_1, \psi_2\rangle$  de colapsar no estado  $|\psi_1\rangle$  (com probabilidade  $a^2$ ) ou  $|\psi_2\rangle$  (com probabilidade  $b^2$ ). Portanto, a soma dos coeficientes elevados ao quadrado (*i.e.*, a soma das probabilidades) será tal que  $a^2 + b^2 = 1$ . Ou seja, os estados descritos pela combinação linear descreverão todos os estados possíveis no qual o sistema combinado  $|\psi_1, \psi_2\rangle$  pode colapsar.

são discerníveis. Haveria, portanto, um caso de objetos que não são idênticos (pois não são *o mesmo*), mas são indiscerníveis – o que negaria a Identidade dos Indiscerníveis.<sup>12</sup>

Há diversas perguntas que podem ser feitas. Um leitor atento poderá, prontamente, questionar a existência de propriedades não descritas (ou capturadas) pelo formalismo. Essa alternativa é chamada de “Interpretação das Variáveis Ocultas” (ou “Interpretação de Bohm”, que faz referência ao físico David Bohm). Todavia, essa interpretação sofre de problemas latentes.<sup>13</sup> Outra pergunta relevante é que as propriedades de *estar no polo sul* e *estar no polo norte* distinguiriam os fótons. Contudo, se a teoria e suas interpretações usuais capturam corretamente os fenômenos quânticos, então todos os estados são descritos naquela combinação linear (inclusive a posição).

Se tais interpretações (que são amplamente aceitas) estão corretas, então há fortes razões para negarmos a Identidade dos Indiscerníveis. Todavia, como vimos antes, tanto a lógica (quanto a matemática) que usamos preserva o Princípio de Identidade. Se temos toda uma teoria da mecânica quântica que afirma que dois fótons podem entrar em estados nos quais eles são indiscerníveis, sobre qual teoria formal deveremos oferecer a semântica para isso? Usualmente fazemos a semântica de uma teoria sobre uma teoria usual dos conjuntos, todavia essas teorias podem ser ditas “leibnizianas”, no sentido que: se  $x$  e  $y$  estão nos mesmos conjuntos (podendo ser entendido como  $x$  e  $y$  *têm as mesmas propriedades*), então  $x$  e  $y$  são idênticos. Se fizermos a semântica das teorias quânticas em uma teoria usual de conjuntos, então não teremos uma interpretação adequada para entidades que são indiscerníveis, mas *não* idênticas.

Uma solução aparentemente viável foi apresentada por Décio Krause, através da teoria chamada de “quase-conjuntos”.<sup>14</sup> Nesta teoria apenas a *pertinência* e a *indiscernibilidade* (e não a identidade) que são tomadas como conceitos primitivos, de modo que a indiscernibilidade não equivale à identidade. Aparentemente essa é uma teoria segura para oferecermos a semântica da mecânica quântica. Contudo, o interessante desta teoria – para nossos propósitos atuais – é que ela é um sistema formal no qual a identidade não é pressuposta (ou tomada como primitiva) pelo sistema. Mas será que é assim mesmo?

---

**12** - Existem diversos casos nos quais, aparentemente, a Identidade dos Indiscerníveis falha. Um exemplo intuitivo é pensarmos no dinheiro em uma conta bancária. Duas unidades de valor são indiscerníveis entre si, ainda que não sejam a *mesma* (pois são *duas*, e não *uma* e a *mesma*). Cf. (FRENCH; KRAUSE, 2006, Cap. 1).

**13** - Cf. (FRENCH, 2015) e (FRENCH; KRAUSE, 2006).

**14** - Cf. (KRAUSE, 1992), (FRENCH; KRAUSE, 2010) e (FRENCH; KRAUSE, 2006, Cap. 7).

### 3 Indiscernibilidade como Conceito Primitivo e Críticas de Bueno

Como resposta à tese de Bueno (2014, p. 325-6), que todo sistema formal assume a identidade, Krause e Arenhart (2015, p. 3) propuseram a teoria de quase-conjuntos como contra-exemplo. Como eles notam, o conceito de indiscernibilidade pode ser analisado sem implicar, necessariamente, no conceito de identidade. Isto é, ao menos em princípio o conceito de indiscernibilidade não é logicamente equivalente ao conceito de identidade. Deste modo, como dito anteriormente, haveria sistemas conceituais nos quais a identidade não é pressuposta, mas sim a indiscernibilidade.

Bueno apresenta dois problemas à estratégia advogada por Krause e Arenhart (BUENO, 2016, p. 3). O primeiro problema afirma que mesmo que o conceito de indiscernibilidade seja tomado como primitivo, isso não implica que o conceito de identidade não seja pressuposto. Isso ocorreria uma vez que a própria tentativa de distinguir a identidade da indiscernibilidade pressupõe a identidade, pois ao dizermos que a identidade não é o mesmo conceito que a indiscernibilidade, estamos apenas dizendo que o conceito de identidade não é idêntico ao conceito de indiscernibilidade.

O segundo problema afirma que mesmo nos sistemas nos quais a indiscernibilidade é tomada como primitiva (e ela não implica na identidade), a própria formulação do conceito de indiscernibilidade naqueles sistemas pressupõe a identidade. Para exemplificar este ponto Bueno apresenta os três primeiros axiomas da teoria de quase-conjuntos, que expressam o conceito primitivo de indiscernibilidade, representado simbolicamente por “ $\equiv$ ”, como sendo uma relação de equivalência:

$$(A1) \quad \forall x(x \equiv x)$$

$$(A2) \quad \forall x \forall y(x \equiv y \rightarrow y \equiv x)$$

$$(A3) \quad \forall x \forall y \forall z((x \equiv y \wedge y \equiv z) \rightarrow x \equiv z)$$

De acordo com Bueno, nos três axiomas o conceito de identidade é pressuposto, uma vez que em (A1), por exemplo, *as mesmas* variáveis precisam ser usadas em cada lado de “ $\equiv$ ”; e nos outros axiomas *as mesmas* variáveis precisam ser ligadas pelos quantificados em cada lado de “ $\rightarrow$ ”. Ao pedir esclarecimentos quanto a esse tópico, Bueno afirmou que o papel funcional [*functional role*] dos símbolos precisam ser o mesmo.<sup>15</sup> Isso implicaria a identidade. Compreende-se a função de expressão de um símbolo como a funcionalidade

---

15 - Os pontos a seguir foram apresentados por Bueno em comunicação pessoal, expostos aqui com sua permissão.

(ou papel) que o símbolo desempenha em uma certa expressão. De acordo com Bueno, podemos dizer que as variáveis denotam algum elemento do domínio de quantificação.

Para sermos mais precisos, consideremos novamente o axioma (A1):  $\forall x(x \equiv x)$ . Vamos separar as ocorrências simbólicas, chamando de “*esq.x*” o símbolo “*x*” a esquerda de “ $\equiv$ ” e “*dir.x*” o símbolo “*x*” a direita. Seja  $\varphi(y)$  a função que designará o que é expresso (interpretado, ou denotado) pelo símbolo “*y*”. Ou seja,  $\varphi(y)$  será uma abreviação para “a função expressiva de *y*”, que será o objeto denotado pela variável *y*.<sup>16</sup> Seguindo as definições anteriores, podemos transcrever o que Bueno afirma como:

$$\varphi(\textit{esq.x}) = \varphi(\textit{dir.x})$$

Podemos nos perguntar agora: por que não é o caso que  $\varphi(\textit{esq.x}) \equiv \varphi(\textit{dir.x})$ ? Ou seja, por que não é o caso que a função expressiva dos símbolos não sejam apenas indiscerníveis, em vez de idênticas?

O problema com essa alternativa, de acordo com Bueno, é que a indiscernibilidade não é “forte o suficiente para caracterizar o conteúdo dos enunciados em questão, já que indiscerníveis podem ser distintos”. Assim, quando afirmamos que a expressão  $(A \vee \neg A)$  é uma verdade lógica (ao menos de acordo com a lógica clássica), é necessário “excluir a possibilidade de que a segunda ocorrência da variável ‘*A*’ seja distinta da primeira”, denotando assim uma proposição diferente daquela que denota a primeira ocorrência de “*A*”. Ou seja, temos de excluir a possibilidade de obter, a partir de  $(A \vee \neg A)$ , a fórmula  $(A \vee \neg B)$  – que evidentemente não é uma verdade lógica. Se as ocorrências de “*A*” forem apenas indiscerníveis, “não temos como garantir a exclusão dessa possibilidade, já que indiscernibilidade não implica identidade”. Deste modo, a identidade é requerida para garantir a própria formulação do conteúdo dos enunciados em apreço.

---

**16** - Devo notar que, de acordo com Bueno, o papel funcional de um símbolo não necessariamente é uma função (no sentido matemático do termo), mas apenas o papel que desempenha numa expressão. Portanto, a utilização que faço de “ $\varphi(y)$ ” não deve ser entendido, diretamente, como uma função metalinguística, mas sim uma simbolização que visa deixar *mais claro* o que estamos tratando. De modo geral, estamos em um nível intuitivo. Contudo, nada nos impede de formalizar tais intuições – pois, basicamente, foi isso que nos levou à lógica formal.

## 4. Objeções a Bueno

Os argumentos apresentados por Bueno, contra a estratégia de Krause e Arenhart, são fortes uma vez que se baseiam em uma análise tradicional (e intuitiva) do modo como se estruturam sistemas formais. Mas essa posição enfrentará alguns problemas interessantes.

### 4.1 Variáveis como Denotadores

A primeira objeção recai sobre a ideia de que variáveis têm como *papel funcional* denotar algum elemento de um dado domínio. As variáveis de uma certa linguagem devem ser entendidas como posições “em branco”, de modo que toda a fórmula só se tornará uma sentença quando tais posições forem preenchidas por termos (que, essas sim, têm papel denotativo).<sup>17</sup> Não devemos tomar as variáveis como tendo papel denotativo – ou como nomes próprios dos elementos do domínio. Enquanto nomes são intersubstituíveis por terem a mesma intenção, duas variáveis devem ser mantidas distintas mesmo que seu domínio seja determinado por um mesmo conceito. Alonzo Church usa do seguinte exemplo para esclarecer tal ponto:

Cada uma das letras  $x$  e  $y$  são variáveis cujo domínio são os números reais, sendo nós obrigados a distinguir as duas desigualdades  $x(x+y) \geq 0$  e  $x(x+x) \geq 0$  como diferentes – de fato, a segunda desigualdade é universalmente verdadeira, e a primeira não. (CHURCH, 1956, p. 10 – *Trad. Nossa*)

#### 4.1.1 Fortalecendo o Argumento de Bueno

Podemos fortalecer o argumento de Bueno, tendo em vista o problema apresentado. Seguindo nossa leitura do raciocínio de Church, assumimos que as desigualdades, por ele mencionadas, são distintas devido ao fato que  $x$  e  $y$  são variáveis distintas (o que implicará no uso da identidade). Também se assume que as três ocorrências de  $x$  em  $x(x+x) \geq 0$ , quando forem substituídas por um termo, devem ser substituídas pelo mesmo termo (e a

<sup>17</sup> - Esta analogia é oferecida por Tarski: “The role of the variable in a sentential function has sometimes been compared very adequately with that of the blanks left in a questionnaire; just as the questionnaire acquires a definite content only after the blanks have been filled in, a sentential function becomes a sentence only after constants have been inserted in place of the variables.” (TARSKI, 1993, p. 5).

identidade estará em jogo de novo). E, por fim, quando oferecermos uma estrutura que modele nossa teoria e uma função interpretação, que interpretará os termos da linguagem em elementos do domínio de nossa estrutura, a ocorrência de um mesmo termo precisa ser interpretado como *o mesmo* elemento do domínio (e eis a identidade novamente).

Isto é, seja  $D$  o domínio da estrutura que modele nossa linguagem,  $\rho$  uma função interpretação e  $a$  e  $b$  duas constantes da linguagem, tais que  $\rho(a) = a^D$  e  $\rho(b) = b^D$  (sendo “ $a^D$ ” e “ $b^D$ ” nomes de elementos de  $D$ ). Usando da estratégia de separar as ocorrências simbólicas, temos o *esq.x*, *meio.x* e *dir.x*, que seriam as três ocorrências de  $x$  na expressão  $x(x + x) \geq 0$ . Precisamos garantir que a substituição de *esq.x*, *meio.x* e *dir.x* seja feita pelo mesmo termo (por exemplo,  $a$ ), tal que  $(esq.x) = (esq.a)$ ,  $(meio.x) = (meio.a)$  e  $(dir.x) = (dir.a)$ , para que então, usando a função interpretação, possamos obter  $\rho(esq.a) = a^D$ ,  $\rho(meio.a) = a^D$  e  $\rho(dir.a) = a^D$ . Ou seja, precisamos usar *duplamente* a identidade, com efeito, primeiro devemos garantir que substituimos as três ocorrências de  $x$  pela *mesma* constante, para que disso elas possam ser interpretadas pelo *mesmo* elemento do domínio. Se substituirmos *esq.x*, *meio.x* e *dir.x* de modo diferente, então podemos obter uma expressão na forma  $a(a + b) \geq 0$  a partir de  $x(x + x) \geq 0$  – e isso só pode ocorrer se já obtivermos antes, como teorema, que  $a = b$  (mas pressupomos a identidade aqui também). Além do mais, caso as interpretações das ocorrências simbólicas sejam diferentes, como  $\rho(esq.a) = a^D$ ,  $\rho(meio.a) = a^D$  e  $\rho(dir.a) = b^D$ , obteremos que  $a^D(a^D + b^D) \geq 0$ . Anão ser que tenhamos que  $a^D = b^D$  (a que pressupõe a identidade novamente, não seja uma relação entre termos da linguagem, mas entre elementos do domínio), isso resultará em problemas também.

Portanto, o argumento inicial de Bueno era que no *papel funcional* das variáveis, em  $\forall x(x \equiv x)$ , as ocorrências de  $x$  precisavam ser as mesmas. No entanto, o argumento dele se torna fortalecido quando falamos também em substituição de variáveis por constantes. Pois agora precisamos garantir que as ocorrências das variáveis  $x$  precisam ser substituídas pelo *mesmo* termo, e que termos iguais precisam ser interpretados por elementos iguais da estrutura.

#### 4.1.2 Respondendo ao Argumento Fortalecido

Podemos distinguir dois níveis (ou tipos de usos) da identidade. (1) Um nível trata da identidade sobre a linguagem, nos referindo ao uso de variáveis e substituição de variáveis por termos. (2) Outro nível trata da identidade dentro das estruturas que modelam a linguagem, nos referindo à interpretação e aos elementos do domínio da estrutura. Nos dois níveis Bueno enfrentará problemas.

(1) No uso da identidade sobre a linguagem, podemos perguntar: qual identidade estamos usando? Aparentemente Bueno poderá formular duas respostas aqui.

(1.a) A identidade que usamos para falar *sobre* a linguagem é uma relação de identidade intuitiva. Ou seja, intuitivamente dizemos que existe uma identidade entre as duas ocorrências da variável  $x$  em  $\forall x (x \equiv x)$ , ou entre duas ocorrências de um termo, como as duas ocorrências da constante  $a$  em  $a = a$ . Mas será que é essa identidade intuitiva que é fundamental? O que determina algo ser ou não intuitivo? E o que garante que a identidade intuitiva é, de fato, uma relação de *identidade* e não uma relação de *indiscernibilidade*?<sup>18</sup>

(1.b) Outra possível resposta é afirmar que a noção de identidade utilizada, quando falamos *sobre* uma linguagem, é uma noção metalinguística de identidade. Ou seja, distinguimos dois níveis de linguagens, a linguagem objeto – onde ocorre o uso das variáveis e termos –, e outro nível é a metalinguagem desta linguagem objeto – que é onde temos a noção de identidade que nos permite afirmar que duas variáveis (ou termos) da linguagem objeto são ou não idênticas. Um problema que Bueno encontraria agora é: podemos oferecer uma metalinguagem onde os termos ou variáveis da linguagem objeto são tomados não como idênticos, mas apenas como indiscerníveis. Isto é, podemos oferecer uma metalinguagem quase-conjuntista para a linguagem que é objeto de nossa análise, o que vai contra a sua tese.

(2) Temos, por fim, o segundo nível no qual tratamos a identidade, *i.e.*, quando nos referimos à identidade entre elementos do domínio da estrutura. Pressupomos essas relações de identidade quando oferecemos uma semântica clássica para a nossa linguagem, que é usualmente feita sobre uma teoria clássica de conjuntos (como *ZF*). Em princípio poderíamos oferecer também uma semântica não-clássica, onde a indiscernibilidade é primitiva (e não a identidade), e os elementos de nosso domínio podem ser apenas elementos indiscerníveis. Nessa semântica (provavelmente erigida sobre uma teoria de quase-conjuntos, ou similar), a indiscernibilidade entre elementos não implicaria na identidade entre eles. Nesta alternativa não precisamos garantir que a interpretação das duas ocorrências de uma constante  $a$  em  $a = a$  sejam tais que  $\rho(esq.a) = \rho(dir.a)$ , mas

---

**18** - Existem problemas com a possibilidade de analisarmos conceitos tomados como primitivos (ou, em outros termos, intuitivos). Bueno assume que conceitos são determinados por suas intenções, de modo que a identidade entre conceitos é determinada por terem ou não a mesma intenção. Em (GRACHER, *Manuscrito*), faço uma análise da posição de Bueno e apresento um conjunto de objeções a ele, sendo algumas dessas objeções pertinentes ao tópico que aqui tratamos.

apenas que  $\rho(esq.a) \equiv \rho(dir.a)$ .

Contudo, Bueno afirma que isso não poderia ocorrer, uma vez que a indiscernibilidade não é “forte o suficiente para caracterizar o conteúdo dos enunciados em questão, já que indiscerníveis podem ser distintos”. Ou seja, o papel funcional dos termos precisam ser tais que denotem entidades diferentes (ou iguais), e não apenas discerníveis (ou indiscerníveis) – uma vez que, segundo Bueno, entidades indiscerníveis poderiam ser diferentes. Mas essa interpretação da indiscernibilidade padecerá de um grave problema que veremos a seguir.

## 4.2 Implicação Modal da Indiscernibilidade

Bueno afirma que entidades indiscerníveis podem ser distintas – ou que a indiscernibilidade *não exclui* a possibilidade das entidades serem distintas. O que se quer afirmar com o termo “podem”? Aparentemente a interpretação modal que podemos oferecer da tese de Bueno é:

$$(x \equiv y) \rightarrow \diamond(x \neq y).$$

Ou seja, se  $x$  e  $y$  são indiscerníveis, então é possível que eles sejam distintos. No entanto esta afirmação sofre de vários problemas. Se  $(x \equiv y) \rightarrow \diamond(x \neq y)$ , e uma vez que é razoável que  $(x \equiv y)$ , então disso se seguirá que  $\diamond(x \neq x)$  - o que não é minimamente razoável. Além disso, se é o caso que  $(x \equiv y) \rightarrow \diamond(x \neq y)$ , então sua contrapositiva também é verdadeira, *i.e.*,  $\Box(x = y) \rightarrow \diamond(x \neq y)$ . Ou seja, se é uma verdade necessária que  $x$  e  $y$  são idênticos, isso implicaria que  $x$  é discernível de  $y$  – o que, novamente, não é razoável. Vejamos dois outros raciocínios, cujas conclusões baseiam-se na interpretação de Bueno para a indiscernibilidade e em qualquer sistema modal usual, como  $K$ :

[1.]  $(x \equiv y) \rightarrow \diamond(x \neq y)$  - Tese de Bueno

[2.]  $(x = y) \rightarrow (x \equiv y)$  - Indiscernibilidade dos Idênticos

[3.]  $(x = y) \rightarrow \diamond(x \neq y)$

Ou seja, se aceitarmos que a indiscernibilidade não exclui a possibilidade das entidades indiscerníveis serem diferentes, e uma vez que a identidade implica na indiscernibilidade, então a própria identidade não garantirá tal exclusão defendida por Bueno. Outro argumento se segue do seguinte modo, aqui exposto em linhas gerais e sem detalhes óbvios:

[1.]  $(x = y) \rightarrow \Box(x = y)$  - Necessidade da Identidade

[2.]  $(x = y) \rightarrow \Diamond(x \neq y)$  – Resultado do Argumento Anterior

[3.]  $\Box(x = y) \rightarrow (x \neq y)$  – Contraposição de 2.

[4.]  $(x = y) \rightarrow (x \neq y)$

Os dois argumentos anteriores se seguem por conta de assumirmos que a identidade implica a indiscernibilidade – Indiscernibilidade dos Idênticos. Uma estratégia possível de Bueno seria afirmar que tal princípio é falso:<sup>19</sup>

$$(x = y) \wedge (x \neq y)$$

Ou seja, ainda que duas entidades sejam idênticas, isso não implicará que elas sejam indiscerníveis. Todavia, Bueno encontrará, novamente, o mesmo problema apontado anteriormente: a identidade não garantirá a exclusão da possibilidade das entidades terem características distintas, pois ainda que sejam idênticas, elas seriam discerníveis – e, se são discerníveis, há algo que diferencia uma da outra, o que iria contra a tese de Bueno.

Com os argumentos anteriores estabelecemos que a indiscernibilidade não implica na possibilidade dos objetos serem diferentes. A única coisa que a indiscernibilidade nos diz é que as entidades indiscerníveis têm, efetivamente, as mesmas características. Há casos nos quais indiscerníveis são objetos diferentes,<sup>20</sup> como há casos nos quais indiscerníveis são idênticos. Portanto, as implicações modais que podem ser inferidas do argumento de Bueno aparentemente não se seguem. Dito isto, parece razoável a empreitada de oferecermos

<sup>19</sup> - Como notou Arenhart (comunicação pessoal), a tese de Bueno vai contra a abordagem usual da identidade em cálculos modais, que mantêm a *necessidade da identidade*. Portanto, outra rota alternativa que Bueno poderia tomar seria adotar cálculos modais com identidade contingente.

<sup>20</sup> - Tal como discutido anteriormente, casos que são descritos por certas interpretações em mecânica quântica.

uma semântica quase-conjuntista para um sistema formal clássico, de modo que a função interpretação das duas ocorrências de uma constante  $a$ , em  $a = a$ , serão tais que  $\rho(esq.a) \equiv \rho(dir.a)$  (e não  $\rho(esq.a) = \rho(dir.a)$ ).<sup>21</sup>

### 4.3 Indiscernibilidade e Verdades Lógicas

Outra objeção ao argumento de Bueno seria a seguinte. Tome duas proposições logicamente equivalentes,  $P$  e  $Q$ . Seja  $P$  a proposição  $(A \rightarrow B)$   $Q$  proposição  $(\neg A \vee B)$ . Indiferente ao modo como expressamos linguisticamente essas proposições,<sup>22</sup>  $P$  e  $Q$  são proposições diferentes –  $P$  é uma proposição condicional enquanto  $Q$  é uma disjunção. No entanto,  $P$  e  $Q$  são indiscerníveis no tocante as características relevantes para afirmarmos que  $(P \vee \neg Q)$  é uma verdade lógica (ao menos na lógica clássica). O que temos de levar em conta não é se as proposições são *as mesmas* (no sentido identitativo), mas sim se suas funções de verdade (ou atribuições) são indiscerníveis – lembrando que indiscernibilidade não implica que seja *possível* haver funções de verdade (ou atribuições) diferentes.

Outro aspecto interessante é pensarmos nas álgebras de Lindenbaum, onde obtemos o conjunto quociente (ou das *classes de equivalência*) de todas as tautologias (que chamaremos de  $\Gamma$ ). Sejam  $X$  e  $Y$  duas proposições pertencentes  $\Gamma$ . Segue-se que  $(X \vee \neg Y)$  é uma tautologia, ainda que o papel funcional de  $X$  seja diferente da de  $Y$ . Fica claro com isso que o importante para ser uma tautologia não é que  $\Phi(X) = \Phi(Y)$  – pois as proposições podem ser completamente diferentes –, mas sim que  $X$  e  $Y$  tenham funções de verdade indiscerníveis. A indiscernibilidade, nesse caso, não trará perda teórica.

## 5 Conclusão

Ainda que os argumentos de Bueno sejam razoáveis, não parece que eles se sustentam de modo geral. A posição de Bueno é forte, uma vez que se baseia em uma análise intuitiva (e clássica) do modo como se estruturam nossos sistemas formais. De fato a tradição atribuiu um papel central para a identidade nos mais diversos aspectos. Contudo, ao analisarmos caridosamente posições não-*standard* (que negam o papel fundamental da identidade, por

21 - Esta noção parece ter sido empregada pelo próprio Bueno (2000) em seu artigo que explora a relação entre a teoria de quase-verdade e quase-conjuntos. Cf. (BUENO, 2000, p. 42).

22 - Devo notar que não assumo uma posição “bourbakista” da lógica ou matemática. Isto é, matemática não é a mera manipulação de símbolos sobre um papel, mas sim algo que corresponde a uma entidade independente da linguagem ou notação usada.

exemplo), podemos ver que há dificuldades latentes a essa posição tradicional face essas posições e aos problemas que elas apresentam.

## Referências

- BÉZIAU, J.-Y. 2000. Quine on identity. *Principia: an international journal of epistemology*, Florianópolis, v. 7, n. 1-2, p. 1-15.
- BRAIDA, C.R.; KRAUSE, D. 2013. *Ontologia II*. 2 ed. Florianópolis: Universidade Federal de Santa Catarina.
- BUENO, O. 2000. Quasi-truth in quasi-set theory. *Synthese, Springer*, v. 125, n. 1-2, p. 33-53.
- BUENO, O. 2014. Why identity is fundamental. *American Philosophical Quarterly*, v. 51, n. 4.
- BUENO, O. 2016. Identity in physics and elsewhere. *Cadernos de História e Filosofia da Ciência (no prelo)*.
- CHURCH, A. 1956. *Introduction to mathematical logic*. Princeton: Princeton University Press.
- FRAENKEL, A.A.; BAR-HILLEL, Y.; LEVY, A. 1973. *Foundations of set theory*. Elsevier.
- FRENCH, S. 2015. Identity and individuality in quantum theory. In: ZALTA, E.N. (ed.). *The Stanford Encyclopedia of Philosophy*. Fall 2015.
- FRENCH, S.; KRAUSE, D. 2006. *Identity and physics: a historical, philosophical and formal analysis*. Oxford: Oxford University Press.
- FRENCH, S.; KRAUSE, D. 2010. Remarks on the theory of quasi-sets. *Studia Logica*, v. 95, n. 1-2, p. 101-124.
- GRACHER, K. 2016. *É a identidade fundamental?* Florianópolis. 113p. Dissertação (Mestrado). Universidade Federal de Santa Catarina.
- GRACHER, K. *Manuscrito*. Identidade e sistemas conceituais.
- KLEENE, S.C. 2002. *Mathematical logic*. Courier Corporation.
- KORFMACHER, C. 2006. Personal identity. *Identity encyclopedia of philosophy*.

- KRAUSE, D. 1992. On a quasi-set theory. *Notre Dame Journal of Formal Logic*, Duke University Presse, v. 33, n. 3, p. 402-11.
- KRAUSE, D. 2002. *Introdução aos fundamentos axiomáticos da ciência*. E.P.U.
- LOCKE, J. 1996. *An essay concerning human understanding*. Abridge: Hackett Publishing Company, Inc.
- MCGINN, C. 2000. *Logical properties: identity, existence, predications, necessity, truth*.
- MENDELSON, E. 2010. *Introduction to mathematical logic*. 5 ed. New York: Wadsworth and Brooks.
- NOONAN, H.; CURTIS, B. Identity. In: ZALTA, E.N. (ed.). *The Stanford Encyclopedia of Philosophy*. Summer 2014.
- QUINE, W.V.O. 1986. *Philosophy of Logic*. Harvard: Harvard University Press.
- RODRIGUEZ-PEREYRA, G. No Prelo. The principles of contradiction, sufficient reason, and identity of indiscernibles. In: ANTOGNAZZA, M.R. (Ed.). *Oxford handbook of Leibniz*. Oxford: Oxford University Press.
- SAVELLOS, E.E. 1990. On defining identity. *Notre Dame Journal of Formal Logic*, University of Notre Dame Press, v. 31, n. 3, p. 476-484.
- SIDER, T.; CONEE, E. 2005. *Riddles of existence: a guided tour of metaphysics*.
- TARSKI, A. 1993. *Introduction to logic and to the methodology of the deductive sciences*. Oxford: Oxford University Press.

# Uma Explicação Pré-semântica das Descrições Referenciais<sup>1</sup>

## *A pre-semantic account of referential descriptions*

Fernando Furtado  
Universidade de Lisboa – LanCog  
CAPES – Brasil

### Resumo

Este artigo se divide em duas partes claramente distintas: uma parte expositiva; visa fornecer ao leitor as bases para a compreensão da discussão acerca das descrições definidas. E uma parte positiva: é apresentada uma proposta de tratamento aos casos problemáticos de Donnellan em defesa da análise russelliana das descrições definidas. Na primeira parte é apresentada uma noção geral do que são descrições definidas e a análise russelliana dessas expressões. Na sequência, são apresentados os contraexemplos de Donnellan que dão suporte a suposição de que as descrições definidas podem, *prima facie*, ser usadas de pelos menos de duas maneiras: *atributivamente* e *referencialmente*. Feitas as devidas considerações, são apresentadas algumas vias de resposta à Donnellan em defesa Russell. A parte mais positiva deste artigo é uma defesa da análise russelliana baseada na ideia que os contraexemplos de Donnellan não estão ligados

---

1 - Ensaio escrito no âmbito da disciplina Tópicos de Epistemologia e Metafísica II: Linguagem e Realidade ministrada pela professora Adriana Silva Graça e pelo professor Pedro Galvão no ano letivo 2014-15 na Faculdade de Letras da Universidade de Lisboa.

2 - Fernando Furtado é estudante de doutoramento na Universidade de Lisboa, membro do LanCog e bolsista CAPES-Brasil. (fernandofurtado@campus.ul.pt).

a eventos propriamente *semânticos*. Seguindo Devitt (1981) e Evans(1982), sugiro que os contraexemplos são de fato *pré-semânticos*. Sendo assim, pode haver uma via de resposta a Donnellan em defesa de Russell ainda talvez pouco explorada.Devitt argumenta que os contraexemplos podem ser tratados no nível pré-semântico, porém pensa que isso é razão para supor que a teoria russelliana é incompleta. Esse resultado é exatamente o oposto do que é pretendido pela proposta oferecida aqui.

## Palavras-chave

Descrições definidas; Contraexemplos de Donnellan; Análise russelliana das descrições.

## Abstract

This paper has two distinct parts: an expositive one, providing the basis for an understanding of theory of definite descriptions. And a positive part: it is presented an account of Donnellan's puzzling cases for Russellian analysis of definite descriptions. First, it is introduced a general account of definite descriptions and the Russellian analysis. Then, Donnellan's counterexamples are introduced in order to support the assumption that definite descriptions can be used *prima facie* in at least two distinct ways: *attributively* and *referentially*. After some considerations, some ways to respond Donnellan, in defense of Russell, are presented. The positive part provides a defense of Russellian analysis based on the ideia that Donnellan's counterexamples don't have to do with properly *semantic* events. Following Devitt (1981) and Evans (1982), I suggest that the counterexamples are in fact *pre-semantic*. So, a new line of response to Donnellan can emerge. Devitt argues that the counterexamples can be treated at a pre-semantic level; however, he thinks this is a reason to assume that Russellian theory is an incomplete one. This result is exactly the opposite my account.

## Key-words

Definite descriptions. Donnellan's counterexamples. Russellian analysis of definite descriptions.

As frases ‘o melhor futebolista do mundo é português’, ‘a capital de Portugal é a cidade mais bonita do país’, ‘o ponto mais ao ocidente da Europa continental está em Cabo da Roca’, ‘o número par e primo é o número 2’ são casos típicos de frases onde correm descrições definidas; expressões que tipicamente desempenham o papel de sujeitos gramaticais nas frases onde ocorrem e têm sua forma geral ‘o  $\alpha$ ’ que aparece em frases como ‘o  $\alpha$  é  $\beta$ ’ ou ‘o  $\alpha$  tal que...’. Um problema que tem afligido os filósofos é fornecer uma análise correta para expressões do tipo ‘o  $\alpha$ ’. Trata-se de termos singulares semelhantes a expressões indexicais como ‘eu’, ‘tu’, ‘ele’ ou há uma análise adequada que suporte a ideia que descrições são termos gerais como ‘português’, ‘brasileiro’, ‘presidente’, ‘estudante’, etc.? Em um primeiro contato talvez seja mais natural pensar que descrições se assemelham mais a termos singulares como ‘eu’. Assim como a palavra ‘eu’ refere-se a uma pessoa particular, a expressão ‘o melhor futebolista do mundo’ também se refere a uma pessoa particular, nomeadamente Cristiano Ronaldo. Por outro lado, expressões como ‘português’ ou ‘brasileiro’ obviamente dizem respeito a muitas pessoas. Desse modo, talvez seja mais natural uma análise das descrições que as trate como termos singulares. É exatamente o oposto do que Bertrand Russell propôs. Após a publicação do seu seminal artigo “On Denoting” (Russell, 1905) passou a ser aceito por quase a totalidade dos filósofos que as descrições definidas são expressões de quantificação analisadas com base em termos gerais. O grande apelo intuitivo e elegância teórica da análise russelliana das descrições definidas fez com que durante várias décadas a proposta de Russell ficasse quase completamente intocada e considerada como paradigma de análise filosófica bem-sucedida. As primeiras grandes objeções apareceram apenas muitos anos depois com a publicação de “On Referring” (Strawson, 1950). Para as finalidades deste artigo iremos considerar principalmente as objeções apresentadas por (Donnellan, 1966) em seu influente “Reference and Definite Description”.

A estratégia de exposição neste artigo será inicialmente apresentar uma noção geral do que são descrições definidas e em seguida apresentar a teoria russelliana das descrições definidas; inclui-se nesta passagem a análise lógica das descrições no modo como é usual encontrar na literatura especializada atualmente. Em seguida serão apresentados os contraexemplos de Donnellan que dão suporte, segundo o autor, a suposição de que as descrições definidas podem, *prima facie*, ser usadas de pelos menos de duas maneiras: *atributivamente* (caso em que a análise russelliana ao que tudo indica está correta) e *referencialmente* (caso que não é previsto pela teoria russelliana).

Feitas as devidas considerações, serão apresentadas algumas vias de resposta à Donnellan

em defesa da análise russelliana das descrições, nomeadamente veremos as ideias de Kripke para acomodar os contraexemplos de Donnellan.

A parte mais positiva deste artigo é uma defesa da análise russelliana baseada na ideia que os contraexemplos de Donnellan não estão ligados a eventos semânticos. Os contraexemplos são de fato pré-semânticos, como sugerido por (Devitt, 1981) e (Evans, 1982).<sup>3</sup> Se este for o caso, pode haver uma via de resposta a Donnellan em defesa de Russell ainda talvez pouco explorada.

## A Teoria das Descrições Definidas

Nesta seção apresento em linhas gerais a teoria das descrições definidas de Bertrand Russell. Essa teoria foi apresentada em seu artigo ‘On Denoting’ de 1905. Uma das ideias principais apresentadas no artigo é que as descrições definidas não são termos singulares como nomes próprios ou demonstrativos. Na verdade, em sua lógica profunda, descrições definidas são expressões de quantificação. A ideia é que termos singulares contribuem para o significado das frases onde ocorrem apenas *indicando* ou *referindo* os objetos dos quais dizem respeito; são como ‘etiquetas coladas’ nos objetos que vez ou outra os falantes lançam mão para referir a objetos. Teorias do tipo mencionada são chamadas ‘teorias da referência direta’. A extensão dessa ideia para as descrições definidas gera estranhos *puzzles* aos quais Russell pretende resistir com a sua teoria das descrições definidas. Desse modo, pode ser razoável antes de avançarmos a teoria das descrições propriamente dita mencionar rapidamente os *puzzles* em causa.

## O caráter informativo de afirmações de identidade

Considere as seguintes frases:

- i) O melhor futebolista do mundo é o atacante português do Real Madrid.
- ii) O melhor futebolista do mundo é o melhor futebolista do mundo.

---

3 - Devitt argumenta a favor da ideia de que os contraexemplos podem ser tratados no nível pré-semântico, porém pensa isso é razão para pensar que a teoria russelliana é incompleta. Isso é exatamente o oposto do que pretendo defender aqui.

Qualquer falante da língua portuguesa sabe que ii) é verdadeira, além disso a frase é completamente trivial; é incapaz de adicionar qualquer informação ao conjunto de crenças de qualquer pessoa que a ouve. O mesmo não acontece com a frase i). Alguém pode saber que Cristiano Ronaldo é o melhor jogador do mundo e não saber para qual clube ele joga (alguém, por exemplo, que se interessa apenas pelos jogos da seleção nacional). Nesse caso, a frase i), se verdadeira, traz nova informação à pessoa que a ouve; ele poderia dizer ao ouvir a frase: - ah, o Ronaldo joga no Real? Não sabia! Mesmo que dissesse tal coisa não diríamos que a pessoa em causa não compreende perfeitamente o que é dito, na verdade ela obtém nova informação exatamente porque é capaz de compreender o que a frase diz. Se supusermos que as descrições definidas são como etiquetas que contribuem para as frases onde ocorrem apenas por indicar os respectivos objetos, então não temos ferramentas para explicar como i) pode ser informativa enquanto ii) não pode, dado que ambas são afirmações de identidade verdadeiras acerca do mesmo objeto.

## O problema do valor de verdade de frases com termos vazios

i) A raiz de  $-2$  é um número inteiro.

Supondo que a raiz de  $-2$  não exista, a frase i) não pode ser verdadeira, dado que a descrição 'a raiz de  $-2$ ' não *apanha* objeto algum (não se pode colar etiquetas em objetos que não existem). Porém, soba suposição que o princípio clássico de bivalência está correto, se uma frase qualquer é falsa, então a sua negação tem que ser verdadeira. Mas a frase 'a raiz de  $-2$  não é um número inteiro' também (e pelas mesmas razões) é falsa. Há muitos casos semelhantes que podem ser formulados para explicar o problema em causa.

## O problema das atitudes proposicionais

O problema das atitudes proposicionais diz respeito ao comportamento de descrições em contextos de atitudes proposicionais. Atitude proposicional é uma relação entre um

sujeito e uma proposição (conteúdo proposicional). *Acreditar que p*, *desejar que p* e *saber que p* são exemplos típicos de atitudes proposicionais quando *p* é uma proposição qualquer. Tipicamente considera-se a cláusula *que* para as atitudes proposicionais; a palavra ‘que’ aparece precedida de um verbo que representa a atitude e seguida por uma proposição que é objeto do verbo. Considere o seguinte exemplo:

i) Filipe III deseja saber se o autor de *Dom Quixote* é Miguel de Cervantes.

ii) Filipe III deseja saber se o autor de *Dom Quixote* é o autor de *Dom Quixote*.

Não é difícil imaginar um contexto no qual i) é verdadeira e ii) é falsa. Basta supor que Filipe III já tinha tido a oportunidade de ler *Dom Quixote*, mas que, por alguma razão irrelevante para nós, não sabia que este livro havia sido escrito por Cervantes. Neste contexto, é bastante razoável supor que i) seja verdadeira. Já no que diz respeito à frase ii), é bastante mais difícil (sem grandes extravagâncias) pensar em um caso no qual ela seja verdadeira. Filipe III dificilmente iria querer saber se o autor de *Dom Quixote* é o autor de *Dom Quixote*, i.e., se o autor de *Dom Quixote* é idêntico a si próprio – uma verdade lógica trivial. Mas como explicar essa diferença? Se a teoria da referência direta é verdadeira, então descrições definidas são apenas etiquetas dos objetos e contribuem para o significado das frases onde ocorrem apenas *apontando* o objeto relevante. Sendo assim, não temos ferramentas para explicar a diferença de valor de verdade entre i) e ii).

Com base nessas motivações e em outras, Russell rejeita a ideia segundo a qual as descrições definidas são termos que referem diretamente os objetos, porém rejeitar essa explicação não basta. Há também de se explicar como as descrições definidas funcionam. E é exatamente o que Russell faz com a sua teoria das descrições definidas.

Segundo Russell, as descrições definidas, apesar de em um primeiro contato parecerem termos singulares, são, em última análise, termos gerais com um certo mecanismo semântico pelo qual é capaz de denotar um e apenas um objeto particular. Por exemplo, ‘o maior atleta do século passado’ é uma descrição que designa apenas um objeto (nota-se pela presença do artigo definido ‘o’) por via de certa característica. Assim, uma frase como ‘o maior atleta do século passado é Pelé’, apesar de superficialmente ser uma afirmação de identidade entre dois termos singulares, em sua *gramática profunda*, traz um termo geral em uma afirmação de quantificação analisada do seguinte modo.

- i) Existe um maior atleta do século passado.
- ii) É apenas um. E,
- iii) É Pelé.

Assim, a frase ‘o maior atleta do século passado é Pelé’ é a conjunção de i)-iii). Em uma formulação geral as descrições definidas têm a forma lógica profunda com a seguinte:

$$\exists x F x [\forall y (Fy \rightarrow x = y) \wedge Gx]$$

Essa fórmula resulta na análise geral das descrições definidas.

$\exists x F x$ : equivale a i). (Critério existencial)

$\forall y (Fy \rightarrow x = y)$ : equivale a ii). (Critério de unicidade)

$Gx$ : equivale a iii) (predicação)

Não aparece em lugar algum na fórmula referência a objetos particulares. Tudo o que há são variáveis ligadas a quantificadores e predicados (termos gerais). Essa, portanto, é a teoria das descrições definidas de Russell, uma teoria que analisa as descrições definidas por meio de termos gerais.<sup>4</sup>

Muito rapidamente podemos ver como a teoria de Russell tem ferramentas capazes de dar conta dos *puzzles* apresentados.<sup>5</sup> No que diz respeito ao caráter informativo de

4 - O objetivo de Russell vai além de fornecer uma análise para as descrições definidas em termos de expressões de quantificação. Russell propõe também a análise dos nomes próprios da linguagem natural em termos de descrições definidas e novamente estas em termos de expressões de quantificação. Se essa estratégia funcionar, então boa parte daquilo que supostamente são os melhores candidatos a termos singulares são, em sua lógica profunda, termos gerais disfarçados. Russell reserva em sua teoria lugar ao que ele chama ‘nomes próprios genuínos’ que são no geral demonstrativos como ‘isto’, ‘este’, etc.

5 - Não é objetivo aqui ser exaustivo acerca das discussões dos puzzles que são ainda hoje muito problemáticos. O objetivo aqui é apenas indicar como as ferramentas da teoria de Russell parecem lidar mais facilmente com os puzzles formulados aqui.

identidades como ‘o melhor futebolista do mundo é o atacante português do Real’ não há mistério algum. Com base na análise russelliana não há pressuposto que as descrições são meras etiquetas dos objetos que denotam, ao contrário, as descrições apanham o objeto por via de uma determinada propriedade (predicado) que um tal particular (objeto) exemplifica (satisfaz) unicamente. Nesse caso, não há mistério em encontrar casos nos quais a afirmação de que duas propriedades são exemplificadas pelo mesmo particular seja verdadeira e relevante. No caso relevante aqui, ‘ser o melhor futebolista do mundo’ e ‘ser o atacante português do Real’ são duas propriedades que Cristiano Ronaldo, e só ele, possui e não é estranho que essa afirmação seja relevante (tenha conteúdo informativo).

As expressões vazias também não serão problemáticas. Como não se pressupõe a referência direta, não se pressupõe que o objeto tenha que existir para poder afirmar coisa acerca dele. As descrições são expressões de quantificação e, como vimos, não aprecem referência a objetos em sua análise. Por essa via podemos dizer coisas acerca de objetos sem pressupor que tal objeto exista.<sup>6</sup>

Essa foi uma rápida apresentação da teoria das descrições definidas de Russell que será discutida ao longo de todo este artigo. O objetivo até aqui não foi ser completamente exaustivo acerca da teoria de Russell e seus detalhes, mas dar alguma noção daquilo que é mais central da teoria e o que será relevante para os propósitos deste ensaio.

## Teoria das descrições de Strawson

Não se pode tratar de descrições definidas sem mencionar a participação de Strawson na discussão. De fato, as primeiras grandes objeções à teoria russelliana foram apresentadas em “On Referring” (Strawson, 1950). Nesse artigo Strawson apresenta um grande conjunto de objeções a proposta de Russell e fornece um tratamento alternativo às descrições definidas. Nesta seção apresento rapidamente o núcleo da proposta de Strawson sem determe calmamente em cada uma das objeções apresentados por ele.

### Um ponto de partida para compreender a diferença entre Strawson e Russell no que diz

6 - O problema aqui não é tão simples assim. Há pelo menos duas leituras razoáveis para a frase ‘a raiz de -2 não é um número inteiro’ com base na teoria de Russell; uma com a negação com âmbito longo e outra com o âmbito curto. Na leitura onde a negação tem âmbito longo a frase parece verdadeira. A forma lógica é a seguinte:  $\neg \exists x Fx [\forall y (Fy \rightarrow x = y) \wedge Gx]$ . Nesse caso a frase diz, sem pressupor que o objeto em causa exista, que não há um objeto tal que tenha tais e tais propriedades e, portanto, parece intuitivamente verdadeira. Mas há uma outra leitura na qual a negação tem âmbito curto e que parece falsa.

A forma lógica é a seguinte:  $\exists x Fx [\forall y (Fy \rightarrow x = y) \wedge \neg Gx]$  e diz alguma coisa como há um objeto tal e apenas um que tem a propriedade G. Esta leitura pressupõe a existência do objeto para falar a respeito dele. Russell rejeita a segunda leitura sem argumentar cuidadosamente nessa direção.

respeito ao tópic das descrições é compreender que cada um tem propósitos diferentes em mente. Enquanto Russell pensa em encontrar a *forma lógica geral* de descrições definidas, Strawson está mais preocupado com o uso de descrições definidas na linguagem natural. Dessa motivação geral surge duas noções quase antagônicas; uma ligada a significado proposicional, tanto quanto possível despreocupada com contextos de uso (Russell). E outra resultante de contextos de elocuições, preocupada com os usos que as pessoas fazem da linguagem para falar (Strawson). De um modo geral, Russell está mais preocupado com a semântica algo descolada das imprecisões típicas das linguagens naturais (ainda que sua teoria pretenda dar conta da linguagem natural) e Strawson quer aproximar a sua teoria da linguagem natural e dos usos que as pessoas fazer das frases e palavras para se expressarem. Com essas diferenças gerais em mente podemos passar aos aspectos de pormenores da teoria de Strawson.

Talvez a distinção mais importante de Strawson seja entre:

- a) Uma frase;
- b) Um uso de uma frase;
- c) E uma elocução de uma frase.

E com exatamente a mesma ideia:

- a) Uma expressão;
- b) Um uso de uma expressão;
- c) E uma elocução de uma expressão.

Segundo essas distinções podemos encarar a expressão ‘o melhor futebolista do mundo’ como uma expressão *tipo* da qual cada elocução da mesma será um *exemplar*, um *espécime* ou um *token*. Toda vez que a expressão ‘o melhor futebolista do mundo’ é usada em uma elocução temos uma nova espécime da expressão tipo. Assim, ‘o melhor futebolista do mundo é português’ e ‘o melhor futebolista do mundo é brasileiro’ são duas ocorrências da mesma expressão tipo, a dizer ‘o melhor futebolista do mundo’. Segundo Strawson, uma frase, tecnicamente, não tem significado. Significado é uma função do uso de uma frase por

uma pessoa em uma elocução. Estritamente, perguntar se a frase ‘o melhor futebolista do mundo é português’ é verdadeira, é precipitado, sem considerar o uso que está sendo feito da frase. O uso de uma frase é determinado por certas convenções explícitas ou não acerca do modo como usamos a linguagem. Assim, segundo uso atual, a frase ‘o melhor futebolista do mundo é português’ é verdadeira. Porém, antes de Cristiano Ronaldo ser eleito, um uso diferente da frase a tornava falsa, dado que àquela altura, segundo aquele uso, a expressão ‘o melhor futebolista do mundo’ se *referia* ao argentino Lionel Messi.

Para Strawson, um determinado uso de uma descrição definida *refere* a um determinado objeto (o objeto que é tal que satisfaz unicamente a descrição); e esta é uma diferença com relação a Russell que vale fazer notar. Para Russell a relação entre um objeto e uma descrição definida não é uma relação de referência, que *pressupõe* a existência do objeto referido e na ausência de tal a frase onde a descrição ocorre fica sem valor de verdade. Para Russell, a relação é mais fraca e compatível com inexistência do objeto, caso em que a frase onde ocorre a descrição será, sob certa leitura, falsa. Segundo Russell, a frase ‘o atual rei da França é careca’ é falsa pelo fato de que não há atual rei da França; atualmente a França não é uma monarquia. Strawson, por sua vez, diria que a frase, sob o uso normal que fazemos atualmente da descrição definida ‘o atual rei da França’, ‘o atual rei da França é careca’ não é verdadeira nem falsa, dado que a descrição definida relevante não possui referente (novamente, sob uso atual da descrição). Sob outro uso, no qual a descrição teria referente, a frase seria verdadeira ou falsa, i.e., teria valor de verdade.

Esses são os aspectos mais importantes a se destacar das diferenças entre as teorias de Russell e Strawson. Há muito mais a dizer principalmente acerca das objeções de Strawson sobre vários aspectos da teoria russelliana que não podem ser discutidos aqui, mas que devem ser mais detidamente estudados por quem pretende se aprofundar no estudo acerca dessa disputa.

## O Uso Referencial de Descrições Definidas

Nesta seção apresento algumas das contribuições mais relevantes de Keith Donnellan para o problema das descrições definidas com base em seu influente artigo (Reference and Definite Description, 1966). Donnellan defende a ideia que descrições definidas podem ser usadas de dois modos distintos – atributivamente e referencialmente – e que as teorias das descrições anteriores não são capazes de explicá-los. A defesa de Donnellan dessa distinção é baseada em inúmeros exemplos que, segundo ele, justificam a distinção. O objetivo desta seção é, além de explicar como funciona a distinção proposta por Donnellan, mostrar

como, ao que parece, tanto Strawson quanto Russell não dão conta de explicar os casos apresentados por ele.

A caracterização da distinção de Donnellan é feita pelo próprio autor do seguinte modo:

1. *Uso atributivo*: “Um falante que usa uma descrição definida atributivamente em uma asserção diz alguma coisa acerca seja de quem for ou do que for que é o tal-e-tal.” (1966, p. 233)<sup>7</sup>

2. *Uso referencial*: “Um falante que usa uma descrição definida referencialmente [...] usa a descrição para capacitar sua audiência a selecionar quem ou o que ele está falando a respeito e diz alguma coisa sobre aquela pessoa ou coisa.” (1966, p. 233)<sup>8</sup>

A ideia geral aqui é que no uso referencial a descrição pode ou não ser satisfeita pelo objeto ao qual a descrição se refere. No uso atributivo, por outro lado, a descrição é usada para dizer alguma coisa acerca de alguém ou alguma coisa que satisfaz uma descrição. Talvez o melhor modo de entender o que Donnellan tem em mente seja com recurso aos exemplos que ele próprio fornece.<sup>9</sup> A estratégia de Donnellan para explicar a distinção é apresentar um exemplo e mostrar casos onde a mesma descrição definida é usada referencialmente e atributivamente. A mesma estratégia será usada agora.

#### Exemplo 1: ‘O assassino de Smith é louco.’

Suponhamos dois cenários diferentes (A) e (B). Em (A) um amigo de Smith chega à sua casa e o vê caído ao chão brutalmente violentado e diz ‘o assassino de Smith é louco’ se baseando no modo como o corpo está e não tem uma pessoa particular em mente. Em um cenário (B), Jones está a ser julgado pelo assassinato de Smith e (pelo seu comportamento indiferente no tribunal) uma mulher que está a ver o julgamento diz ‘o assassino de Smith é louco’. Em (A), a descrição é sobre uma pessoa qualquer que tenha assassinado Smith e a descrição é acerca do objeto que a satisfizer. Em (B), sustenta Donnellan, o referente da

---

7 - Tradução livre de responsabilidade do autor deste artigo.

8 - Idem nota anterior.

9 - Os exemplos a seguir têm pequenas variações de estilo e adaptações para o português e não diferem no que é relevante aos exemplos apresentados originalmente pelo autor.

descrição é Jones, sendo ele ou não o assassino de Smith. Enquanto em (A) a descrição denota o objeto por via de determinada propriedade e apanha seja quem for que satisfaça a descrição, em (B) a descrição tem uma função genuinamente referencial, a mulher tem *uma pessoa* específica em mente (Jones) e usa a descrição para referi-la. Segundo Donnellan, em (A) a descrição definida é usada atributivamente e em (B) é usada referencialmente.

### Exemplo 2: ‘Quem é o homem bebendo Martini?’

Este segundo exemplo é um caso análogo ao anterior, mas que ocorre na forma de uma pergunta. Novamente suponhamos dois cenários distintos (A) e (B). Em (A) pense em duas mulheres em um bar conversando de onde se vê ao lado oposto do salão um homem. Uma das mulheres demonstrando interesse pergunta para a outra: ‘quem é o homem bebendo Martini?’. A intuição de Donnellan é que mesmo no caso no qual o homem em questão não está bebendo Martini, a pergunta pode ser respondida corretamente, ou seja, mesmo que a descrição ‘o homem bebendo Martini’ não seja satisfeita, a conversa das amigas funciona perfeitamente e a descrição cumpre seu papel na frase. Para Donnellan esse caso só pode ser explicado se o uso da descrição definida for entendido referencialmente. Agora, em um cenário (B), suponha que está a ocorrer uma confraternização anual dos Alcoólicos Anônimos e chega aos ouvidos do presidente do grupo que há alguém na festa que está bebendo Martini. Imediatamente o presidente chama o responsável pela segurança e pergunta ‘quem é o homem que está bebendo Martini?’. Nesse cenário, o presidente não tem uma pessoa particular em mente, ao contrário, ele quer saber alguma coisa acerca de quem quer que seja que for o homem que está bebendo Martini para eventualmente poder expulsá-lo da festa. Em (B), portanto, a descrição ‘o homem bebendo Martini’ é usada atributivamente para apanhar seja quem for que satisfaça a descrição.

### Exemplo 3: ‘O marido dela é gentil com ela.’

Este exemplo é particularmente interessante porque há várias coisas envolvidas, mas vamos tentar tratar do que é mais relevante para uso referencial e uso atributivo da descrição definida em causa. O caso aqui é que dependendo do modo como a frase é interpretada (referencialmente ou atributivamente) pode dizer coisas profundamente diferentes, na realidade pode dizer uma falsidade ou uma verdade. O cenário em causa é o seguinte:

alguém está a ver um casal se divertindo em um parque e pelo modo como o homem trata a mulher alguém diz 'o marido dela é gentil com ela.' Porém, o que ocorre na realidade é que o homem que está no parque é o amante e que o marido da mulher não é nada gentil com ela. Assim, temos duas interpretações possíveis: uma (A) na qual a frase é falsa (atributiva) e outra (B) na qual a frase é verdadeira (referencial). O interessante nesse exemplo é que a frase só é verdadeira na interpretação segundo a qual a descrição apanha o objeto que não satisfaz a descrição, ainda que a descrição seja satisfeita por um único objeto que não é o objeto referido. Assim, sob a interpretação (B), a frase diz alguma coisa verdadeira acerca de alguém que não satisfaz a descrição. E, sob a interpretação (A), diz alguma coisa falsa acerca de alguém que satisfaz unicamente a descrição.

Donnellan avança várias críticas tanto a Russell quanto a Strawson que poderiam ser exploradas em detalhe. Para as finalidades deste ensaio, uma visão geral das críticas de Donnellan será o suficiente.

No que diz respeito a Russell, Donnellan pensa que sua teoria pode ser aplicada corretamente a muitos dos casos de usos atributivos de descrições, mas falha no que diz respeito ao uso referencial. Russell pode dar conta de um dos usos, mas não de ambos. Segundo Donnellan, Russell não reconhece o uso referencial de descrições definidas e esta é uma deficiência incontornável para a sua teoria das descrições. Ainda que possa ser correta ao tratar dos usos atributivos de descrições, é incompleta, dado que sua teoria não pode explicar os usos referenciais.

Contra Strawson, Donnellan acredita que sua teoria acomoda o uso referencial de descrições, mas vai longe demais nessa direção. Strawson não permite que o uso referencial ocorra adequadamente quando a descrição não é satisfeita. E isso, pensa Donnellan, seus exemplos mostram que pode ocorrer em usos referenciais de descrições.

Donnellan argumenta ainda que Russell e Strawson compartilham uma suposição (ou pressuposição) que à luz de seus exemplos parece falsa. Ambos acreditam que quando nada se encaixa na descrição definida o valor de verdade da frase na qual ela ocorre é afetado. Isto é, se não há  $\alpha$ , o valor de verdade de um uso da frase 'o  $\alpha$  é  $\beta$ ' será afetado. Para Russell, a frase será sempre falsa. E para Strawson, a frase não terá valor de verdade. E, novamente à luz de seus exemplos, casos de usos referenciais de descrições mostram que esta suposição é falsa ou, ao menos, implausível.

Para Donnellan, tanto Russell quanto Strawson estão errados acerca do uso referencial. Um falante pode usar 'o  $\alpha$  é  $\beta$ ' para dizer alguma coisa verdadeira mesmo quando nada satisfaz a descrição 'o  $\alpha$ ' (i.e., não há  $\alpha$ ).

A ideia central que para Donnellan nem Russell nem Strawson conseguiu captar é que alguém pode referir alguma coisa usando uma descrição que não descreve corretamente a coisa a qual se está a referir. Podemos considerar uma pequena variação no exemplo 2, novamente no uso referencial de ‘o assassino de Smith é louco’. Se Jones não é culpado – na realidade, Smith cometeu suicídio –, então a descrição ‘o assassino de Smith’ se aplica a ninguém. Nem Jones nem outra pessoa qualquer se encaixa na descrição. Mesmo assim, afirma Donnellan, o falante pode usar a frase para dizer alguma coisa verdadeira: ele diz de Jones, o homem que ele refere pela descrição ‘o assassino de Smith’, que ele é louco. Se Jones for realmente louco, o falante diz alguma coisa verdadeira. E esse é exatamente o ponto que tanto Russell quanto Strawson é incapaz de explicar.

## Uma Abordagem Pré-semântica

Nesta seção apresento a parte mais positiva deste ensaio. Vou defender uma abordagem pré-semântica ao uso referencial de descrições definidas segunda a qual o uso referencial de uma descrição definida é algo semelhante a uma convenção linguística, ainda que ocorra implicitamente por via do contexto de elocução.

Antes de seguirmos para a estratégia que pretendo defender propriamente dita, faz-se necessário algumas distinções com a finalidade de estabelecer o vocabulário. Quero distinguir entre três momentos linguísticos: pré-semântico, semântico e pós-semântico.

- i) Pré-semântico: convenções linguísticas, etc.
- ii) Semântico: saturação, desambiguação, etc.
- iii) Pós-semântico: implicaturas, etc.

Com base em exemplos pode ficar mais fácil localizar, pelo menos intuitivamente, como funciona a distinção. Suponhamos por exemplo (claro que não precisa ser verdade) que durante algum tempo Steve Jobs e Steve Wozniak não possuíam um nome para a empresa que estavam criando, isso certamente seria um problema para eles até que um dia Jobs diz para Woz ‘Nossa empresa vai ser chamar “Apple”’. No exato momento em que esse evento ocorre, certa convenção linguística é feita. Uma convenção segundo a qual a palavra ‘Apple’ será usada para designar a empresa recém-criada por Woz e Jobs. Esse foi um evento *pré-semântico* envolvendo certo uso da palavra ‘Apple’. Anos após a criação da empresa (e de certa convenção

acerca de certo uso da palavra ‘Apple’) um tabloide de notícias estampa em sua primeira página ‘A Apple é a maior empresa de tecnologia do mundo’, fazendo referência à primeira vez que a empresa alcançou a posição de empresa mais valiosa do mundo. A palavra ‘Apple’ é ambígua em inglês e uma notícia semelhante escrita em inglês não levaria de modo algum o leitor a pensar: será que a notícia é sobre a empresa de tecnologia ou sobre aquela fruta que tanto gosto? Certamente, o contexto no qual a palavra ‘Apple’ é usada não deixa menor espaço para a ambiguidade no uso do termo. Essa desambiguação ocorre tipicamente no nível *semântico*. Em uma outra situação, duas amigas estão conversando acerca de seus *laptops* e uma delas está reclamando que seu computador está muito ruim e a fez perder um trabalho de escola que estava por terminar. Ao que a outra amiga responde: ‘eu tenho um Apple’. Nesse contexto, a amiga que tem um *laptop* ruim não teve problema algum em compreender que a sua amiga queria dizer que os computadores da Apple tipicamente não apresentam problemas. Mas nada do que foi compreendido apareceu na elocução da frase ‘eu tenho um Apple’. A compreensão de seu conteúdo ocorreu por via de uma *implicatura conversacional*. Esse é um caso típico de compreensão com recurso ao nível *pós-semântico*.

Com base nessas distinções podemos pensar acerca dos casos de descrições definidas referenciais e perguntar em que nível está a distinção. Descrições definidas são ambíguas? São como convenções? Implicaturas conversacionais? Parece que não há respostas óbvias aqui. Os três caminhos parecem plenamente viáveis. Donnellan por si próprio não é completamente claro acerca do problema, mas na seguinte passagem parece apontar para uma distinção pragmática.

“It does not seem possible to say categorically of a definite description in a particular sentence that it is a referring expression (of course, one could say this if he meant that it might be used to refer). In general, whether or not a definite description is used referentially or attributively is a function of the speaker’s intentions in a particular case. [...] Nor does it seem at all attractive to suppose an ambiguity in the meaning of the words; it does not appear to be semantically ambiguous. (Perhaps we could say that the sentence is pragmatically ambiguous[...].)” (Donnellan, 1966, p. 272)

Donnellan parece estar sugerindo uma distinção no nível pós-semântico. Mas, sendo esse o caso, é argumentável que uma teoria semelhante a teoria russelliana seja capaz de responder a todos os fatos semânticos relativos a descrições definidas e as críticas de Donnellan parecem não serem capazes de refutar Russell. A respeito da posição de

Donnellan, Kripkesugere:

“Donnellan’s denial that he advocates a semantic ambiguity in definite descriptions seems inconsistent with much of his paper, we can try ignoring the denial, and take his paper to be arguing for such an ambiguity. Then we may ask: has Donnellan established a (semantic) ambiguity inconsistent with Russell’s theory?” (Kripke, 1977, p. 262)

Kripke (1977) tenta argumentar a favor de uma resposta negativa para a pergunta. Para tanto, tenta mostrar que o uso referencial de descrições pode ser acomodado em uma teoria do tipo russelliana defendendo a ideia que a distinção é pragmática (pós-semântica, na terminologia usada deste ensaio).

Influenciado por Grice (1975), Kripke lança mão da distinção entre *referente do falante* e *referente semântico* (ou *referente da descrição*). Para clarificar a distinção, Kripke sugere dois tipos de intenções: *intenção geral* e *intenção específica*. A intenção geral está ligada ao referente semântico da descrição independentemente de quando a descrição é usada, baseado em certas convenções linguísticas de certos dialetos. A intenção específica aparece, em certa ocasião, para referir um certo objeto. A intenção específica está ligada ao referente do falante.

Se o falante acredita que o objeto acerca do qual quer falar em certa ocasião satisfaz as condições para ser o referente semântico, então sua intenção específica coincide com a sua intenção geral. Para Kripke, o caso onde há coincidência entre a intenção específica e a intenção geral é o uso atributivo de Donnellan. E quando não há coincidência, é o uso referencial.

Para Kripke a teoria de Russell funciona corretamente para explicar o referente semântico e isso é tudo que há de semanticamente relevante para ser explicado, ainda que possa haver uso de descrições onde o referente do falante é que vale (casos nos quais a intenção específica e a geral não coincidem). Nesses casos, uma teoria pragmática (pós-semântica) irá explicar o fenômeno com base nas intenções do falante.

Ao que parece, os últimos parágrafos acerca das ideias de Kripke não são inconsistentes com a citação de Donnellan. Afinal, ele reconhece o papel das intenções do falante e até sugere que a distinção seja pragmática. Porém, se esse for caso, parece razoável pensar que não há ambiguidade (no nível semântico) e que, portanto, a teoria de Russell explica corretamente o *sentido*, a *semântica*, das descrições definidas. Ainda que descrições definidas possam ser *usadas* (nível pós-semântico) para dizer coisas diferentes daquilo que

é semanticamente o típico.

Há outra possibilidade de tratar da distinção entre uso referencial e uso atributivo de descrições definidas. É tentar mostrar que o uso referencial de descrições definidas não é um evento semântico nem um evento pós-semântico, mas sim um evento pré-semântico; está no nível das convenções linguísticas. Há disponível na literatura teorias com essa motivação como, por exemplo, os trabalhos de Michael Devitt (*Designation*, 1981) e Gareth Evans (*The Varieties of Reference*, 1982), entre outros.

## A teoria causal da referência

Usar uma descrição definida referencialmente é algo semelhante a inaugurar o uso de certa expressão com certo significado particular que até então não havia sido usado ou como usar certo nome que ainda não havia sido usado para referir certo objeto. A imagem da cerimônia de batismo funciona aqui também. Quando usamos uma descrição definida referencialmente em certo contexto, estamos a fazer certa convenção (mesmo que não explícita) que aquela descrição irá passar a referir determinado objeto em situações posteriores. Um fenômeno muito semelhante ao que acontece quando batizamos uma pessoa ou coisa com certo nome. Ou mesmo quando usamos uma certa palavra com certo significado que a palavra não possuía antes (uma boa imagem aqui é o uso de gírias). Os falantes passam com frequência a usar palavras com significados novos que até então nunca tinham sido atribuídos a determinada expressão.

O exemplo de Devitt é útil para explicar a chamada ‘teoria causal da referência’. Devitt pede para considerarmos o caso do seu gato. E afirma que sua esposa disse “Vamos chamá-la ‘Nana.’” E ele concordou. Desse modo, Nana foi nomeada. Para esse cenário Devitt fornece a seguinte explicação:

What happened to those present at naming Nana? They *perceived* the ceremony, using at least their eyes and ears. To perceive something is to be causally affected by it. As a result of the effect it had on them, they were in a position to use the name ‘Nana’ later to designate the cat. What they gained at the ceremony was “an ability to designate Nana by ‘Nana.’” (1981, p. 23)

A habilidade de designar *Nana* por '*Nana*' pode ser passada de uma pessoa para outra por via de uma cadeia causal que irá explicar os usos posteriores do nome, mesmo por quem nunca tenha estado em contato com o gato. Este é um caso que envolve um nome e não uma descrição definida, mas para Devitt (e estou de acordo) imagens semelhantes podem ser desenvolvidas para explicar as descrições definidas usadas referencialmente.

Quando usamos a descrição 'o assassino de Smith' em contexto apropriado podemos estar de certo modo a estipular que a descrição passa a referir a um determinado objeto, ainda que esse objeto não satisfaça a descrição. Na verdade, para o fenômeno encarado de modo mais geral, satisfazer ou não a descrição é algo completamente irrelevante.

Podemos estipular certas descrições inclusive em casos onde a descrição é intencionalmente não satisfeita. Podemos estipular, que a descrição 'a menina mais bonita da sala' se refira a menina que é na verdade a menina mais feia. E quando usamos a descrição ela se refere de fato a menina mais feia e não a menina mais bonita, que seria o caso do uso atributivo.

Eu concordo com a ideia geral do projeto do Devitt que talvez o uso referencial de descrição deva ser encarado como um evento pré-semântico, mas penso que a sua teoria não está completamente correta. Primeiro, porque penso que há casos que a sua proposta não é capaz de responder. E depois penso que sua teoria não explica corretamente o que ocorre no momento do batismo inicial. Penso que Devitt, por motivações filosóficas mais gerais, quer explicar o fenômeno em termos que são incapazes de fazê-lo. No restante deste ensaio, passo a apresentar os argumentos contra a teoria de Devitt e depois apresento a teoria que proponho que pode explicar o fenômeno do uso referencial das descrições definidas de modo mais completo e com maior parcimônia.

## **Contraexemplo à teoria causal**

Há um conjunto de exemplos que são difíceis de explicar para qualquer teoria causal da referência que funcione com base em contato perceptual com o objeto. Um desses exemplos que é recorrente na literatura é o seguinte: imagine que a certa altura o rei de Portugal teve seu trono tomado por um usurpador e o rei foi mandado para as masmorras. O usurpador resolve dar uma festa para a qual os nobres do reino foram convidados. Um desses convidados ao chegar a festa em conversa com o guarda diz "o rei me convidou". A descrição 'o rei' não está funcionando atributivamente (o rei está nas masmorras preso e não

tem condições de convidar ninguém) e não funciona segundo a teoria do Devitt (podemos imaginar que o convidado nunca tenha estado em contato perceptual com o usurpador para poder usar a descrição ‘o rei’ para referi-lo), mas esse parece um uso referencial adequado de descrições definidas.

Certamente esse exemplo não é definitivo e podem haver respostas viáveis para a versão conforme apresentada, mas parece possível ir modificando ligeiramente o exemplo de modo a sempre gerar problemas para a teoria causal da referência como imaginada por Devitt. O objetivo aqui não é apresentar um contraexemplo definitivo, mas mostrar a dificuldade da teoria causal com exemplos relativamente recorrentes na linguagem usual. Com algum trabalho imaginativo podemos reproduzi-los a infinidade.

## O problema da descrição do primeiro uso

Uma outra objeção possível diz respeito a descrição do primeiro uso. O problema aqui é saber que elementos devem estar presentes no primeiro uso de um nome ou uma descrição definida usada referencialmente. Um dos elementos fundamentais da proposta de Devitt é contato perceptual com o objeto que será submetido à cerimônia de batismo. A ideia é que há uma certa cerimônia que dá conta do caso “ela vai chamar ‘Nana’” devido ao contato perceptual que os agentes que participam da cerimônia têm com o objeto, nesse caso *Nana*. Através dessa cerimônia e do efeito causado pelo objeto, o gato passa a se chamar ‘Nana’. Mas para quem? Para a pessoa que o batizou? Uma pessoa é o bastante? Se duas, como o batismo é da mesma coisa? Parece razoável pensar em uma situação dessas onde algumas das pessoas presentes ‘batizam’ coisas diferentes. A percepção é muito complexa e parece razoável pensar em situações nas quais todos estão em contato perceptual com o objeto, mas ‘batizam’ coisas diferentes (estão em contato com mais de uma coisa). Essa dificuldade parece ainda mais evidente em casos de descrições definidas usadas referencialmente. Tipicamente não haverá um evento explícito e será ainda mais fácil haver confusão com relação ao evento de batismo. Ao que parece uma descrição dos eventos de batismo (das convenções) que elimina estados mentais intencionais dos agentes envolvidos, não pode responder completamente às perguntas acima.

Uma resposta viável é tentar mostrar que no contexto as pessoas envolvidas sabem exatamente qual objeto está sendo batizado. Mas o que exatamente *ésabera* aqui? Reconhecer a *intenção de referir*? Se for isso, qual o papel *fundamental* que a relação causal perceptual desempenha? Parece que o papel da percepção é um papel secundário. Primeiro vem a

*intenção de referir* e o *reconhecimento* desta intenção, depois vem a relação perceptual (se é que esta é necessária). Se relação perceptual não é fundamental, ao contrário das crenças e intenções dos agentes envolvidos, talvez possa ser dispensada no curso da explicação.

Mesmo como a relação causal perceptual, o nome pode ter ‘batizado’ coisas diferentes. Ainda permanece o mistério da referência. No meu ponto de vista, o batismo, ainda que relacionado com a percepção, precisa de *intenções de referir* e *reconhecimento* das mesmas. Não há nada na percepção que possa separar o gato de tudo o mais que está a ser percebido para que o batismo seja possível pela percepção apenas. A intenção de referir é fundamental.

A ideia geral do meu ponto de vista talvez já tenha sido alcançada pelo leitor. Mas pode ainda faltar detalhes de como essa via teórica pode funcionar. Como podemos fazer essas convenções? Como as convenções funcionam? O que é necessário para que as convenções funcionem? Todos os detalhes serão apresentados a seguir. Não há nada além daquilo que foi apresentado até aqui, tudo o que é feito a seguir é uma sistematização teórica das ideias apresentadas anteriormente.

## A solução

A solução que proponho é análoga à proposta de Devitt, porém não pressupõe a necessidade da relação percepção entre o objeto referido e o agente que usa determinada expressão para referi-lo. Na versão que proponho também haverá dois tipos de referência: um primitivo, ligado ao uso da convenção ou batismo, e outro derivado, ligado a certo uso posterior de uma expressão com a referência previamente convencionada. Esses dois momentos explicam o uso de certa expressão com certa referência pela primeira vez (em uma convenção) e a posterior distribuição de certa convenção em uma comunidade linguística.

*Núcleo de referência* é o conjunto de pessoas que compartilham a convenção de usar a expressão referencial  $j$  para referir  $x$ . Para todo  $j$  há pelo menos um *núcleo de referência* formado pelas pessoas que partilham a mesma convenção.<sup>10</sup>

---

Essa formulação permite, ao mesmo tempo, que hajam expressões que tenham núcleos

**10** - Note que se não há um núcleo de referência associado a uma expressão referencial  $\varphi$ , então, por definição,  $\varphi$  não é uma expressão referencial.

de referência com apenas dois membros (o locutor e o interlocutor) e a mesma expressão tenha diferentes núcleos de referência. Sendo que núcleo de referência é, por definição, uma função que vai de uma convenção específica para uma expressão a um conjunto de pessoas.

*Princípio geral:* Uma expressão referencial usada por um locutor refere se, e somente se, há um núcleo de referência vinculado à expressão.

Há duas maneiras de uma expressão referencial possuir núcleo de referência; primitivamente (NRP) (inauguração de convenção) derivadamente (NRD) (intenção de usar uma convenção anterior).

(NRP) Uma expressão referencial *j* usada pelo locutor *L* com intenção de referir *x* refere se, e somente se, *I* (um interlocutor) consegue reconhecer que *L* usa *j* com a intenção de referir *x*.<sup>11</sup>

Esse é o princípio que governa o uso primitivo de uma expressão referencial.

*Note:* o papel do objeto é subsidiário. A intenção de referir eo reconhecimento da intenção é que possuem papéis fundamentais. A relação perceptual com o objeto pode desempenhar um importante papel para o reconhecimento da intenção de referir, mas esse é um papel secundário.

Uma expressão usada por um locutor com uma intenção de referir alguma coisa, tem referência primitiva se, e somente se, há um núcleo de referência formado pelo locutor e pelo menos um interlocutor e este interlocutor percebe a intenção do locutor de referir essa coisa com essa expressão.

(NRD) Uma expressão referencial *j* usada pelo locutor *L* com intenção de referir *x* refere derivadamente se, e somente se, há um núcleo de referência derivado causalmente de um núcleo de referência primitivo e a convenção desse núcleo é compatível com a intenção de *L* de referir *x*.

Em termos mais diretos, uma expressão referencial usada por um locutor com a intenção

---

*11* - Uma expressão pode ter inúmeras convenções associadas. Porém, quando usada é usada com uma intenção específica que, se reconhecida pelo interlocutor, torna a referência possível.

de referir alguma coisa tem referência derivada se, e somente se, há um núcleo de referência derivado causalmente de uma convenção de referência primitiva e a referência desse núcleo é compatível com a intenção do locutor ao usar a expressão.

É fácil ver como a solução que proponho tem maior facilidade em tratar dos problemas apresentados à proposta de Devitt. O caso da frase ‘o rei me convidou’, por exemplo, onde não há relação perceptual com o objeto referido, não levanta problema especiais para a proposta apresentada. Há uma intenção de referir que é, no contexto adequado, corretamente reconhecida pelo interlocutor de modo que a descrição ‘o rei’ se refere ao usurpador e não ao rei.

O outro problema levantado diz respeito às condições necessárias para o batismo (ou convenção). A proposta de Devitt talvez tenha dificuldades em explicar como a relação perceptual é suficiente para isolar corretamente o objeto que está a ser nomeado de tudo o mais no campo perceptual e gerar acordo entre aqueles que participam do evento de nomeação. Pela proposta que acabo de apresentar, esse problema é mais facilmente resolvido, dado que a intenção de referir faz o trabalho de isolar corretamente o objeto de referência no momento do batismo ou convenção.

Supondo que leitor aceite o que foi apresentado até aqui, o que dizer acerca da distinção de Donnellan? As descrições definidas são ambíguas ou não? Há uso genuinamente referencial de descrições definidas? Em um certo sentido quase trivialmente segue-se que sim. Mas há um sentido no qual as descrições definidas usadas referencialmente não são genuinamente descrições, são mais semelhantes a nomes próprios. Aqui então há duas vias. Na primeira, o leitor que aceitou uma solução semelhante a que proponho, pode defender que há uma convenção sociolinguística que diz que descrições podem ser usadas eventualmente de modo referencial. Algo semelhante à convenção que o nome ‘João’ é ambíguo e pode se referir a muitas pessoas e quem vai decidir é o contexto. A outra via é não aceitar que haja tal convenção e ter uma posição que o uso referencial de descrições não tem nenhuma relação semântica especial com as próprias descrições. Nesse sentido, uso referencial de descrições definidas estão muito mais ligadas ao contexto e as intenções do falante que a alguma noção semântica envolvendo descrições; de certa forma a descrição é um mero dispositivo que desempenha certa função em certo contexto linguístico que poderia ser desempenhado de modo semelhante por outra coisa qualquer – linguística ou mesmo extra-linguística. Eu, particularmente, estou mais inclinado a esta última noção. Mas acredito que a intuição aqui está ligada a intuições filosóficas mais gerais acerca da linguagem; principalmente acerca do papel semântico do contexto – contextualistas tenderia a segunda via e um minimalista a primeira.

## Bibliografia

- Devitt, M. (1981). *Designation*. Columbia University Press.
- Donnellan, K. (1966). Reference and Definite Description. *Philosophical Review*, pp. 281-304.
- Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.
- Frege, G. (1892). Sobre o Sentido e a Referência. *Fundamento - Revista de Pesquisa em Filosofia*, pp. 21-44 Trad. Sérgio Miranda.
- Grice, P. (1975). Logic and Conversation. In G. Morgan, & P. Cole, *Syntax and Semantics 3: Speech Acts* (pp. 41-58). New York: Academic Press.
- Kaplan, D. (1977). Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and other Indexicals.
- Kripke, S. (1977). Speaker's Reference and Semantic Reference. *Midwest Studies in Philosophy*, pp. 255-276.
- Kripke, S. (1980). *Naming and Necessity*. Oxford: Blackwell.
- Russell, B. (1905). On Denoting. *Mind*, 14, pp. 479-493.
- Strawson, P. (1950). On Referring. *Mind*, pp. 320-344.



# Vagueness: Tolerance and Incoherence

## *Vagueza: Tolerância e Incoerência*

Sagid Salles  
UFRJ-PPGLM

### **Abstract**

*In this paper I will argue that to accept the principle of tolerance does not provide us with a good explanation of the phenomena of vagueness. I will be mainly concerned with the incoherentist strategy, which accepts tolerance and the consequent incoherence of vague predicates. In fact, incoherentism seems to be the most plausible way of accepting tolerance. Because of this, the rejection of incoherentism might be seen as a way to rescue the alternative theories from the objection that they do not respect the principle of tolerance as a literal truth about vagueness. Despite this, I will conclude that neither incoherentism nor the main alternatives satisfy the criteria for an ideal theory of vagueness.*

### **Key-words**

*Vagueness. Principle of tolerance. Incoherentism.*

## Resumo

Neste artigo defendo que a aceitação do princípio de tolerância não resulta numa explicação satisfatória do fenômeno da vagueza. Meu interesse principal é com a estratégia incoerentista, que aceita a tolerância e a consequente incoerência dos predicados vagos. De fato, o incoerentismo parece ser a forma mais plausível de aceitar o princípio de tolerância. Devido a isto, sua rejeição pode ser vista como um modo de resgatar as teorias alternativas da objeção de que não respeitam o princípio de tolerância como uma verdade literal sobre a vagueza. A despeito disso, concluirei que nem o incoerentismo nem as principais alternativas satisfazem os critérios para uma teoria ideal da vagueza.

## Palavras-chave

Vagueza. Princípio da tolerância. Incoerentismo.

## Introduction<sup>1</sup>

It is sometimes supposed that vague predicates are tolerant to very slight changes. Furthermore to be tolerant is at least a necessary condition for a predicate to be vague. The point is that very slight changes cannot make any difference to the application of vague predicates. For instance, if someone is bald, then the mere addition of one hair cannot make him non-bald; if something is a heap of salt, one grain less cannot make it a non-heap of salt; and the examples can be easily multiplied. The *principle of tolerance* says that very small changes cannot make any difference to the application of these predicates.

Suppose a sequence like  $B_0, B_1, B_2, \dots, B_{10^0}$ , where “ $B_0$ ” should be interpreted as “a person with 0 hairs on her head is bald”, “ $B_1$ ” as “a person with 1 hair on her head is bald”, and so on. A formulation of the principle of tolerance for “bald” says that for every number  $n$ , if a person with  $n$  hairs on her head is bald, then a person with  $n+1$  hairs on her head is

---

*I* - I would like to thank Guido Imaguire and Juliana Faccio Lima for helpful suggestions and criticisms on earlier drafts of this paper. Thanks also to Iago Bozza Francisco and Otávio Bueno for the opportunity for several discussions about vagueness. Finally, I would like to acknowledge the support provided by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior).

also bald:  $\forall n (B_n \rightarrow B_{n+1})$ .

This leads us to the well-known sorites paradox, which can be formulated as follows:

(1)  $B_0$

(2)  $\forall n (B_n \rightarrow B_{n+1})$

---

(3)  $B_{10^{10}}$

The argument would take us to the conclusion that someone with  $10^{10}$  hairs on her head is bald. In other words, it shows that someone who is clearly not bald is bald. The same kind of argument could be used to show that a person with 0 hairs on her head is not bald. In this case, the argument would show that someone who is clearly bald is not bald.

Although there are different ways of avoiding the paradox, to reject the principle of tolerance is, at first sight, the more plausible one. After all, (1) seems to be obvious and the argument does not depend on any controversial rules of inference.

However, even this strategy is initially problematic. Consider the principle of tolerance as applied to “bald”:  $\forall n(B_n \rightarrow B_{n+1})$ . To refuse this amount to accept that at least one instance of the universally quantified sentence is not true. Let us suppose this instance is  $B_{9.999} \rightarrow B_{10.000}$ . Given that this instance is not true, the antecedent is true and the consequent is not true. In other words,  $B_{9.999}$  is true and  $B_{10.000}$  is not true. If this is so, then there is a sharp division between bald and not bald people. There will be a last case of bald in the sorites sequence, and all the others are not bald. Finally, if “bald” has a sharp boundary, then “bald” is a precise predicate and not a vague one. This is what Horgan (1994, p.162) called the *problem of precisification*, and it is serious a challenge to any theory of vagueness that rejects tolerance.

There have been many theories – such as many-valued theories, supervaluational theories, degrees of truth theories, contextualism, etc – trying to solve the sorites by rejecting the principle of tolerance. However, in so doing, it is suggested that they make vague predicates non-vague in the above sense. The problem here is that these theories accept in some sense that there is a number  $n$  such that  $B_n$  is true and  $B_{n+1}$  is not. Degrees of truth theories, for example, will accept the existence of a number  $n$  such that  $B_n$  is true in degree 1 and  $B_{n+1}$  is not true in degree 1 (is true in degree less than 1). The supervaluationist, in her turn, will

accept that there is a number  $n$  such that  $B_n$  is true for all precisifications and  $B_{n+1}$  is not true for all precisifications. Contextualists will also accept the existence of a boundary, and will explain vagueness – or at least the apparent tolerance of vague words – by appealing to some *similarity constraint* which says, roughly, that the boundary of vague words are never where we are looking for. In short, all these theories seem to postulate the existence of “a fact of the matter about the truth-value transitions in sorites sequences” (Horgan, 1994, p.163).

A possible reply on behalf of those theories is to say that tolerance should be seen as a kind of metaphor, not as a literal truth about vague predicates. In fact, it seems to me that in all these theories there is a sense in which we could say that vague predicates do not have boundaries. Many-valued and supervaluationist theories explain the inexistence of a boundary by appealing to the existence of indeterminate cases; degrees of truth theories do postulate a last case of true in degree 1, but they do not postulate a boundary between a last case of true and a first case of not true in the ordinary way; contextualists explain the metaphor of the inexistence of a boundary by appealing to a similarity constraint; and so on. After all, the principle of tolerance, together with other very plausible suppositions, implies that vague words are incoherent. It should be no surprise to accept this principle only as a kind of metaphor. The problem, then, would be whether some of these theories provide an acceptable explanation for the metaphor, and which one provides the better explanation.

In this paper I will argue that to accept the principle of tolerance does not provide us with a good explanation of the phenomena of vagueness. I will be mainly concerned with the incoherentist strategy, which accepts tolerance and the consequent incoherence of vague predicates<sup>2</sup>. In fact, incoherentism seems to be the most plausible way of accepting tolerance. Because of this, the rejection of incoherentism might be seen as a way to rescue the above theories from the objection that they do not respect the principle of tolerance as a literal truth about vagueness. Despite this, I will conclude that neither incoherentism nor the aforementioned theories satisfy the criteria for an ideal theory of vagueness.

---

2 - It is not totally clear what it would be to say that a predicate is incoherent. I will discuss at least two senses in which predicates might be incoherent: (i) they might involve inconsistent rules of use, and (ii) their mastering might involve dispositions to accept inconsistent rules of use.

# 1 Tolerance and Incoherence

## 1.1 The non-restriction consequence

If the principle of tolerance is true for “bald”, then it is true that  $\forall n (B_n \rightarrow B_{n+1})$ . So, for any item  $O$  in the sorites sequence, if you call it “bald”, you have to call every item with more hairs than  $O$  “bald” too. By contraposition, it is also true that  $\forall n (\neg B_{n+1} \rightarrow \neg B_n)$ . Thus, for any item  $O$  in the sorites sequence, if you call it “non-bald”, you have to call every other with less hair on the head “non-bald” too.<sup>3</sup> Let us express this by saying that the principle of tolerance, in this case, is a rule for the use of “bald”; and the same for the other vague predicates.

Whether or not the principle of tolerance is true, it is obviously true that  $\forall n (B_n \rightarrow B_{n-1})$ . That is to say, for any item  $O$  in the sorites sequence, if you call it “bald”, you have to call every other item with less hair on the head than  $O$  “bald” too. By contraposition, it is also true that  $\forall n (\neg B_{n-1} \rightarrow \neg B_n)$ . That is to say, for any item  $O$  in the sorites sequence, if you call it “non-bald”, then you have to call every item with more hairs on the head than  $O$  “non-bald” too. I will not be long about this, given that I think it is obvious that this is a minimal requirement on our uses of “bald”. The same kind of constraint applies to “tall”, “big”, “very much”, “young”, and many other vague words.

It should be noted that tolerance, plus the minimal requirement aforementioned, implies the unrestricted application of vague predicates. Suppose a sequence of persons (or, if you like, a temporal sequence of the same person) where the first is 0-haired, the second is 1-haired, the third is 2-haired, and so on. The result is this: if “bald” applies to one, then “bald” applies to all other elements in the sequence; and if “non-bald” applies to one, then “non-bald” applies to all other elements in the sequence. If someone applies “bald” to a person with  $n$  hairs on the head, she *must* apply “bald” unrestrictedly, and the same goes for “non-bald”. This is what we should expect if tolerance was true for “bald”. Let me call this the *non-restriction consequence* of principle of tolerance.

---

3 - This formulation seems to be committed to the thesis that meaning is normative. Throughout this paper, I will say that a speaker *have to* (or *should*) apply the predicate so and so. It seems, then, that I am committed to the thesis that the rules or norms which are constitutive of the meaning of vague predicates (i) classify the uses as correct and incorrect, and (ii) determine that speakers should use them correctly. Anti-normativists would accept (i) only. It seems to me that Wright presupposes some kind of normativism when he says that speakers are *forced* to apply vague predicates so and so (1975, p. 329, 333). Since I am concerned with Wright’s view, this presupposition is not a problem. I do not think, however, that my arguments depend on this presupposition. I could rather say that a speaker *should* apply the vague predicate so and so *if he wants* (*intends, etc.*) *to preserve its meaning* (*to communicate, etc.*). In this case, I would not be committed to (ii) because the *ought* (to use the predicate so and so) would depend on non-semantic or non-meaning-constitutive facts (for instance, intentions to preserve meanings or intentions to communicate). About the normativity of meaning, see De Souza Filho (2013).

## 1.2 A rule that no one follows

It is a clear fact about the ordinary uses of “bald” that speakers do not apply “bald” in such unrestricted way. They apply “bald” only to some objects and “non-bald” only to some others (it is irrelevant here if there are borderline cases). So ordinary speakers are not literally following the principle of tolerance, they are not applying “bald” in strict accordance with tolerance. The same will be true with respect to most (if not all) of the sorites predicates. Those who accept that the principle of tolerance is a rule for the uses of vague predicates will have to explain why ordinary speakers do not apply vague predicates in such unrestricted way.

Wright (1975, 1976) thinks that there is a behavioristic account of how people manage to apply “bald” only to some objects and “non-bald” also only to some objects, and how it is possible a large degree of agreement between them about who is or is not bald. Horgan (1994), in turn, thinks that this account is pragmatic. Ordinary speakers call some objects “bald” and some others “non-bald”, but they do not call all of them both “bald” and “non-bald” (as tolerance would imply in this case). He then goes on to argue that there must be pragmatic ways of insulating the incoherence of “bald” in such a way that it does not make too much effect on the ordinary uses of “bald”.

I will call this part of the explanation – that part which is supposed to explain how speakers manage to apply vague predicates in the way they *really* do – the *Other Half Theory*. It does not matter whether this theory is pragmatic, behavioristic or causal (Burns: 1991, p.95). The point is that those who accept the principle of tolerance will need such an additional theory.

The principle of tolerance, plus the minimal requirement aforementioned, implies that if you apply a vague predicate to one item in the sorites sequence, you *have to* apply it to any other item too. The only way we have to restrict our application of vague predicates is violating the principle of tolerance. The Other Half, whatever it is, must predict that people do not apply vague predicates in such unrestricted way and, hence, it must predict that people systematically violate the principle of tolerance. We first accept that tolerance is a rule for vague predicates and, then, we have to explain that it is a rule which is systematically violated by ordinary speakers.

In fact we can safely suppose that if tolerance is a rule for the uses of vague predicates, it is one that no competent user strictly follows. Imagine a speaker who applies “bald” (or “tall”, “heap”, “big”, etc.) in such unrestricted way. Would you call her a competent user of “bald” (“tall”, “heap”, “big”, etc.)? It seems that the best way out here is to just deny that

tolerance is a rule for vague predicates. In this case we still have to explain why tolerance seems to be true at first sight, but this is a different problem (the psychological problem of vagueness), and there has been several attempts to do that.

### 1.3 Wright's Incoherentism

Is there a way to conciliate the thesis that the principle of tolerance is a rule for vague predicates with the thesis that it is a rule which is systematically violated (so systematically that no competent speaker applies vague predicate strictly in accordance with this rule)? The advocates of tolerance need to address this question. One solution is to appeal to the existence of inconsistent rules. On the one hand, tolerance is a rule for vague predicates. On the other, there are rules which are inconsistent with tolerance. Ordinary speakers must choose when they will act in accordance with one or other. Crispin Wright (1975, p.334) has accepted this perspective.<sup>4</sup>

Wright's thesis: Whereas large enough differences in *F*'s parameter of application sometimes matter to the justice with which it is applied, some small enough difference never thus matters.

On the one hand, he accepts that tolerance is true and, hence, that a speaker is forced to apply vague predicates in an unrestricted way. On the other hand, he accepts a rule which is inconsistent with tolerance. The point is quite simple: while very slight differences force you to apply vague predicates unrestrictedly (tolerance), substantial differences do justify the application of vague predicates in a restricted way (the rule which is inconsistent with tolerance).

Let *a* and *b* be two items in the sorites sequence for "bald" which are such that (i) *a* is less haired than *b* and (ii) *a* and *b* are *substantially different* in the relevant aspect (that is, the difference of number of hairs between them is large enough to justify your application of "bald" to one and "non-bald" to the other). Suppose you apply "bald" to *a*. Given that "bald" is tolerant, you should apply "bald" to *b* too. However, given that *b* is substantially different from *a*, you are allowed to apply "non-bald" to *b* (even though you have applied "bald" to *a*). The same works for the case in which you begin by applying "non-bald" to *b*. In this case, tolerance will force you to apply "non-bald" to *a* too; but the other rules will allow you to

<sup>4</sup> - I use Weatherson's formulation for both Wright and Eklund's thesis (2009, p.85).

apply “bald” to *a* (even though you have applied “non-bald” to *b*). In sum, tolerance implies that you are not allowed to apply “bald” to *a* and “non-bald” to *b*. The rule for substantial differences, in turn, implies that you are allowed to apply “bald” to *a* and “non-bald” to *b*. The result is that you are and aren’t allowed to apply “bald” to *a* and “non-bald” to *b*. This makes sense of how tolerance may be a rule for the use of vague predicates even though it is so systematically violated. It is true that tolerance is systematically violated, but the reason is that there is an operative rule which is inconsistent with tolerance. Would this kind of incoherentism be plausible?

Here are some initial problems. This version of incoherentism is saying that both propositions bellow are true:

- (1) It is not allowed to apply “bald” to *a* and “non-bald” to *b*.
- (2) It is allowed to apply “bald” to *a* and “non-bald” to *b*.

To begin with, the incoherentist owes us an explanation of what makes it true that (1) and (2). Nonetheless the answer could not be that it is the fact that people use “bald” in strict accordance with (1) and (2). Given that these rules are inconsistent, it cannot be the case that people use “bald” in strict accordance with them (in each context, either you follow one or you follow the other). What else, then, makes both (1) and (2) true? Additionally, could we say that these inconsistent rules are both true concerning a linguistic practice such as, for example, the practice of using “bald”? It is just not clear what it would be the correct answer here. It might be, of course, that we have a set of rules containing inconsistent instructions such as (1) and (2). But the linguistic game of “bald” is not merely a set of rules, it is a set of rules which gives us instructions about how to use “bald”. As we will see bellow, inconsistent rules do not provide us with such instructions.

If a game is such that its instructions allow you to do *A* and do not allow you to do *A*, then there will be no effective instructions with respect to *A*. In the game of chess, you are allowed to move the horse in *L*. Suppose now that we add a rule saying that you are not allowed to move the horse in *L*. In this case, would you be allowed to move the horse in *L*? The answer is *yes and no*. As an instruction, this answer is not better than no answer at all. So it would be as if you had no instruction about whether or not you are allowed to move the horse in *L*. Now, think about the case of vague predicates. Are you allowed to apply vague predicates in a restricted way? The answer is *yes and no*. Once again, as an instruction, this

is not better than no answer at all. It appears that inconsistent instructions are not really informative about which move to do. They have no real role in the explanation of our uses of vague predicates; at least, no more than their simple inexistence would have.

Imagine a Martian who has no previous knowledge of how to use the word “bald” but she has learned all of its rules. She is not acquainted with our ordinary uses. In this situation, the Martian would still not be able to use “bald” as us. In other words, he would not know how to use “bald”. Wright seems to accept this result:

But while it may be true that the authority of the rules can be cited for any of the moves the community actually makes, it is plain that the rules alone do not provide a satisfactory account of the practice of the game. For someone could master the rules yet still not be able to join in the game, because he was unable to guess what sort of eclectic application of them an opponent was likely to make in relation to any given move.<sup>5</sup>

The rules of the game do not provide an account of how the game is played, for it is possible that someone might grasp them yet be unable to participate.<sup>6</sup>

Although Wright is disposed to accept this, it is not a plausible result. The reason why the result is implausible is that it implies that a person who explicitly acknowledges all the rules for the use of “bald” may not be able to apply this predicate to any singular case at all.<sup>7</sup> As far as the rules tell her, she will not be able to decide whether or not to apply “bald” to anyone (if you like, you can suppose that the conditions are the ideal ones). The instructions for the use of “bald” would teach nothing to her. In knowing the rules for “bald” the Martian would have no informative instruction to make even a single move in the linguistic game for “bald”. This is too radical.

On the one hand, Wright’s incoherentism does not make clear what it would be for both (1) and (2) being true, or even whether they could be both seen as true rules concerning

---

5 - Wright (1975, p. 362)

6 - Wright (1975, p. 362)

7 - Does Wright’s theory really have this implication? It is true that someone who knows (1) and (2) may still not know how to use “bald”. But there might be other rules for the use of “bald” and those might be informative. Someone who knows all the rules for the use of “bald”, for example, will know that this predicate is about amount of hair on the head and not about colors or horses; she would also know that if she applies “bald” to a person with  $n$  hairs on the head, then she has to apply it to a person with  $n-1$  hairs on the head too; and so on. All this is correct, but this person would still not be able to decide, based on the rules for the use of “bald”, whether or not anyone is bald. In other words, as far as the rules can tell her, she would not be able to apply “bald” to any single item in the sorites sequence.

the uses of “bald”. On the other, even supposing that (1) and (2) could both be true, they would have no role in the explanation of our uses of vague predicates; no more than their inexistence would have. It seems to me that Wright’s inconsistent rules are not really explaining vagueness, they are indeed explaining nothing.

Wright’s goal is to argue against what he calls the *governing view*. The governing view is a conjunction of two thesis: (i) our use of language can be correctly viewed as a practice in which the admissibility of a move is determined by rules; (ii) the general properties of these rules can be discovered by means of intuition and “working knowledge of the language from the inside, as users of it” (Burns: 1991, p.85). He thinks that arguments such as those I have provided above show that (ii) is false (Wright: 1975, p.363). But Wright cannot solve these problems only by refusing (ii). These problems are not problems about how we manage to discover in which rules our uses of vague predicates are rested on. What gives rise to all the objections discussed so far is the fact that Wright accepts tolerance as a rule for vague predicates, and also the fact that he accepts that vague predicates involve inconsistent rules. On the other hand, if we refuse also (i), then we have no reason to believe in tolerance anymore. In this case it would be better to just explain vague predicates entirely by the Other Half, without postulating any rule at all; that is, without postulating tolerance.

## 1.4 Eklund’s Incoherentism

Eklund (2002, 2005) has provided a different form of incoherentism for vague predicates.

Eklund’s thesis: Competent speakers are disposed to accept that whereas large enough differences in *F*’s parameter of application sometimes matter to the justice with which it is applied, some small enough difference never thus matters.

Eklund do not think that the principle of tolerance is true. In fact, the talk about dispositions and competence is meant to explain the very fact that tolerance is not true, even though we are disposed to accept it is. It is arguably plausible to suppose that competent users of vague predicates are disposed to accept an untrue principle. Although tolerance is not true, it is part of the semantic competence with vague predicates to accept tolerance; and it is also part of that competence to accept that large enough differences matter to a proper application of a

vague predicate. That is to say, competence with these expressions will involve dispositions “to accept some things that in fact lead to inconsistency” (2005, p. 41).

It should be noted that competent speakers are disposed to apply vague predicates in a restricted way only. They always do so and unrestricted applications of these predicates are not acceptable in ordinary uses of them (I will give one more reason for this unacceptability later). Even Eklund should accept this, on the pain of being empirically wrong. Could we reasonably claim that competent speakers are also disposed to accept tolerance? In this case the disposition to accept tolerance cannot be the same as the disposition to apply vague predicates unrestrictedly. I suspect that Eklund understands the first one as the disposition to accept each step in the step-by-step version of the sorites argument. Suppose that a competent speaker is facing a step-by-step version of the sorites argument for “bald”. She would feel forced to accept each of the steps, and this happens because on each step she would feel that if she calls a  $n$ -haired person “bald”, she has to call the  $n+1$ -haired “bald” too. In other words, she would feel that the mere difference of one hair on the head cannot change her application of “bald”. To be disposed to accept each step of the step-by-step version of the sorites argument is not the same as to be disposed to apply vague predicates as if each step was correct. In sum, competent speakers are disposed to accept tolerance but not to apply vague predicates unrestrictedly.

Eklund’s theory is more plausible than Wright’s version of incoherentism. But it has its own problems. According to Eklund’s view, Timothy Williamson is not a competent user of vague predicates. Given that Williamson is an advocate of epistemicism, he would not be disposed, I presume, to accept tolerance or each one of the conditionals in the step-by-step version of sorites argument. Because of this, he would not have the inconsistent dispositions. In reading Williamson’s papers about vagueness we can perfectly tell that he is a competent user of vague predicates, whether or not he believes in a wrong theory.

Additionally, it seems that there is no special problem in there being disagreement about whether or not some predicates are vague (Weatherson: 2009, p.87). We can discuss whether “good” is vague (perhaps to be good is to maximize utility in a vague way) without being incompetent users of this predicate. Those who disagree that “good” is vague may well be not disposed to accept each step of the step-by-step version of the sorites without thereby being incompetent users of “good”. We do not need to know (neither to agree) that a vague word is vague in order to competently use it.

Finally, Eklund talks about competent speakers being disposed “to accept that whereas large enough differences in  $F$ ’s parameter of application...”. As Weatherson notices, this formulation is ambiguous. Let us say that the parameter of application for “bald” is  $\Psi$ . Now

we can read Eklund's thesis in two different ways. On the first reading, he would be saying that competent users of "bald" are disposed to accept: that whereas large enough differences in  $\Psi$  matter, small enough differences in  $\Psi$  never thus matter. On the other reading, he would be saying that competent speakers are disposed to accept that: whereas large enough differences in the parameters of application of "bald" (whatever they are) matter, small enough differences never thus matter. Both readings have the problems mentioned above. But there are some problems specifically to each one of them.

On the first reading, competent speakers cannot have false beliefs about which is the relevant parameter for a vague predicate. If S has a false belief about which is the relevant parameter for "heap", for example, she will not be a competent user of our predicate "heap". Suppose the parameter for "heap" is  $\Psi$  but Mary believes it is  $\emptyset$ . In this case, Mary might be not disposed to accept that *whereas large enough differences in  $\Psi$  matter, small enough differences in  $\Psi$  never thus matter*. After all, she not even believes that  $\Psi$  is the relevant parameter for "heap". Aside this, it is hard to believe the parameters of application for vague predicates are generally easy to access. I have been treating "bald", for example, as if its only relevant parameter of application was the number of hairs on the head. But this is not the case. The distribution of hairs on the head is an important parameter too. To take another example, the parameter of application for "thin" is a set of measurements. But what specific measurements are included in this set? It is too much to require that speakers know the specific parameter which is at stake for each vague predicate in order to be able to use them in a competent way. The second reading does not have this problem, but it requires people having some concept of *parameter* before being a competent user of vague predicates. As Weatherson (2009, p. 87) notices, it is implausible to think that ordinary users should have some concept of a parameter before being able to use vague predicates. It seems to me that this supposition is even more implausible than the first reading.

All these problems have the same source: Eklund is making the conditions on our uses of vague predicates too strong. My intuition is that we should go to the opposite direction. The conditions on our uses of precise predicates might be strong, but the conditions on the use of vague words should be rather weak.

## 1.5 The vagueness of "substantial difference"

A problem for both Wright and Eklund's view is the fact that the notion of *substantial differences* – the difference between  $a$  and  $b$  which is large enough to make difference to

the application of a vague predicate – is itself vague. How many hairs more than  $a$   $b$  has to have in order for there to be a substantial difference between  $a$  and  $b$ ? It is not hard to find a plausible (but not correct, I think) answer in some cases. Think about “bald”. A plausible answer in this case is that there will be a substantial difference between  $a$  and  $b$  with respect to the number of hairs if  $a$  can be discriminated from  $b$  only by means of casual observation (Wright seem to suggest this). For other cases there is no clear answer. What would count as a substantial difference between a young man and a not young one? I can classify people as young or not young only by knowing their ages. Suppose I classify  $a$  as young. How older should  $b$  be in order to be substantially different from  $a$ ? The same problem will arise for several others vague words: “old”, “many” (as in “I go to the movies many times a week”), “much” (as in “I do not have much money”), “rich”, “poor”, “far away” (as in “ $a$  is far away from  $b$ ”), and so on. In all these cases, it seems that there will be a small enough difference that never matters to the justice with which the predicate is applied (this is the intuition behind the principle of tolerance). However, in all those cases it is vague what would count as a substantial difference. One cent is not a substantial difference between a rich and a non-rich person. If 1 cent is not a substantial difference,  $1+1$  is not too. If  $1+1$  is not, then  $1+1+1$  is not too... The notion of *substantial differences* (the “large enough” differences) is itself vague, at least as applied to many vague predicates. Being so it is tolerant.

This notion plays a crucial role in both Wright’s and Eklund’s view. Recall that Wright’s strategy was to say that we are allowed to restrict our application of a vague predicate  $F$ , applying  $F$  to an object  $x$  and its negation to  $y$ , when there is a substantial difference between  $x$  and  $y$ . If “substantial difference” is itself a vague predicate (or better, a relation), then it should be unrestrictedly applied. Now we will have to explain why this is not so. We would need an explanation for the fact that ordinary users of vague predicates do not do so. After all, they do not consider any difference as a (non)substantial one. Eklund’s strategy is to explain one side of the inconsistent dispositions by appealing to the substantial differences. If “substantial difference” is itself vague, then it should be explained by appealing to inconsistent dispositions which will require once again the notion of “substantial differences”. Now we are back to the beginning.

## 1.6 The positive case against unrestricted application

Let me now consider a positive argument for the thesis that at least some vague predicates must be restrictedly applied and cannot be unrestrictedly applied. I think there is a simple

argument to show this. In fact, I think we can (i) show that the predicate “bald” definitely applies to a 0-haired person and (ii) develop a non-vague notion of *ideal case* of “bald”. Consider the following sequence of people ordered by the number of hairs in their head.

0-haired person, 1-haired person, 2-haired person, ..., n-haired person

Let us suppose that “non-bald” applies to the first person, the 0-haired person. The minimal requirement explained at the beginning of this section says that *for all number n: if “non-bald” applies to a n-haired person, it applies to a n+1-haired person too*. Given this, “non-bald” applies to all the items in the sorites sequence. That is to say, “non-bald” will apply to everyone. Independently of how many hairs on the head a person has, she is non-bald. If this is true, then “bald” is not a predicate about lack of hairs. Of course “bald” is a predicate about lack of hair. Therefore, the supposition that “non-bald” applies to a 0-haired person is false. No one can correctly apply “non-bald” to 0-haired person. We can safely say that our predicate “bald” applies to a 0-haired person and it cannot be the case that its negation applies to a 0-haired person. That is to say, “bald” definitely applies to 0-haired person. We can formulate a similar argument to show that “non-bald” definitely applies to a person who has a hundred percent of her head covered by hair.

Now we cannot formulate this argument to show that “non-bald” cannot be applied to a 1-haired person. From the supposition that “non-bald” applies to a 1-haired person it does not follow that “bald” is not about lack of hair. It might be that “non-bald” applies to a 1-haired person and yet “bald” is about lack of hair (in this case, about total lack of hair). So this argument shows that “bald” definitely applies to 0-haired people, but not that it definitely applies in any other case. A 0-haired person is what I call, following Imaguire (2008, p.123), an *ideal case* of “bald”. The notion of ideal case is, at least for “bald”, a precise one. There is a sharp boundary between the ideal cases of “bald” and the non-ideal cases of “bald”; namely, only 0-haired people are ideal cases of “bald”. The same will apply to the ideal cases of non-bald.

This argument is quite plausible and it does not depend on any really controversial assumption. However, it has some interesting consequences. First, there is at least a case to which “bald” definitely applies. From this it follows that “non-bald” cannot be unrestrictedly applied. No sorites reasoning can force me to apply “non-bald” to a 0-haired person. Similarly, no sorites reasoning can force me to apply “bald” to a person who has a

hundred percent of her head covered by hair.<sup>8</sup> Against Wright, the principle of tolerance is just not true for “bald”. Against Eklund, no competent speaker needs to be disposed to accept tolerance or the step-by-step version of the sorites argument for “bald”.

Crispin Wright has said that incoherentists should accept that “there are no determinate F’s and non-F’s” (1994, p.141), where F is a vague predicate. At least for “bald”, the incoherentist cannot accept this. At any rate the existence of ideal cases is not a knock down objection against incoherentism or the principle of tolerance. This argument above shows that incoherentism, such as presented so far, is false. However, it is possible to defend a restricted version of the principle of tolerance for some vague predicates along these lines. The principle of tolerance does not force you to accept, for example, that a 0-haired person is non-bald, but it can be applied to all the other items in the sorites sequence (for simplicity, I will ignore the ideal case of “non-bald”). That is to say: *for all n such that n>0: if a person with n hairs is bald, then a person with n+1 hairs is bald too*. In other words, they could formulate tolerance in such a way as to accommodate the existence of ideal cases. Eklund, in his turn, might say that although we do not have to be disposed to accept each step of the step-by-step version of the sorites argument, we do have to be disposed to accept each step until an ideal case is in place; that is, we would be disposed to accept a restrict version of the principle of tolerance. This would be a more complicated and an even more implausible version of incoherentism, but it can be done.

## 1.7 Vague predicates and sorites susceptible predicates

I have been discussing the sorites predicates so far. In other words, I have been discussing predicates to which the sorites argument is compelling. According to incoherentism to be sorites susceptible is an essential property of vague predicates (for vague predicates are tolerant or at least we are disposed to accept that they are). Nonetheless, as Weatherson (2009, p.80) notices, there are predicates which are intuitively vague but not sorites susceptible. Take the predicate “has few children”. Now take the following sorites argument (I just repeat Weatherson’s formulation of the argument):

8 - It is not quite clear how important the notion of ideal cases will be to an account of vagueness. The most of the recent theories seem to depend on a less restrict notion of *clear case*, one which allows that “bald” definitely applies not only to 0-haired people, but also to many other cases. This is what gives rise to the problem of precisification and makes these theories so problematic. But it might be that some theory of vagueness can be formulated in such a way as to depend only on the more restrict notion of *ideal cases* (perhaps degrees-of-truth theories could say that only ideal cases of “bald” are bald in degree 1; contextualists could say that all the other cases are determined by context; or it might be that we can formulate a new theory which is construed out of the notion of ideal cases; or maybe this notion is not so important at all). In any case, the argument provides one more reason against incoherentism.

- (1) An academic with one child has few children.
- (2) If an academic with one child has few children, then an academic with two children has few children.
- (3) If an academic with two children has few children, then an academic with three children has few children.
- (4) If an academic with three children has few children, then an academic with four children has few children.
- (5) If an academic with four children has few children, then an academic with five children has few children.
- (6) So an academic with five children has few children.

Would you feel forced to accept this argument is sound? I don't think so. It is intuitive that "has few children" is a vague predicate. There is something special with its boundary and *this something* must be explained by a theory of vagueness. Yet it is not a sorites susceptible predicate. A competent user of "has few children" does not need (to be disposed) to accept this reasoning. Suppose, for example, that someone deny (3). Would we be allowed to say that he is "missing something we are able to see" (Eklund, 2005, p.41)? Contra Eklund, this clearly is not the case. And the same is true of other vague phrases such as "go to the movies few times a week", "watch movies very often", etc. It is not clear that a competent speaker would be missing something by categorically refusing a sorites argument for each of these cases.

It is true that many vague words are sorites susceptible, but we should be careful in drawing the conclusion that to be sorites susceptible is essential to vagueness. So far as our intuition can tell us, it is not. Following Soames (1998, p.217), I think the most plausible would be to distinguish the sorites predicates from vague predicates, in such a way that the first kind is a subset of the second one. Incoherentism does not provide us with any means by which we could do that.

## 2 Tolerance and the problem of precisification

The principle of tolerance (together with some non-controversial constraints) is supposed to explain the alleged fact that the difference of only one hair does not make difference to the application of “bald”. If I call a  $n$ -haired person bald, I have to call a  $n+1$ -haired person bald too. If I call her non-bald, I have to call a  $n-1$ -haired person non-bald too. Whatever a  $n$ -haired person is (bald or non-bald), her immediate neighbors in the sorites sequence will be too. Generalizing, tolerance is the explanation for the fact that very slight changes do not make difference to the application of vague predicates. We must recognize that the principle of tolerance makes sense of the intuition about very slight changes.

However, we have seen that tolerance is highly problematic. Ordinary speakers do not apply vague predicates as if they were strictly tolerant, their uses systematically violate tolerance. To appeal to some form of incoherentism, in turn, would also be implausible. In the face of all that, one might think that the best strategy is to deny that the principle of tolerance is true of vague predicates and look for some alternative way to explain the phenomenon of vagueness. Those who take this latter venue will have to deal with the problem of precisification.

### 2.1 Does tolerance explain imprecision?

It should be noticed that the principle of tolerance, by itself, does not obviously explain the intuition that vague predicates are imprecise. A precise predicate is such that there will be a boundary between the things to which it applies and the cases of any other kind. A non-precise predicate is such that it has no boundary, “no such distinction may be drawn between cases where it is definitely correct to apply the predicate and cases of *any* other sort.” (Wright, 1975, p.330). Let us take this for grant and ask if tolerance is a good way to explain this intuition.

Suppose the principle of tolerance is true for “bald”. Then, we have four options, which form what Williamson (1994, p.167) called the *all or nothing view*. (i) there is a case of bald but not of non-bald, (ii) there is a case of non-bald but not of bald, (iii) there is no case of bald and no case of non-bald and (iv) there is both a case of bald and a case of non-bald. It is hard to see which of these options would provide us with the relevant sense of imprecision.

The cases of (i) and (ii) are obvious. In (i) “bald” applies to all objects in the sorites sequence and “non-bald” is empty. In (ii) “bald” applies to nothing and “non-bald” applies

to all objects in the sorites sequence. In both cases “bald” would be a paradigmatic example of a precise predicate. In (iv) both “bald” and “non-bald” apply to every object in the sorites sequence, in which case “bald” and “non-bald” have the same extension. Again, both ‘bald’ and ‘non-bald’ would clearly have precise extensions. Finally, the principle of tolerance together with (iii) only implies that both “bald” and its negation are empty. It should be clear that to be empty is not a sufficient condition for being imprecise.

To be tolerant is not a sufficient condition for being imprecise. The reason why the principle of tolerance fails is what I have called the *non-restriction consequence*. Tolerance implies unrestricted application of vague predicates, it does not allow for the situation in which the vague predicate is applied to *a* and its negation to *b*. The result is the *all or nothing view* and none of its options gives us an intuitive picture about what imprecision is.

We could then argue that the problem of precisification, whatever is its source, is not due the rejection of tolerance. Even those who accept tolerance will face this problem. Incoherentists presuppose that tolerance is the source of imprecision and their explanations of the phenomenon of vagueness take us to a set of problems. They might be wrong (and I suspect they are wrong) about this. This is one more reason why we should take tolerance only as a metaphor to something else.

## 2.2 Backing to the traditional theories?

Since the principle of tolerance does not give us a plausible theory of vagueness, it might be that the best path is to refuse tolerance and accept that vague predicates do have boundaries (the theories that follow this strategy are what I call the *traditional theories*). In this case we will have to explain why we are so inclined to accept tolerance. Both contextualism and epistemicism follow this strategy. According to Williamson’s epistemicism, for example, we could not know where the boundary of a vague predicate is, although they have boundaries. We could never find out where the boundary of a vague predicate is, or even imagine how we could find out the exact boundary of a vague predicate. Because of this, it is tempting to believe that there is no boundary or that the existence of such a boundary is inconceivable. Yet this is a “fallacy of the imagination” (1994, p.3), and he goes on to argue that vague predicates do have a boundaries. According to Fara’s contextualism (2000), the intuition behind the principle of tolerance, roughly, comes from the fact that once we look at a pair of objects that are very similar to each other, the boundary of the vague predicates changes. This is why we are inclined to accept each particular instance of  $n$  ( $B_n \rightarrow B_{n+1}$ ) even

though we hesitate to think that the sorites argument is sound (if we do not hesitate, then the argument will not sound as a paradox at all). Thus both contextualism and epistemicism deny the principle of tolerance, but they do provide acceptable explanations for why we are so inclined to accept it. It is not clear to me how the advocates of supervaluationism or degree-of-truth theories could explain the intuition for tolerance, but there is no reason why think this is impossible.

This strategy is not more implausible than to accept tolerance and incoherentism. First, it is not even obvious that the principle of tolerance provides us with a good explanation for the imprecision of vague predicates. Second, incoherentism is not a plausible theory. Whether or not some of the above theories is right, the mere fact that they refuse tolerance is not enough to refute them.

## **Conclusion: The unsolvable problem of vagueness**

For those inclined to accept one of the traditional theories of vagueness, the diagnosis of this paper might be positive. After all, to refuse tolerance and the imprecision of vague predicates would not be less plausible than to accept it. For those who, like me, are really impressed by the problem of precisification, the result is a negative one. Neither incoherentism nor the traditional theories are correct. An ideal theory of vagueness should satisfy all of the following three criteria: (i) to explain vague predicates without making them precise, (ii) to explain vague predicates without making them incoherent and (iii) to explain vague predicates in a way which solves the sorites paradox. I suspect that the way to solve this problem is to realize that imprecision is not the same thing as tolerance. To be an imprecise predicate is not the same as to be a tolerant predicate. As we have seen, tolerance not only take us to incoherence, it not even provides us with a clear explanation of imprecision. What we have to do is to find out a way of explaining imprecision without appealing to tolerance and without violating the other two criteria for an ideal theory of vagueness. If there is no such way, then the problem is unsolvable one.

## **References**

Burns, Linda Claire. 1991. *Vagueness: an investigation into natural languages and the sorites*

*paradox*. Springer Science+Business Media Dordrecht.

Horgan, Terence. 1994. Robust Vagueness and the Forced-March Sorites Paradox. *Philosophical Perspectives*, 8, Logic and Language.

Eklund, Matti. 2002. Inconsistent Languages. *Philosophy and Phenomenological Research*. Vol. LXIV, No. 2

Eklund, Matti. 2005. What Vagueness Consists in. In: *Philosophical Studies* 125: 27–60

De Souza Filho, S. 2013. *Seguir Regras e Naturalismo Semântico*. Disponível em: <https://ppglm.files.wordpress.com/2008/12/dissertacao-ppglm-sergio-souza.pdf>

Fara, Dalia Graff. 2008. Profiling Interest Relativity. *Analysis* 68.4, October 2008, pp. 326–35.

Fara, Dalia Graff. 2000. Shifting Sands: An Interest-Relative Theory of Vagueness. *Philosophical Topics*. Vol. 28, NO.1

Imaguire, Guido. 2008. “A Vagueza e os Paradoxos Sorites”. In: Imaguire, Guido & Schirn, Matthias. *Estudo em Filosofia da Linguagem*. Edições Loyola, São Paulo, 2008.

Rafman, Diana. 2005. How to understand contextualism about vagueness: reply to Stanley. *Analysis*, 65.3, pp.244-48.

Soames, Scott. 1998. Vagueness, Partiality and the Sorites Paradox. In *Understanding Truth*, Oxford University Press, 1999, pp. 203-227.

Weatherson, Brian 2009. Vagueness as Indetermination. In: Dietz, Richard & Moruzzi, Sebastiano (ed.). *Cuts and Clouds: Vagueness, Its Nature, and Its logic*. Oxford University Press, pp. 77-90.

Williamson, Timothy 1994. *Vagueness*. Routledge.

Wright, Crispin. 1975. On The Coherence of Vague Predicates. *Synthese* 30 (1975) 325-365.

Wright, Crispin. 1976. Language-Mastery and the Sorites Paradox. In: Evans, Gareth & McDowell, John (1976) *Truth and meaning: Essays in Semantics*. Oxford.

Wright, Crispin. 1994. The Epistemic Conception of Vagueness. *The Southern Journal of Philosophy* (1994) Vol. XXXIII, Supplement.

# Traduções



# O Desvio Metafórico<sup>1\*</sup>

*Monroe C. Beardsley*

*Tradução de Carlos André Fernandes  
Revisão de L. H. Marques Segundo*

De todas as questões sobre a metáfora que interessam aos teóricos da literatura ou aos estetas, a principal – isto é, a primeira e fundamental – é, sem dúvidas: O que ela é? Dar uma explicação adequada da metáfora enquanto fenômeno linguístico, e que sirva como base para uma explicação enquanto fenômeno poético, é dizer o que é peculiar às expressões metafóricas, de que modo elas diferem de expressões literais, como as reconhecemos e como sabemos o que elas significam.

Não é fácil dizer exatamente quais são as questões sobre este problema. Há vários modos de se descrever a metáfora, alguns deles remontando a tempos antigos, que são tão familiares e com tanta confiança ecoaram de um escritor a outro que todas elas têm o ar de serem mais ou menos equivalentes. Mas há, acredito, uma distinção importante entre elas, e parte da minha intenção aqui é provocar uma cisão – separar as coisas mais claramente do que tem sido feito. Pretendo distinguir o que pode ser chamado de abordagem do objeto e abordagem da palavra para o problema da análise da metáfora.

Segundo uma dessas perspectivas, considerada amplamente, o modificador (como o

---

1 - \* The Metaphorical Twist, publicado originalmente em *Philosophy and Phenomenological Research*, Vol. 22, N°3 (Mar., 1962), pp. 293-307

chamo) na metáfora – por exemplo, a palavra “raivoso” em “O sol raivoso” – preserva seu papel designativo padrão ao entrar na metáfora e, portanto, continua nesse contexto a denotar os mesmos objetos que denota em contextos literais. Por conseguinte, a metáfora é uma comparação implícita, uma símile elíptica, e diz, com efeito, que o sol é como uma pessoa raivosa. A pessoa raivosa é referida, nesse contexto, assim como o sol – há dois objetos. A metáfora, como Johnson disse, “nos dá duas idéias ao invés de uma.” Mr. John Crowe Ransom classificou as metáforas como “Importadores” que introduzem “objetos estranhos” na “situação”<sup>2</sup>) – Acho que ele estava pensando naqueles importadores extravagantes de víveres exóticos, como trufas e abelhas cristalizadas. A metáfora, em sua perspectiva, arrasta um objeto estranho e impróprio para dentro do contexto (deleitándonos, diriam os teóricos do século dezoito, por seu charme e novidade), e, assim, acrescenta a ela a “textura local de irrelevância” que Ransom considera tão essencial à poesia.

Chamar-lhe-emos de Teoria da Metáfora da Comparação Objetual. Segundo sua rival, a Teoria da Oposição Verbal, nenhuma importação ou comparação ocorre de todo em todo, mas, ao invés, um feito especial da linguagem, ou um jogo verbal, envolvendo dois níveis de significado no próprio modificador. Quando um predicado é metaforicamente unido a um sujeito, ele perde sua extensão comum, pois adquire uma nova intensão – talvez uma intensão que não teria em nenhum outro contexto. E essa mudança de significado é forçada por tensões, ou oposições, inerentes à própria metáfora.

Proponho-me a oferecer razões para rejeitar a Teoria da Comparação Objetual tanto em sua forma geral quanto em sua forma especial recentemente oferecida. Em seguida explicarei mais pormenorizadamente a Teoria da Oposição Verbal, e a defenderei contra algumas possíveis objeções.

## I

Admito, até certo ponto, que não importa se você fala acerca das metáforas no modo objetual ou no modo semântico. Mas até certo ponto. Suponha, pois, que a palavra “sarça” seja introduzida metaforicamente em certo contexto, como, digamos, em “East Coker” – a referência a

frígidos fogos do purgatório

---

2 - “William Wordsworth: Notes Toward an Understanding of Poetry,” *Kenyon Review*, XII (Summer 1950): pp. 498-519.

cujas chamas são rosas, e a fumaça sarças.<sup>3</sup>

Você pode começar a sua explicação tanto na linguagem-objeto (falando sobre as características das sarças) quanto na metalinguagem (falando sobre a conotação da *palavra* “sarça”). Você pode dizer, “As sarças têm a capacidade de arranhar as pessoas, retardar o seu progresso, ser transformadas em cachimbo,” e assim por diante. Ou você pode dizer, “A palavra ‘sarça’ conota propriedades como a de arranhar, retardar o progresso ser transformada em cachimbo,” e assim por diante. Contudo, embora esses dois modos de falar se sobreponham, uma vez que em parte a conotação da *palavra* deriva do que é comumente verdadeiro sobre os *objetos*, elas não coincidem completamente.

Pois as conotações são controladas não só pelas propriedades que os objetos efetivamente têm, mas por aquelas que amplamente acreditamos que possuem – ainda que a crença seja falsa. Este é então o meu primeiro argumento contra a Teoria da Comparação Objetual – uma adesão consistente a essa teoria produziria explicações incorretas e incompletas das metáforas em casos em que o modificador tivesse conotações, aplicáveis naquele contexto, que não fossem características acidentais comuns dos objetos denotados. Por exemplo, um dos importantes significados marginais de “sarças” no poema de Eliot vem, é claro, do modo como a coroa de espinhos aparece na narrativa cristã. E, completamente à parte de sua verdade histórica, a existência dessa religião é suficiente para dar à palavra esse significado. Se ao explicá-la nos limitarmos ao que sabemos sobre sarças, não a compreenderíamos completamente.

Meu segundo argumento contra a Teoria da Comparação |Objetual é que uma vez que nos comprometemos a encontrar, ou fornecer, um objeto para ser comparado ao sujeito da metáfora (isto é, nos termos de I. A. Richards, um “veículo” que o fará avançar) abrimos o caminho para esse fluxo de imagens idiossincráticas que é uma das sérias barreiras entre um leitor e um poema. Considere um exemplo também discutido por Mr. Ransom, a passagem sobre a adaga de Brutus no discurso de Antônio:

Observai com cuidado como sangue de César o seguiu,  
Como querendo vir para a porta, a fim de convencer-se  
Se era Brutus, realmente, quem batia por modo tão grosseiro

---

3 - T. S. Eliot, *Four Quarters*, New York: Harcourt, Brace, 1953, p. 16.

(III, ii, 178). Ransom fala da “mudança” do teor (o sangue) para um “criado” abrindo a porta, sendo tal criado o veículo.<sup>4</sup> Não há, obviamente, qualquer criado nessa passagem, não mais que há um chefe de família rudemente despertado ou um agricultor em apuros a ser alarmado por Paul Revere. De onde vem o criado? A terminologia teor-veículo, com seu pressuposto subjacente que a metáfora deve ser uma comparação, instiga o explicador a inventar, onde ele não pode encontrar um veículo, e, assim, obtermos o criado. Mas “querendo vir para a porta” não é exatamente sinônimo de “o criado querendo vir para fora”, como fora aplicado ao sangue de César. E esse é o primeiro significado para o qual o teórico dá atenção, sem qualquer significado maior importado – uma boa descrição – arbitrariamente. Citando uma metáfora característica de Samuel Johnson, “O tempo é, de todos os modos de existência, o mais servil à imaginação,” William K. Wimsatt, Jr., nota, “Não precisamos imaginar o Tempo como um servo curvando-se a seu mestre, a Imaginação.”<sup>5</sup>

Meu terceiro argumento contra a Teoria da Comparação Objetual é que ela tende a conduzir à infeliz doutrina da “adequação.” Se uma metáfora é uma comparação, é possível perguntar se a comparação é “adequada” ou “inadequada.” Vemos isso no quarto tipo de mau gosto de Aristóteles, sua objeção à expressão de Górgias, “eventos que são frescos e repletos de sangue.”<sup>6</sup> Se tomarmos as palavras de Macbeth, (II, iii),

suas adagas

monstruosamente cobertas com sangue

está comparando o sangue das adagas e pernas cobertas, e, se investigarmos sua adequação, estaríamos propensos a dizer, como o crítico do século dezenove citado por Cleanth Brooks,<sup>7</sup> que Shakespeare “nos enoja com a sua tentativa de comparação.” Mas a pergunta correta é o que *se quer dizer* por essas palavras - quais propriedades são atribuídas a adagas através do significado marginal do atributo metafórico. Não importa no momento se adagas ensangüentadas em geral são mais bem descritas assim; a pergunta é: o que aprendemos dessa descrição sobre *essas* adagas e seu papel na estória completa – ou sobre o

4 - “Poetry: I. The Formal Analysis; II. The Final Cause,” *Kenyon Review*, IX (Summer 1947): 436-56, (Autumn 1947): pp. 640-58.

5 - *The Prose Style of Samuel Johnson*, New Haven: Yale University Press, 1941 (Yale Studies in English, Vol. 94), p. 66.

6 - Lane Cooper, trans., *The Rhetoric of Aristotle*, New York: Appleton, 1932, p. 192.

7 - *The Well Wrought Urn*, New York: Reynal and Hitchcock, 1947, p. 29. The New Variorum Edition, ed. H. H. Furness, 5<sup>th</sup> ed., Philadelphia: Lippincott, 1915, pp. 160-61, mostra divertidamente o que um quebra-cabeça lanciante teria sido para os teóricos Shakesperianos.

falante que a descreveu desse modo.

Dito de modo mais geral: Suponha que o poeta comente, “A minha amada é o meu Schopenhauer.” Na Teoria Comparativista, perguntaríamos o que a sua amada e Schopenhauer têm em comum. Mas não *conhecemos* a sua amada, e, por isso, como podemos responder a essa pergunta até que ele nos diga, pela própria metáfora, como ela é? A pergunta correta é: quais os possíveis significados “Schopenhauer” pode atribuir à amada, e que não são excluídas pelo contexto.

## II

A minha objeção geral à Teoria da Comparação Objetual aplica-se, acredito, com força semelhante à forma muito interessante dessa teoria que foi oferecida por Paul Henle: a Teoria Icônica do Significado.<sup>8</sup> O Sr. Henle atualmente parece sustentar ambas as teorias às quais me referi. A sua versão da Teoria da Oposição Verbal, contudo, é descrita em termos da resposta do leitor – sua “surpresa” no “conflito de significado.”<sup>9</sup> Prefiro apresentar a teoria, não como uma teoria sobre o efeito da metáfora, mas como uma teoria acerca da estrutura lingüística que causa o efeito – sobre o “conflito de significado” em si. O Sr. Henle disse pouco sobre isso, e ele não explica sua relação com sua outra, e principal, teoria – de fato, ele não diz por que deveria haver qualquer surpresa, ou conflito, se a outra está correta.

O Sr. Henle sustenta que “há um elemento icônico na metáfora,” e propõe analisar as atribuições metafóricas em termos do conceito de signos icônicos. Em seu exemplo extraído de Keats, “pensamentos de ódio envolvem minha alma na escuridão,” ele diz haver duas relações: primeiro, a palavra “envolver” designa certa ação – “envolvimento em um manto.” Segundo, essa ação passa a ser um signo icônico de escuridão.<sup>10</sup> “Em uma metáfora alguns termos simbolizam o ícone e outros simbolizam o que é iconizado.”<sup>11</sup>

Podemos começar por perguntar como o manto entra nessa explicação. A Teoria Icônica parece importar um objeto estranho de algum tipo – como o criado na caso do

---

8 - Paul Henle, em seu capítulo sobre metáfora em, Ed., *Language, Thought, and Culture*, Ann Arbor: University of Michigan Press, 1958, ch. 7, tem um desenvolvimento e também uma modificação se sua primeira perspectiva brevemente definida diante de seu discurso Presidencial para a Divisão Ocidental da American Philosophical Association, “The Problem of Meaning,” in *Proceedings and Address of the American Philosophical Association*, 1953-54, Vol. XXVIII, Yellow Springs: Antioch Press, 1954.

9 - *Language, Thought, and Culture*, pp. 182-83.

10 - *Ibid.*, pp. 177-79.

11 - *Ibid.*, p. 181.

Sr. Ransom – e isso enfrenta dificuldades na teoria da importação. O Sr. Henlen aceitou ainda a doutrina da “conveniência” que, como sugeri, a Teoria Comparativista, ao menos, torna atraente. Assim, ele diz que “não teria falado de pensamento de ódio *capturando* a alma na escuridão,” porque armadilhas “são todas mordazes, com as bordas definidas, e isso prejudica a correspondência com a escuridão.”<sup>12</sup> Talvez eu não devesse insistir tanto nessa observação, mas devo dizer que tal se generaliza num princípio crítico ainda mais surpreendente. Em meu entender, penso que o problema de se mantas e armadilhas são melhores signos icônicos de escuridão é uma questão completamente irrespondível, e, felizmente, não necessita de ser respondida. Se o falante no poema tivesse sido preso na escuridão, ao invés de envolto nela, simplesmente algo diferente teria sido dito sobre ele, e sobre como ele se sentiu, e veio assim a se sentir – o que poderia ter feito um poema pior, ou melhor, dependendo de diversas outras coisas.

O Sr. Henle cita um dos exemplos de Aristóteles sobre o modo pelo qual “analogias proporcionais” podem ser invertidas: podemos dizer uma coisa ou outra, ou que o escudo é a taça de Ares, ou a taça é o escudo de Dionísio.<sup>13</sup> “Que tal inversão seja possível, certamente, é uma consequência do caráter icônico da metáfora,” diz o Sr. Henle.<sup>14</sup> E talvez isso se siga de alguma Teoria da Comparação Objetual, pois, se A pode ser comparado a B, por que B não pode ser comparado a A? E um enunciado de semelhança é equivalente à sua própria conversa. Mas, se isso se segue, então é uma objeção fatal à teoria. Ora, o Sr. Henle percebeu que há uma dificuldade, e por isso ele diz que embora as metáforas sejam sempre reversíveis, às vezes o “tom sentimental é diferente.” Não acredito que seja assim: a diferença entre “este homem é um leão” e “aquele leão é um homem”<sup>15</sup> está naquilo que os diferentes modificadores metafóricos atribuem aos dois objetos. Na Teoria da Oposição Verbal não se segue que porque A é metaforicamente B, B é, portanto, metaforicamente A. Essa é exatamente a diferença entre uma metáfora e a analogia proporcional de Aristóteles, ou um símile relacional – ainda que Aristóteles pensasse que a diferença não fosse grande. E certamente a Teoria da Oposição Verbal está correta sobre essa consequência, enquanto que a Teoria Icônica, se implica que ao chamar homens de leões e leões de homens estamos em ambos os casos atribuindo a mesma propriedade, é claramente falsa.

Há outra objeção que pode, penso eu, ser justamente levantada contra a Teoria Icônica do Significado. Dever-se-ia contar como um mérito numa teoria da metáfora a

---

12 - *Ibid.*, p. 180.

13 - Lane Cooper, trans., *op. cit.*, p. 193.

14 - Henle, *op. cit.*, p. 190.

15 - Eu uso seu exemplo, mas não sua explicação, de R. P. Blackmur, "Notes on Four Categories in Criticism," *Sewanee Review*, LIV (October 1946): PP. 576-89. Seria, a propósito, interessante ouvir uma defesa da reversibilidade de Ezra Pound em seu "Your mind and you are our Sargasso sea."

sua capacidade de analisar a metáfora nos mesmos termos que faria com os oximoros. Isso ajuda na economia da teoria, e encaixa com as afinidades evidentemente profundas entre esses dois tipos de locuções. Ora, a Teoria Icônica é um tanto deficiente aqui, pois não parece funcionar bem para oximoros. Em “choro mudo” (se isso pode ser tomado como um exemplo), deveríamos dizer que uma pessoa muda foi feita de signo icônico de algo que não é mudo: a falta de som tornou-se um signo de som. Isso não é muito convincente. A verdade parece ser, ao invés, que no oximoro temos o arquétipo, a forma mais aparente e intensa da oposição verbal.

### III

Se nos voltarmos aos objetos referidos por uma metáfora, e considerarmos os significados das palavras em si, deveríamos procurar pela metaforicidade da metáfora, por assim dizer, em algum tipo de conflito que está ausente na expressão literal. Uma direção na qual esse conflito tem sido procurado pode prontamente ser dada como sem saída, assim me parece. Essa abordagem contrasta o significado da expressão em si e a idéia na mente do falante (ou autor). Chamar A de B metaforicamente, nessa perspectiva, é dizer que A é um B sem significá-lo – a metáfora é uma forma de ironia.<sup>16</sup> Apela-se aqui implicitamente à intenção, e a teoria sofre de todo o mal associado a essa noção. Não decidimos que uma palavra em um poema é usada metaforicamente porque sabemos o que o poeta estava pensando; ao invés, sabemos o que ele estava pensando porque vemos que a palavra é usada metaforicamente. As pistas para esse fato têm de estar de alguma maneira no próprio poema, ou raramente seríamos capazes de ler poesia.

Há uma alusão de uma perspectiva similar na excelente abordagem à metáfora no livro recente de Isabel Hungerland.<sup>17</sup> Na metáfora, ela diz, “Tem de haver algum ponto determinável no desvio ou violação do uso comum – dito de outro modo, a violação tem de ser deliberada.” A Sra. Hungerland disse uma vez que a segunda cláusula aqui foi inadvertida; mentionei-a apenas para enfatizar que as duas cláusulas estão certamente longe de serem

**16** - Anthony Nemetz, em um artigo recente, “Metaphor: The Daedalus of Discourse,” *Thought*, XXXIII (Autumn 1958): pp. 417-42, baseia seu argumento sobre a fórmula que “uma metáfora consiste de duas partes: 1. O que é dito; 2. O que se entende” (419); a questão é, então, qual é a relação entre elas? Mas essa formulação recebe a investigação na direção errada. Uma metáfora é um “enunciado” como uma expressão literal é: Podemos dizer coisas ou literalmente ou metaforicamente, e em ambos os casos só se poder ser entendido à medida do que podemos dizer. Em uma observação sarcástica, o que é sugerido é oposto do que é afirmado, mas se não deixarmos a palavra “enunciado” abranger tudo, estamos certos de pensar que interpretar a observação é um processo de procurar por uma intenção escondida por trás.

**17** - *Poetic Discourse*, Berkeley and Los Angeles: University of California Press, 1958 (University of California Publications in Philosophy, No. 33), pp. 108 – 110.

equivalentes, pois metáforas acidentais ou não intencionais são perfeitamente possíveis.<sup>18</sup>

A oposição que torna metafórica uma expressão está, então, no interior da estrutura do próprio significado. A característica central da Teoria da Oposição verbal eu já expus em outro lugar,<sup>19</sup> mas eu gostaria de recapitulá-la brevemente aqui. Nessa versão, eu disse que a possibilidade da performance metafórica – as oportunidades que uma língua viva apresenta para brincar com o significado neste modo particular – depende de uma diferença notada entre dois conjuntos de propriedades na intensão, ou significado, de um termo geral: primeiro, essas propriedades que (ao menos em um dado tipo de contexto) são tomadas como condições necessárias para a aplicação correta do termo em um sentido particular (essas são as propriedades definidoras, ou designadas, ou o significado central do termo nesse tipo de contexto); segundo, essas propriedades que pertencem ao significado marginal do termo, ou (no sentido da palavra do crítico literário) sua conotação – propriedades que um falante pode, em um contexto apropriado, mostrar que atribui a um objeto por meio do uso desse termo sem pretender seguir uma regra que não aplicaria ao termo para esse tipo de objeto caso não tivesse essa propriedade. Eu disse que quando um termo é combinado com outros de modo que haveria uma oposição lógica entre seu significado central e o significado dos outros termos, ocorreria uma mudança do significado central para o significado marginal que nos mostraria que a palavra tem de ser entendida de forma metafórica. Este é o único modo em que ela pode ser tomada sem absurdidade. O termo “oposição lógica” aqui inclui tanto a incompatibilidade direta das propriedades designadas quanto a incompatibilidade indireta entre as pressuposições dos termos – como quando nosso conceito de sol exclui a possibilidade do comportamento voluntário que é pressuposto pelo termo “raivoso”. A oposição lógica é o que dá ao modificador o seu desvio metafórico.

Uma atribuição metafórica, então, inclui dois ingredientes: uma distinção semântica entre dois níveis de significado, e uma oposição lógica em um nível. Desse modo, não há problemas de “raivoso”, em um contexto metafórico, denotar uma pessoa raivosa e introduzi-la com o propósito de comparação; o preço a se pagar pela admissão desse contexto é que sua função ali é para significar somente suas características conotadas.

Assim é a Simples Teoria da Oposição Verbal como eu a tenho definido, e parece-me estar correta até aqui. Isto é, acredito que o fenômeno descrito, a mudança da designação para a conotação, ocorre de fato. Mas receio que não seja o bastante. Penso que algo mais

**18** - Veja Walker Percy, “Metaphor as Mistake” *Sewanee Review*, LXVI (Winter 1958): pp. 79 – 99. Percy mostra de maneira interessante como pode haver “erros que... resultaram em uma experiência estética poética” (80). No entanto, ele parece também enfraquece no fim, quando diz “que a autoridade e a intenção de Nomear, é o elemento essencial do significado da situação” (93).

**19** - Veja *Aesthetics: Problems in the Philosophy of Criticism*, New York: Harcourt, Brace, 1958, Ch. III.

importante também ocorre em pelo menos algumas metáforas.<sup>20</sup> E para explicá-la, temos de fazer (ou tornar mais explícito do que na primeira versão) duas distinções.

As conotações de uma palavra para objetos de certo tipo, acordar-se-á, são extraídas do conjunto total de propriedades acidentais encontradas ou atribuídas em tais objetos. Chamemos esse conjunto de propriedades acidentais de *âmbito de conotação potencial* da palavra. Em um dado momento da história da palavra, contudo, nem todas essas propriedades terão sido usadas. Pense, por exemplo, em um número de propriedades características das árvores, embora não necessariamente presente em todas: ter folhas, fazer sombra, ter raiz, ser alta, ser fina, ter casca, ser flexível no vento, força, e assim por diante. Algumas dessas, tais como ter folhas, fazer sombra, ser alta, estão claramente dentre as conotações reconhecidas de “árvore,” sendo prontamente evocadas em metáforas familiares. Elas podem ser chamadas de *conotações principais*. Outras propriedades, tais como talvez ser fina e ter casca, não parecem ser conotações principais, embora possam ser suficientemente características das árvores e estar disponível no âmbito de conotação potencial. Elas podem esperar, por assim dizer, ocultas na natureza das coisas, pela realização – aguardando serem capturadas por pela palavra “árvore” como parte do seu significado em algum contexto futuro.

A primeira distinção, pois, é entre dois conjuntos de propriedades acidentais – não é uma distinção nítida, nem uma distinção que possa sempre ser lapidada com confiança, mas ainda assim objetivamente determinável. A segunda distinção é entre dois tipos de metáfora – e está sujeita a qualificações similares.

Suponha que comecemos por dividir as metáforas em duas classes. Tentemos colocar na Classe I, metáforas como “sol sorridente” e “a lua espiando atrás de uma nuvem.” Note que não são metáforas mortas – esse não é o problema aqui São metáforas vivas, mas de alguma maneira elas são diferentes daquelas que poderíamos por na Classe II: “o sol raivoso,” “sol rebelde,” “sol fiel,” “lua inconstante.” Reconhecemos, parece seguro dizer, que as metáforas na Classe II são mais interessantes que aquelas na Classe I – que não é dizer, é claro, que elas são melhores em todo contexto poético. Mas qual é a diferença?

Ora, nos termos da Teoria Simples da Oposição Verbal, algo pode ser dito em explicação. As metáforas da Classe II são mais complexas do que as da Classe I. Elas parecem dizer mais sobre o objeto. Assim, enquanto descrição, elas são mais precisas, mais penetrantes. Falar do sol como “sorridente” é pressupor um amplo contraste com um sol que não obteve

---

<sup>20</sup> - A parte que é nova (para mim) na minha explicação de agora da metáfora não me ocorreu após, e à luz de, os artigos de Mr. Henle e Mrs Hungerland, que foram meus colegas de simpósio quando o presente artigo, em uma forma diferente, foi lido perante o 17th encontro anual da American Society for Aesthetics, Cincinnati, Ohio, Oct. 29-31, 1959. As críticas de Mr. Henle à Teoria da Oposição-verbal como incapaz de explicar o elemento de novidade presente no significado metafórico, e a discussão que se seguiu dos artigos, deixou-me a presente linha de pensamento.

sucesso em sorrir, talvez, ou que está nervoso e está a pino sobre o deserto. Mas falar do sol como “bravio” é pressupor uma distinção clara entre essa qualidade e outras qualidades concebidas com especificidade comparável: obediência, pontualidade, deferência para com seus desejos. Ora, a Teoria da Oposição Verbal, mesmo em sua forma simples, permite graus de complexidade, e por isso, talvez, ela possa, pelo menos parcialmente, explicar a diferença entre as duas classes. Contudo, parece haver algo mais a ser dito.

É neste ponto que nos deparamos com um problema um tanto complicado. Pois, uma sugestão que parece demasiadamente óbvia, é esta: as metáforas da Classe I são triviais e banais; as da Classe II são estimulantes e novas. Se há verdade nesta descrição, então ela só pode ser alcançada e mantida por meio de manobras hábeis em torno de algo pouco profundo e ilusório. Em primeiro lugar penso que não temos de supor que isso seja uma questão de mera repetição. Talvez “sol sorridente” tenha sido muito mais dito que “lua inconstante,” mas ainda que estivéssemos a repetir a expressão de *Romeu e Julieta* um sem número de vezes até que estivéssemos cansados disso, e, por conseguinte, não estivéssemos em posição de prestar atenção ao seu significado, isso apenas não a tornaria trivial. Em todo caso, se ser trivial é uma noção que diz respeito à frequência, então não é ela o que faz a diferença aqui. No entanto, em segundo lugar, a natureza de uma metáfora particular não pode ser inteiramente independente de seu período na história da literatura Inglesa. Pois, o que ela significa, ou pode querer dizer, em um dado momento deve depender, até certo ponto, em que outros contextos as palavras apareceram, e que expressões análogas ou paralelas existem nessa língua.

## IV

Suponhamos que quando a metáfora “a lua inconstante” foi formulada pela primeira vez em português, foi a primeira vez que “inconstante” fora usada metaforicamente – ou, pelo menos, a primeira vez que foi aplicada a um objeto inanimado. (Isso certamente não impede a possibilidade de poder ter sido aplicada originalmente a *apenas* objetos inanimados, digamos, ao seu movimento rotacional; se em algum momento ela veio a ter o significado psicológico ou comportamental, como seu significado primário, então podemos falar do primeiro uso metafórico após esse período). Neste momento a palavra “inconstante” não *tem* conotações. Quando, portanto, nos deparamos com “lua inconstante,” pegamos adequadamente a oposição verbal, mas ao procurar as conotações relevantes tropeçamos. Como podemos, então, explicá-la? Dados a sintaxe circundante e o tom predominante,

ela claramente faz sentido; temos, portanto, de tentar dar-lhe sentido. Assim, procuramos dentre as propriedades acidentais ou contingentes de pessoas inconstantes em geral, e atribuímos essas propriedades, ou tantas quanto pudermos, à lua. E essas propriedades, pelo menos por enquanto, tornar-se-iam parte do significado de “inconstante,” embora anteriormente elas fossem propriedades apenas daquelas pessoas. Poderíamos, então, dizer que a metáfora transforma uma *propriedade* (efetiva ou atribuída) em um *sentido*. E se, beneficiando-se dessa licença, outros poetas estivessem a procurar outras aplicações metafóricas para “inconstante,” que empregasse as mesmas propriedades e criasse sentidos similares, ou sobrepostos, então esses sentidos poderiam conectar-se muito proximamente à palavra de modo que se fixariam relativamente como conotações dessa palavra. Desse modo, as metáforas não só efetivariam uma conotação potencial, mas a estabeleceriam como um grampo.

É neste ponto que a Teoria da Comparação Objetual faz de fato a sua contribuição. Pois ela está correta em dizer que às vezes ao explicar metáforas temos de considerar as propriedades dos objetos denotados pelo modificador. Mas tais objetos não são referidos para comparação: são referidos de modo que se possa dar um novo status a algumas de suas propriedades relevantes enquanto elementos de significado verbal.

Suponhamos que em um dado momento na história da língua portuguesa tivéssemos disponível tais metáforas como “céu sorridente,” “mar sorridente,” e “jardim sorridente.” O modificador não pode, é claro, significar exatamente a mesma coisa em todos esses contextos, mas tem de haver algum significado em comum. E suponhamos que este significado comum já esteja estabelecido com a conotação de “sorridente.” Quando um poeta pela primeira vez fala de um “sol sorridente,” o que acontecerá? A oposição lógica é clara, de modo que nos voltamos primeiro para a conotação afixada de “sorridente” e aplicamo-la ao sol (como a Teoria Simples da Oposição-verbal diz). Mas não iremos tão longe. Talvez não possamos ir tão longe; talvez apenas não sejamos forçados a ir. Em todo caso, vemos que é uma metáfora, e podemos interpretá-la corretamente, embora não a tomamos como *criando* significado do mesmo modo como as metáforas da Classe II. Isso é simplesmente tomar emprestado o seu sentido, baseando-se no que já está estabelecido e disponível.

A Teoria Revisada pode muito bem ser ilustrada por meio de uma metáfora bastante interessante à qual tomei emprestada de Paul Henle. Em um de seus trabalhos religiosos, Jeremy Taylor disse que “Casamentos castos são honráveis e agradam a Deus,” que a viuvez pode ser “amável e graciosa quando é adornada com gravidade e pureza,” porém que “a virgindade é uma vida de anjos, o esmalte da alma...”<sup>21</sup> Este não foi o primeiro uso

21 - *Of Holy Living*, Ch. II, sec. 3, in *Works*, ed. C.P. Eden, Vol. III, London 1847, p. 56.

Mr. Henle usou este exemplo em seu artigo apresentado no simpósio.

metafórico de “esmalte;” aprendemos no *NED* que Donne, em 1631, já teria usado a expressão “esmaltado com essa bela Doutrina das boas Obras,” e que Evelyn, em 1670, teria usado a expressão “esmalte seus caracteres.” Além disso, o próprio Taylor, na dedicatória de seus *Sermões*, falou “daquelas verdades que são o esmalte e a beleza de nossas igrejas.” Talvez tais usos já tenham estabelecido algumas das propriedades do esmalte como conotações fixadas da palavra; talvez não. Teríamos de saber isso a fim de saber o quão exatamente estava na Classe II a metáfora “o esmalte da alma” no contexto do *Of Holy Living* de Taylor. Mas da nossa própria época podemos fazer um julgamento mais seguro. O esmalte é rígido, resistente a colisão e a arranhões, é aplicado com labor e perícia, e é decorativo. Eu pensaria que algumas dessas características não são completamente estabelecidas como conotações reconhecidas da palavra. Contudo, falar da virgindade como o esmalte da alma é dizer seguramente (como o Sr. Henle mostrou) que ela é uma proteção para a alma, e que é o toque final de adereço sobre o que já está bem-feito. Assim, essa metáfora não introduz simplesmente conotações latentes no primeiro plano do significado, mas traz à cena algumas propriedades que não eram anteriormente significadas por ela.

Parece-me provavelmente que temos de distinguir pelo menos três estágios nessa metamorfose do significado verbal, muito embora o ponto de transição não esteja claramente destacados. No primeiro estágio temos uma palavra e as propriedades que não fazem parte definitivamente da intensão dessa palavra. Algumas dessas propriedades são elegíveis a se tornarem parte da intensão, a juntar-se ao âmbito da conotação. A fim de serem elegíveis, elas têm de ser propriedades (efetivas ou imputadas) claramente comuns, propriedades típicas – não só no sentido estatístico, mas presente normalmente ou caracteristicamente nos objetos denotados pela palavra. Desse modo, suponha, por exemplo, que alguém diga que a brancura pudesse se tornar uma conotação de “esmalte.” Isto poderia acontecer, penso eu, se a maioria ou todos os esmaltes fossem brancos, ou se os esmaltes fossem geralmente brancos exceto quando afetados por condições externas, ou se o melhor esmalte fosse branco, ou se as coisas mais brancas dentre as coisas brancas fossem esmaltadas.

Quando a palavra começa a ser usada metaforicamente em um determinado tipo de contexto, o que antes era apenas uma propriedade torna-se, pelo menos temporariamente, um significado. E uma familiaridade ampla com essa metáfora, ou metáforas similares, pode fixar a propriedade como uma parte estabelecida do significado. Ela não é, ainda, uma condição necessária a ser aplicada na palavra, nesse segundo estágio. Ainda que “árvore” conotasse altura, não há contradição em dizer que uma árvore é baixa ou atrofiada. Entretanto, se alguém dissesse que sua árvore é uma árvore “no sentido pleno” (compare com “Ele é em todos os sentidos um homem,”) penso que estaríamos justificados em considerá-lo como dizendo, dentre outras coisas, que ela atingiu uma boa altura, ao menos para sua espécie.

Quando uma conotação se torna tão padronizada para certos tipos de contexto, ela pode ser adquirir um novo status, no qual ela se torna uma condição necessária para ser aplicada à palavra nesse contexto. Ela então passa a constituir um novo sentido padrão. Esse terceiro estágio é ilustrado por meio de metáfora morta: “rabo”, usada em relação a luzes de automóveis, atualmente nada tem a ver com rabos de animais, e o seu significado pode ser aprendido por alguém que nunca ouviu que animais têm rabos. Certamente que nem todas as conotações passam por esse estágio, mas algumas sempre o farão.

Talvez alguma parte dessa história possa ser descoberta em palavras como “caloroso” e “duro” que são tiradas do campo sensorial e aplicadas à personalidade humana – como acontece aparentemente em muitas línguas.<sup>22</sup> Penso que a primeira aplicação de “caloroso” a uma pessoa transformou algumas propriedades acidentais de coisas quentes em parte de um novo significado da palavra, embora agora pensemos facilmente nessas propriedades como conotações de “caloroso” – por exemplo, afável, aprazível no contato, acolhedor. Essas qualidades eram parte do âmbito de conotação potencial de “caloroso” mesmo antes delas terem sido observadas em objetos quentes, o que pode não ter ocorrido até terem sido notadas nas pessoas e até que alguém, pensando numa palavra que descrevesse metaforicamente essas pessoas, tivesse encontrado “caloroso.” Mas antes que essas qualidades pudessem vir a pertencer à conotação fixada de “caloroso,” foi preciso descobrir aquilo que elas poderiam *significar* na palavra quando usada em uma metáfora apropriada. Finalmente, apesar disso ainda não ter acontecido, “pessoa calorosa” pode vir a perder sua característica metafórica com a conotação presente de “caloroso,” transformando-se em uma nova designação. Ela, então, passaria a ser uma metáfora morta.

Se a Teoria Revisada da Oposição Verbal estiver correta, ela daria conta de muita coisa. Ela faz melhor do que a teoria simples ao explicar a notável extensão a qual a metáfora pode expandir nosso repertório verbal além dos recursos da linguagem literal.<sup>23</sup> Ela permite

22 - Estas metáforas de personalidade têm sido investigadas de modo bastante interessante por Solomon E. Asch, "On The Use of Metaphor in the Description of Persons," in H. Werner, ed., *On Expressive Language*, Worcester: Clark University Press, 1955, e "The Metaphor: A Psychological Inquiry," in R. Tagiuri and L. Petrullo, eds., *Person Perception and Interpersonal Behavior*, Stanford: Stanford University Press, 1958. Veja também Roger Brown, *Words and Things*, Glencoe, III.: Free Press, 1958, pp. 145 – 54.

23 - Este é o modo como interpreto o poema de Wallace Stevens, "The Motive for Metaphor" (*Collected Poems*, New York: Knopf, 1955, p. 286): metáfora nos permite descrever, estabelecer e preservar, as sutilezas da experiência e altera ("a metade das cores de um quarto de coisas" na primavera), ao passo que as palavras em suas designações padrões no dicionário possam apenas lidar com

A influência das cores primárias do meio dia,  
 O A B C de ser,  
 O temperamento ruborizado, o martelo  
 De vermelho e azul...

novidade, mudança de significado, e mesmo mudança radical. Ela admite a imprevisibilidade da metáfora, as idéias surpreendentes que possam surgir mesmo devido a justaposições casuais das palavras. Ela mostra que uma metáfora pode ser objetivamente explicada, pois as propriedades das coisas e as conotações das palavras são publicamente detectáveis, e as controvérsias sobre ela são em princípio solucionáveis. Ela também explica a relativa obscuridade, ou confusão momentânea da metáfora da Classe II, que pode levar tempo para ser entendida completamente.

## V

Parece-me que, em sua forma revisada, a Teoria da Oposição Verbal pode percorrer um longo caminho em direção a uma explicação satisfatória da metáfora – se ela puder ser defendida contra duas possíveis linhas de objeções que são sugeridas por desenvolvimentos recentes.

A primeira objeção pode ser levantada por aqueles que estão comprometidos como uma teoria extensionalista do significado como oposta a uma teoria intensionalista. A Teoria da Oposição Verbal não pode ser formulada sem que se fale de propriedades (isto é, qualidades e relações) que são incompatíveis umas com as outras, mas o extensionalista não acredita que há tais coisas como propriedades. Não poderíamos, poderia ele perguntar, passar sem o conceito de incompatibilidade e tratar as metáforas simplesmente como um caso especial de enunciados materialmente *falsos*? É claro que há uma diferença entre dizer que alguém é careca quando não é, e dizer que é um leão quando não é. Mas talvez a diferença seja simplesmente que a última é mais surpreendente, mais evidente e certamente mais reconhecida como falsa. Vemos como o falante pode cometer um erro acerca da calvície, mas não vemos como ele poderia confundir um homem com um leão, e por isso é a clara improbabilidade da última observação (à luz do conhecimento comum) que nos faz rejeitá-la literalmente e tomá-la metaforicamente – ao invés de qualquer oposição interna do significado.

Poderíamos construir um argumento a favor dessa Teoria da Improbabilidade da metáfora, e podemos ainda sustentá-la por meio de exemplos de certos casos de metáforas degeneradas que poderiam ser analisadas dessa forma. Por exemplo, o piadista diz, “Eu estive na Filadélfia uma vez, mas ela estava fechada.”<sup>24</sup> Isso é realmente auto-contraditório? Parece-me completamente correto dizer que novas metáforas alargam nossos recursos lingüístico, ainda que elas não “expandam o significado” no sentido estrito objetado por J. Szrednicki, “On Metaphor,” *The Philosophical Quarterly*, X (July 1960): pp. 228 – 37.

24 - Outros exemplos teriam sido dados por Kenneth Burke, “Semantic and Poetic Meaning,” in *The Philosophy of Literary Form*, Baton Rouge: Louisiana University Press, 1941, p. 144 - “A cidade de Nova Iorque está em Iowa” pode significar que a influência de Nova Iorque se estende, como suas vias férreas, até o oeste.

É verdade que a palavra “fechado” é normal, e muito apropriadamente, aplicada a empreendimentos individuais, como lojas e museus, que possuem portas que podem ser fechadas e trancadas. Mas talvez sem estender demais o termo, uma cidade inteira também poderia ser literalmente fechada. Suponhamos que assim é. Nesse caso, o efeito metafórico peculiar – a difamação da vitalidade da vida noturna da Filadélfia – tem de depender da nossa rejeição do enunciado como falso como fora de mão, pois ele é absurdamente improvável. Ainda que se conceda que essa manobra verbal ocorra, ela não cobre todos os casos. No extremo oposto está oximoro. Um crítico no *Reporter* há um tempo atrás descreveu as figuras literárias do Movimento Beat como “escritores que não escrevem o que escrevem.” Isso não é apenas improvável. E me parece que as metáforas, em sua maioria, têm alguma coisa dessa auto-controvérsia embutida, completamente distinguível da piada da Filadélfia. Certamente deve haver casos de fronteira, em que há uma possibilidade não tão remota de o modificador ser considerado de uma maneira que literalmente será aplicado ao sujeito: por exemplo, a expressão “recozido pelo gelo” em *A Tempestade* de Shakespeare (I.ii,256), onde “recozido” poderia significar “castigado,” e assim a expressão completa poderia ter sido literal na época de Shakespeare.<sup>25</sup>

A segunda objeção à Teoria da Oposição-verbal poderia ser posta deste modo: ainda que haja propriedades que sejam opostas, elas não serão, em linguagem comum, fixadas na designação dos termos gerais de modo que ocorram contradições nítidas e claras. Aceitase que “irmão” e “irmão do sexo masculino” possam ser praticamente sinônimos perfeitos no tocante a seus significados centrais (ignorando suas conotações), e, por isso, “irmão do sexo feminino” é internamente contraditório – apesar de não ser bem uma metáfora, é claro. Mas a tese é que para grande parte das palavras que interessam, as regras não são bem definidas, e por isso quando essas palavras são usadas metaforicamente não pode ser porque detectamos uma incompatibilidade de significado no nível da designação.

O Professor Michael Scriven<sup>26</sup> argumentou que a palavra “limão” não tem, na verdade, quaisquer propriedades definidoras no sentido tradicional – isto é, propriedades que *teriam de* estar presentes no caso da palavra estar sendo corretamente aplicada a um objeto. Ele cita a definição do dicionário Webster, “a fruta ácida de uma árvore (*citrus limonia*), da mesma família da laranja,” e isso não parece dar as condições necessárias para algo ser um limão, já que não seria uma contradição dizer que um limão cresceu de uma bananeira, ou de nenhuma árvore. Scriven, porém, vai adiante, e afirma que não há uma única propriedade

25 - Veja o interessante artigo de Allan Gilbert, “Shakespeare Amazing Words,” *Kenyon Review*, XI (Summer 1949): 484 – 88, e Andrew Schiller, “Shakespeare’s Amazing Words,” *Kenyon Review*, XI (Winter 1949): pp. 43 – 49.

26 - “Definitions, Explanations, and Theories,” in Herbert Feigl, Michael Scriven, and Grover Maxwell, eds., *Minnesota Studies in the Philosophy of Science*, Vol.II: *Concepts, Theories, and the Mind-Body Problem*, Minneapolis: University of Minnesota Press, 1958, pp.105 – 7.

de limões que seja individualmente necessária, embora muitas outras estejam presentes. Ele também sustenta que o mesmo vale para muitos termos gerais de uso comum. Eles designam aquilo que ele chama de “conceitos agregativos,” e que embora tenha “critérios” de aplicação, não têm condições necessárias.

Essa importante idéia, caso possa ser sustentada, exigiria alguma reformulação na Teoria da Oposição Verbal que apresentei anteriormente. Isso não destruiria a teoria por pressupor que se a teoria é verdadeira, então a palavra “limão” não pode ser usada metaforicamente – como evidentemente pode. Scriven fala do significado literal como tendo “uma fronteira móvel para além da qual apenas o mau uso e metáfora se encontram.”<sup>27</sup> Se “limão” não tem condições necessárias, então não pode ser posta em um contexto verbal em que alguma condição necessária é logicamente excluída, embora possa ser posta em um contexto no qual muitos de seus critérios são excluídos de modo que não pode ser literalmente aplicada – como quando um carro usado vem a ser um limão.

Não estou convencido de que “limão” e outras palavras comuns *não* possuam condições necessárias, e o Sr. Scriven, presentemente, sustenta sua primeira posição somente em uma forma modificada. Penso, por exemplo, que ter certa textura orgânica – ao invés de ser feito de madeira ou cera – seria uma condição necessária para algo ser um limão. Seguramente, ser um objeto material é uma condição necessária – um “limão espiritual” ou não seria literalmente um limão ou não seria literalmente espiritual. As questões envolvidas aqui são sutis, bastante sutis para a ocasião. Por exemplo, se de repente eu estivesse com um objeto que cresceu exatamente como um limão, porém com seis palmos de diâmetro, suponho que eu poderia ser persuadido a chamá-lo de um “limão” gigante – na verdade, não me convenci disso. Isso mostra que neste momento uso a propriedade de ser pequeno apenas como um “critério”, mas não como uma propriedade definidora, de limão? Talvez sim – no entanto, se alguém diz que um objeto é um limão *sem* incluir qualquer comentário sobre seu tamanho incomum, estou justificado, penso, em deduzir que ele é pequeno. Talvez pudéssemos seguir uma sugestão do finado Arthur Pap<sup>28</sup> e outros que tenham discutido a “textura aberta”<sup>29</sup> de termos empíricos, e ponderar o critério como mais ou como menos exigido: distinguindo “graus de significado.” Poderíamos, então, identificar um modificador metafórico como uma expressão colocada em um contexto no qual uma de suas condições mais rigorosamente necessária é excluída. Ainda que o tamanho pequeno não seja uma propriedade indispensável de limões, ela pode ser bastante central, e nesse caso um contexto que fosse oposto a essa propriedade seria o

27 - *Ibid.*, p.119.

28 - *Semantics and Necessary Truth*, New Haven: Yale University Press, 1958, p. 327.

29 - Friedrich Waismann, “Verifiability,” *Proceedings of Aristotelian Society*, Supplementary Vol. XIX, London, 1945, pp. 119 – 50. Cf. Georg Henrik Von Wright, *A Treatise on Induction and Probability*, London: Routledge and Kegan Paul, 1951, ch. 6, §2, and Pap, *op. cit.*, chs. 5, 11.

bastante para conduzir a palavra a um estado metafórico.

Deixo em aberto a questão, dando-me por satisfeito no momento caso tenha mostrado que a Teoria da Oposição Verbal não só explica muito bem diversas características reconhecidas das metáforas, como também não faz suposições que uma filosofia da linguagem sólida estaria relutante em aceitar.

