LINGUASAGEM

ISAAC ASIMOV E A TEORIA DA OTIMIDADE

Gabriel de Ávila Othero

Lendo uma vez um artigo que fazia críticas à Teoria da Otimidade (o texto de Rennison, 2000), julguei bem interessante a seguinte crítica: “a OT é uma teoria que pode ser aplicada a qualquer coisa e a todas as coisas” (p. 140). Essa “crítica” me parece ser a mais pura verdade. Quem começa a estudar a Teoria da Otimidade logo percebe que alguns de seus princípios podem realmente ser aplicados a diversas esferas do conhecimento. Mais do que isso: os princípios da OT parecem ser aplicáveis a praticamente qualquer tipo de decisão que temos de tomar em nossas vidas. Esse foi o primeiro aspecto que me chamou a atenção para a teoria – e também foi o que achei mais interessante. Afinal, a OT trata sobre **como resolver situações de conflitos entre exigências diferentes**. E essas exigências não precisam necessariamente fazer parte apenas da gramática das línguas. Podemos aplicar a OT a vários outros domínios.

Por exemplo, creio que minha primeira aproximação à OT foi através de um texto de Elan Dresher (Dresher, 1996), em que ele conta uma anedota sobre alguns anciães judeus do século I d. C. que estavam debatendo sobre rituais judaicos. Sua história nos mostra como o “espírito” da OT pôde ajudar os velhos sábios judeus a resolver seus conflitos enquanto discutiam sobre quais rituais deveriam ou não ser executados durante o Sabbath.

Mais tarde, eu também li em um *blog* de Lingüística que a OT “não é apenas uma teoria de fonologia ou de sintaxe, mas também de filosofia de vida. A vida nos faz exigências conflitantes e, para satisfazer algumas, precisamos violar outras”. E esse, acredito, é um grande mérito da OT. Afinal de contas, não seria incrível se pudéssemos explicar fenômenos da linguagem seguindo os mesmos princípios e raciocínios que estamos acostumados a aplicar para resolver nossos problemas cotidianos?

A OT é justamente uma teoria que lida com **restrições** (ou princípios) que podem ser violadas se houver necessidade. Por exemplo, se um princípio A diz “durma oito horas por noite” e um princípio B diz “acorde às 7h para ir trabalhar”, eu defronto com três alternativas logicamente possíveis:

i. posso obedecer aos dois princípios e ir deitar às 23h para acordar às 7h;
ii. posso (ir a uma festa, voltar muito cansado, ir dormir às 2h e decidir) obedecer ao princípio A. Assim, durmo as minhas oito horas de sono, acordo às 10h e violo o princípio B;
iii. posso (ir a uma festa, voltar muito cansado, ir dormir às 2h e decidir) obedecer ao princípio B. Assim, acordo às 7h, durmo apenas cinco horas e violo o princípio A.

Tudo vai depender de qual princípio é o mais importante para a **hierarquia** em questão. Ou seja, devo organizar um **ranqueamento** das restrições e respeitar a mais importante. Pode ser **A >> B** (A é mais importante que B) ou **B >> A** (B é mais importante que A). Uma maneira de representar isso graficamente pode ser a seguinte:

Candidatos		**A** (ter oito horas de sono)	**B** (acordar às 7h)
i.	Deitarei às 23h e acordarei às 7h		
ii.	Deitarei às 2h e acordarei às 10h		*
iii.	Deitarei às 2h e acordarei às 7h	*	

Se eu julgar que as duas restrições estão em pé de igualdade (**A <<>> B**), então, o cenário expresso pelo candidato (i) é o melhor, e eu irei de dormir às 23h. No entanto, se for o caso de eu ter ido para a cama às 2h (como mostram os candidatos (ii) e (iii)), terei de fazer uma escolha: se eu achar que a restrição

A é mais importante, então o candidato (ii) será o melhor (ainda que eu tenha de violar algum outro princípio, como B); finalmente, se eu julgar que a restrição B é a mais importante, o candidato ótimo será (iii).

A moral da história é a seguinte: as restrições devem ser **ranqueadas**, e é possível **violar** uma restrição menos importante se isso for necessário para obedecer a alguma outra restrição **mais alta no ranking**.

Ora, isso parece fazer bastante sentido. Mas o que o Isaac Asimov do título tem a ver com tudo isso? Asimov criou as famosas **três leis da robótica**:

1) Um robô não pode prejudicar um ser humano, ou, por omissão, deixar um ser humano sofrer dano.
2) Um robô deve obedecer às ordens recebidas dos seres humanos, a menos que contradigam à primeira lei.
3) Um robô tem de proteger sua própria existência, desde que esta lei não entre em conflito com as duas primeiras leis.

À primeira vista, essas leis parecem triviais. Mas elas são, na verdade, muito bem elaboradas. Da maneira como estão expressas, elas logo remetem à "maneira OT de pensar". Vejamos por quê.

As três leis de Asimov poderiam ter sido expressas como regras lógicas, tais como (1a), (2a) e (3a) abaixo:

"x, x(robô), x deve
a) não prejudicar um ser humano;
b) obedecer às ordens recebidas dos seres humanos;
c) proteger sua própria existência.

As instruções acima parecem estar claras o suficiente para que uma máquina possa compreender. Mas imaginemos o seguinte cenário: eu tenho um robô e peço a ele que se jogue pela janela. O que o robô iria fazer? Obedeceria às minhas ordens, tal como estabelece (b), e se jogaria da janela? Ou não faria nada, por tentar proteger sua própria existência, seguindo (c)? E se eu mandasse o robô machucar alguém? Será que machucaria, obedecendo às minhas

ordens? Ou será que seguiria a regra (a) de não prejudicar um ser humano? E o que aconteceria se eu mandasse o robô me ferir?

Se as leis da robótica fossem implementadas da maneira como vemos em (a), (b) e (c), provavelmente o robô entraria em *looping* e não faria absolutamente nada nessas situações, pois tentaria obedecer a uma regra, mas seria barrado por outra. As regras são contraditórias entre si e poderiam acabar fundindo o cérebro positrônico do robô.

O que as leis de Asimov, da maneira como ele as formulou, têm de interessante é a parte em que dizem "obedeça a essa lei **a menos que entre em conflito com uma lei mais importante**". Ou seja, Asimov usou o mesmo tipo de raciocínio que usamos para resolver conflitos em OT. Entre as leis, há uma relação de dominância: **Lei 1 >> Lei 2 >> Lei 3**.

Por isso, seu eu pedir a um robô que se jogue pela janela, ele se jogará. Podemos representar isso com a seguinte tabela (ou ***tableau***). O que acontece se eu mandar o robô se jogar pela janela?

Candidatos		Lei 1	Lei 2	Lei 3
a.	F O robô se joga da janela			*
b.	O robô não se joga da janela		*	
c.	O robô *me* joga da janela	*	*	
d.	O robô não faz nada		*	

Na análise acima, percebemos que o robô não ficará confuso com as regras que lhe exigem ações contraditórias. Ele deverá selecionar o candidato **ótimo**, aquele que apresenta menos violações às regras, obedecendo ao ranqueamento de suas restrições. O candidato (a) é o ideal, pois viola apenas a terceira lei, a mais baixa entre as ranqueadas. Os candidatos (b) e (d) violam a segunda lei. O candidato (c) viola duas leis, a primeira e a segunda.

E o que aconteceria seu eu mandasse o robô machucar alguém?

Candidatos	Lei 1	Lei 2	Lei 3

a.	O robô se fere		*	*
b.	O robô fere outra pessoa	*		
c.	O robô *me* fere	*	*	
d.	F O robô não faz nada		*	

A resposta: o robô não faria nada. Ainda que essa atitude viole a segunda lei da robótica, os candidatos (b) e (c) violam uma lei mais importante, mais **alta no ranking**, e o candidato (a) viola duas leis.

Finalmente, o que aconteceria seu eu pedisse para o robô me ferir?

Candidatos	Lei 1	Lei 2	Lei 3
O robô se fere		*	*
O robô fere outra pessoa	*		
O robô me fere	*	*	
F O robô não faz nada		*	

Novamente, o robô não iria fazer nada. Essa é a única escolha lógica, em uma análise em OT. Programando os robôs dessa maneira, Asimov garantiu que suas máquinas fossem completamente inofensivas aos humanos. Assim, é logicamente impossível que um robô programado com as três leis da Robótica de Asimov faça mal a um ser humano. Mesmo antes de seu tempo, Asimov já expressava a “maneira OT de pensar”.

Por onde seguir? Há poucos materiais (até a data deste texto pelo menos) publicados em português sobre a Teoria da Otimidade. Um bom livro introdutório é Costa (2001). Alguns bons manuais introdutórios estão em inglês, como Archangeli & Langendoen (1997) e McCarthy (2002). Uma excelente fonte de material produzido em OT é o site do ROA (*Rutgers Optimality Archive*), em http://roa.rutgers.edu.

REFERÊNCIAS

ARCHANGELI, Diana; LANGENDOEN, D. Terence (eds.) *Optimality Theory: an overview*. Oxford: Blackwell, 1997.
COSTA, João. *Gramática, Conflitos e Violações: Introdução à Teoria da Optimidade*. Editorial Caminho, Lisboa, 2001.
DRESHER, Elan. The Rise of Optimality Theory in First Century Palestine. In *GLOT International* 2, 1/2, January/February 1996.
McCARTHY, John. *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press, 2002.
RENNISON, John. OT and TO: On the status of OT as a theory and a formalism. *The Linguistic Review* 17, 2-4, 2000.

Agradeço à Leda Bisol por comentários à primeira versão do texto e ao pessoal do Círculo de Estudos Lingüísticos da UFRGS pelos muitos debates interessantes.
Pós-doutorando em Lingüística pela Universidade Federal do Rio Grande do Sul – UFRGS. E-mail: gabriel_othero@terra.com.br .
Trecho original: “OT is a theory of anything and everything that it is applied to”.
Do *blog* Shadow, em www.garyfeng.com/wordpress/2006/02/21/the-rise-of-optimality-theory-in-first-century-palestine/. Cf. trecho original: “OT is not so much a theory of phonology or syntax as a philosophy of life. Life makes conflicting demands, and to satisfy some we must violate others”.
O asterisco significa uma violação ao princípio.
Essas leis ficaram famosas após o lançamento do filme *Eu, robô*, em 2004. Elas foram primeiramente publicadas por Asimov em seu conto *Círculo Vicioso*, de 1942.
Ou: “para todo x, se x é um robô, x deve”.
O símbolo F representa o candidato ótimo, a opção que será escolhida como a melhor pelo robô.
Mas o que aconteceu então com “a” maquiavélica robô do filme *Eu, robô*, que organiza uma revolução das máquinas contra os humanos? Ora, não posso revelar aqui... vá assistir ao filme!
Março de 2009.