A Flexibilização da Lógica em Direção a uma Melhor Modelagem da Mente pela IA

The Flexibility of Logic Toward a Best AI Modelling of Mind.

OLIVEIRA, Luis F.
Programa de Pós-graduação em Filosofia,
área de concentração em Ciência Cognitiva e Filosofia da Mente.
Faculdade de Filosofia e Ciências
UNESP – Campus de Marília
Av. Hygino Muzzi Filho, 737
Caixa Postal 420 – 17525-900
Marília – SP

luisfol@bol.com.br

ZAMPRONHA, Edson. S.
Departamento de Música
Instituto de Artes
UNESP – Campus de São Paulo
R. Dom Luis Lasagna, 400
CEP. 04266-030
São Paulo - SP - Brasil

zampra@osite.com.br

A Flexibilização da Lógica em Direção a uma Melhor Modelagem da Mente pela IA

The Flexibility of Logic Toward a Best AI Modelling of Mind.

Resumo:

A lógica multidimensional é uma modelagem lógica paraconsistente que pode simular características humanas que envolvem contradições. Este trabalho apresenta como a IA (inteligência artificial) pode ter seus procedimentos lógicos flexibilizados através desta lógica. Para tanto utiliza-se o modelo proposto por Gershenson. Na análise desta proposta mostra-se que um dos problemas fundamentais a serem superados, o da verificação da veracidade das premissas adotadas pela lógica, continua presente. Aponta-se o caminho para a superação deste problema no estabelecimento de relações motivadas, não arbitrárias, entre premissas adotadas pela lógica e aquilo que elas representam, e afirma-se que este é um dos caminhos para que a IA possa ser mais que uma simulação e possa se aproximar efetivamente a uma modelagem da mente humana.

Unitermos: Inteligência artificial; Lógica paraconsistente; Lógica multidimensional.

1. Lógica clássica e a inteligência artificial

O intuito de provar que o mecanicismo é condizente com a natureza dos fenômenos mentais ainda não foi satisfatoriamente alcançado pela Ciência

Cognitiva, principalmente pelo viés do computacionalismo clássico. Os modelos propostos pela Inteligência Artificial, tal como o computador serial digital desenvolvido por von Neumman, são implementações da Máquina de Turing. E esta, por sua vez, é uma função lógica. Estando o domínio computacional da Inteligência Artificial baseado na lógica clássica, tal domínio deve necessariamente apresentar as características típicas de um sistema formal para que possa ser devidamente implementado. Particularmente, ele deve estar livre de ambigüidades e contradições.

A lógica clássica possui uma relação muito próxima com a linguagem natural. No entanto, algumas características da linguagem natural não se adequadam a um procedimento formal. Por exemplo, o fato da linguagem natural ser permeada de contradições. Por essa razão Frege, fundador da lógica moderna, buscou a elaboração de uma linguagem artificial mais econômica e exata (sem ambigüidades) (WADLER, 2000). Segundo Feitosa e Paulovich (2001), um sistema formal deve apresentar:

(1) um conjunto qualquer de símbolos, ou alfabeto;

(2) um conjunto de expressões “bem formadas”;

(3) um conjunto de axiomas;

(4) um conjunto finito de regras.

Além disso, a lógica clássica (LC), a princípio, trabalha com dois valores: verdade e falsidade. Desse modo, um predicado pode ser verdadeiro ou falso, mas nunca simultaneamente verdadeiro e falso. Na LC não se está preocupado com o fato de uma expressão ser realmente uma verdade ou não para a ciência ou filosofia. Seus procedimentos funcionam independente desta veracidade. Em outras palavras, o que está sob o domínio dessa lógica são procedimentos formais que permitem partir de premissas e alcançar um resultado. A correspondência entre este resultado e algo externo à própria lógica não é uma questão que a LC se proponha.

A IA normalmente se baseia na LC para gerar um modelo do funcionamento da mente, e nesse sentido a Máquina de Turing é um modelo lógico abstrato da mente. Mas é possível perguntar até que ponto esse modelo é realmente adequado. Ou, ainda, quais aspectos da mente humana são evidenciados através deste modelo. A resposta a essas perguntas envolve não somente aspectos filosóficos, mas também computacionais e, nesse caso, lógicos.

2. A diferença entre humanos e máquinas no tratamento de problemas

Suponhamos um sistema artificial baseado no computacionalismo clássico o qual, portanto, opera a partir da lógica clássica. É possível ordenar a tal sistema a realização de uma determinada tarefa. Por exemplo, determinar quantos quilômetros um carro deve percorrer para ir de São Paulo a Tóquio. Tal proposta envolve sutilezas as quais, talvez, uma máquina ainda não possa computar. A princípio, a máquina calcularia a distância percorrida de maneira mais rápida e exata que um ser humano. Por outro lado, um ser humano poderia responder que não é possível ir de carro de São Paulo a Tóquio. Este exemplo, apesar de banal, nos mostra algumas divergências interessantes na forma de tratamento dos problemas por humanos e por máquinas.

Seres humanos costumam levar em conta a veracidade das premissas com que trabalham. Não que isso impeça a realização de um procedimento puramente formal, mas a verificação ou não da veracidade das premissas pode mudar significativamente a relação entre indivíduo e problema. Segundo a perspectiva de Vygotsky, comprovada empiricamente por Luria, existem distintas etapas para a solução de problemas por seres humanos (FRAWLEY, 2000). Em primeiro lugar, existe a localização do problema perante o universo histórico-social do indivíduo. O indivíduo cria uma estrutura conceitual (*frame*) que permite tratar os dados em questão, normalmente utilizando a linguagem como ferramenta de controle. É uma "questão de como o raciocício

não-monotônico, como a adição de informações influencia a situação comprobatória das conclusões, é capaz de inibir ou descartar opções inferências" (FRAWLEY, 2000, p.38-39). Em segundo lugar, o indivíduo realiza as atividades computacionais propriamente ditas. Ele faz o cálculo em questão através de um procedimento formal. A importância dessa descrição vygotskiana é colocar o contexto (externo) sócio-histórico na perspectiva da resolução do problema, e isto significa que antes de qualquer computação formal e lógica o indivíduo provavelmente irá verificar a pertinência e a veracidade das premissas apresentadas.

No caso de sistemas artificiais, como ocorre na IA, existe apenas a computação lógica. Um sistema artificial não está, a principio, apto a estabelecer relações sócio-históricas[1]. Ou seja, ele não tem a possibilidade de situar o problema em um contexto histórico-social individual. Mesmo assim, um sistema poderia até concluir que não é possível viajar de carro de São Paulo a Tóquio, mas tal solução seria o resultado de uma programação computacional mais completa, e não da verificação da pertinência das premissas.

É possível que um sistema artificial possa até simular, e de maneira eficiente, os procedimentos realizados por humanos. Porém isso não parece ser suficiente para explicar o funcionamento da mente, já que o computador

continua realizando apenas operações sintáticas, sem verificação da pertinência de suas premissas e conclusões. Esse modelo é, sim, uma boa ferramenta para a melhor compreensão da natureza dos processos mentais, um artefato que permite testar empiricamente hipóteses e teorias sobre a mente e que reproduz certas partes do seu funcionamento, em particular seu raciocínio lógico-formal, seu modo de funcionamento dedutivo. Mas, por outro lado, é difícil sustentar que este modelo seja possuidor de uma mente tal qual a mente humana devido aos limites da lógica clássica e da não consideração de outros tipos de raciocínio possíveis de serem realizados.

Enfim, por um lado a diferença consiste no fato de humanos verificarem a pertinência das premissas com que a lógica trabalha, sua relação com respeito ao que é exterior a ela mesma. Por outro lado, o problema é interno à própria lógica, limitada a uma lógica formal que não permite contradições nem ambigüidades. Nesse sentido, a utilização de uma lógica que permita uma maior proximidade com a mente humana e sua linguagem natural tem grande interesse. Em particular uma lógica que permita estados intermediários entre o verdadeiro e o falso, e mesmo que permita o aparecimento de contradições.

Um avanço já concretizado no *metier* científico é a utilização da lógica *fuzzy*[2]. A vantagem desta ferramenta é tornar possível a utilização de valores intermediários contínuos entre 0 e 1 (ou falso e verdadeiro). De certa forma,

com a lógica *fuzzy* já é possível conceber uma computação mais flexível, mais próxima da realidade da mente humana, e isso consiste certamente em um grande avanço Mesmo assim, ainda não se pode computar contradições, embora essa lógica permita operações que envolvem ambigüidades, vaguidez, imprecisões, ruídos e *inputs* incompletos.

3. A proposta da lógica multidimensional

Carlos Gershenson, em seu artigo *Modelling Emotions with Multidimensional Logic*, publicado em 1999, propõe um tipo de modelagem lógica que possibilita o uso de contradições por ser paraconsistente. Por lógica paraconsistente entende-se uma lógica na qual uma conclusão pode ser obtida a partir de premissas contraditórias.

A lógica proposta por Gershenson é multidimensional. Utiliza três operadores típicos da lógica *fuzzy*: *e*, *ou* e *negação*.

$$\begin{aligned} &\mu A \wedge \mu B = \min(\mu A, \mu B) \\ &\mu A \vee \mu B = \max(\mu A, \mu B) \\ &\neg \mu A = 1 - \mu A \end{aligned} \qquad (1)$$

Além desses, Gershenson introduz outros três: *equivalência*, *grau de contradição* e *projeção*. *Equivalência* é definida como (o autor utilizou um espaço bi-dimensional para as explicações):

$$E(x,y) = (1-y,1-x) \qquad (2)$$

Conforme a Figura 1, equivalentes são os pontos simétricos em relação à linha *fuzzy*, por exemplo (1, 1) e (0, 0) ou (0.25, 1) e (0, 0.75). Pontos sem contradição (ao longo da linha *fuzzy*) não têm equivalentes além deles mesmos.

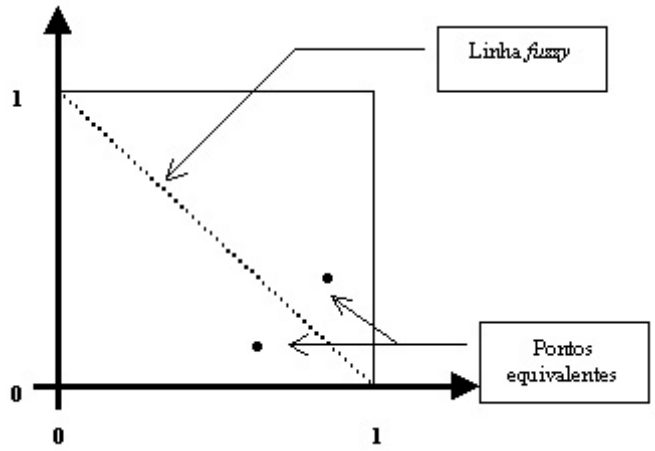

Figura 1. Pontos equivalentes.

Já a *Contradição* é definida através de um coeficiente que permite determinar quão contraditória é uma proposição. Esse coeficiente é definido através da seguinte equação 3:

$$C(x,y) = |(x+y)-1| \qquad (3)$$

Se C = 0, o ponto pertence à linha *fuzzy* e não apresenta contradição. Conforme o valor de C aumenta (ver Figura 2), mais o ponto se afasta da linha *fuzzy* e maior é a contradição, até o limite em que C = 1, o que indica uma total contradição.

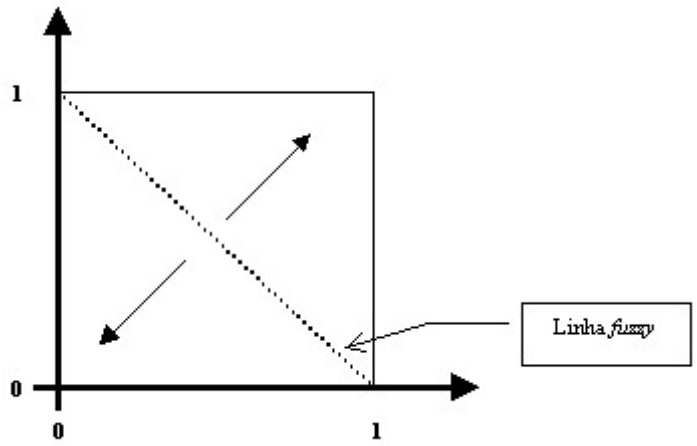

Figura 2. Contradição.

Por fim, a *projeção* obedece a seguinte equação 4:

$$P(x,y) = \left(\left(\frac{(x-y)+1}{2}\right),\left(\frac{(y-x)+1}{2}\right)\right) \qquad (4)$$

Todos os pontos pertencentes a uma linha perpendicular à linha *fuzzy* são representados, na projeção, por um único ponto. Isso permite que a

projeção que gera o ponto (0.5, 0.5) sobre a linha *fuzzy* possa ser o resultado de qualquer ponto pertencente à diagonal (0, 0)-(1, 1). Assim, esta projeção representa um ponto que pode ter como propriedade o fato de ser simultaneamente 100% branco e 100% preto, ou 0% branco e 0% preto, ou ainda qualquer estado intermediário nesta diagonal, tal como ilustrado na Figura 3.

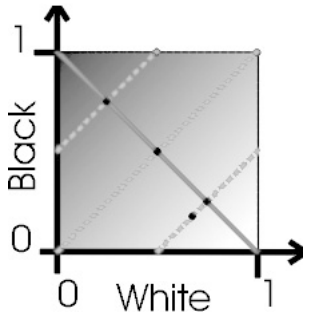

Figura 3. Projeção. (In: GERSHENSON, 1999, p.43)

Basicamente os operadores utilizados por Gershenson ampliam o domínio da lógica fuzzy, mas não alteram suas propriedades fundamentais. Ele utiliza esta lógica para a modelagem de emoções sobrepondo variáveis lógicas bidimensionais as quais representam pares de emoções. É possível utilizar quantos pares que se julgue necessário. Este modelo pode ser visualizado no seguinte gráfico tridimensional (Figura 4):

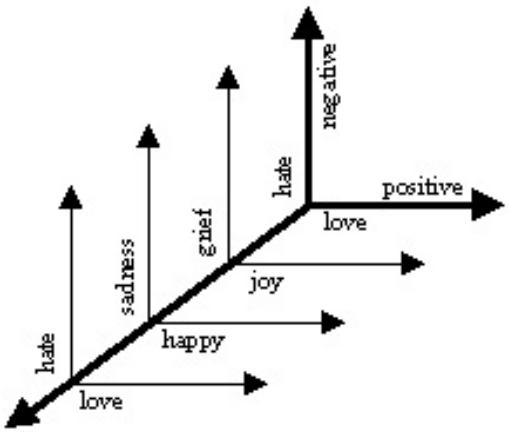

Figura 4. Modelo de emoções (In: GERSHENSON, 1999, p.44)

O eixo *x* apresenta emoções positivas, o *y* as negativas e, no *z* estão arbitrariamente situadas as emoções específicas. O par *love/hate* foi apresentado duas vezes sobre o eixo *z* para possibilitar tanto as relações entre *love/hate* e *joy/grief* e, ainda, entre *love/hate* e *happy/sadness*.

Desenvolvendo a idéia acima Gershenson amplia o quadro de emoções que seu sistema supostamente modela. Para isso ele determina uma tabela de emoções (Tabela 1) e suas respectivas posições no espaço tridimensional mostrado na Figura 4.

Emotion	***Vector***
Love	*Near* (1, 0, 0)
Hate	*Near* (0, 1, 0)
Indifference	*Near* (0.5, 0.5, 0)
Joy	*Near* (1, 0, 1)
Grief	*Near* (0, 1, 1)
Boredom	*Near* (0.5, 0.5, 1)
Happiness	*Near* (1, 0, 2)

Sadness	*Near* (0, 1, 2)
Conformity	*Near* (0.5, 0.5, 2)
Ecstasy	*Near* (1, 0.5, 0.3)
Rage	*Near* (0.4, 1, 1.5)
Pride	*Near* (1, 0, 1.5)
Shame	*Near* (0, 1, 1.5)
Anger	*Near* (0, 1, 2.5)
Jealousy	*Near* (1, 0.8, 0)
Lust	*Near* (0.8, 0.3, 2.7)
Sorrow	*Near* (0, 1, 0.5)
Revenge	*Near* (0.5, 1, 0.5)
Tenderness	*Near* (1, 0.2, 2.3)
Enthusiasm	*Near* (1, 0, 1.3)
Surprise	*Near* (0.8, 0.8, 2)

Tabela 1. Emoções representadas com o modelo. (In: GERSHENSON, 1999, p.44-45)

Determinado o ponto onde se situa cada emoção, Gershenson define um espaço tridimensional, uma esfera de raio arbitrário em torno de cada ponto. Assim, "se a distância de *A* para *X* é menor que ou igual a 0.1, *A* é considerado como pertencente à emoção *X*. Então, conforme a distância aumente, *A* é considerado ainda pertencente à emoção *X*, mas em menor grau, até que, finalmente, ele não tenha nenhum grau de *X*" (Gershenson, 1999). O conjunto de equações utilizado é o seguinte:

$$\begin{cases} 1 \rightarrow |X - A| \leq 0{,}1 \\ 1 - (10 \times (|X - A| - 0{,}1)) \rightarrow 0{,}1 < |X - A| < 0{,}2 \\ 0 \rightarrow 0{,}2 \leq |X - A| \end{cases} \qquad (5)$$

Gershenson considera, ainda, que os pontos da tabela que localizam especificamente as emoções não são determinados de maneira estrita. São valores aproximados. Segundo ele, a utilidade dessa imprecisão está na possibilidade, em uma simulação de comunidades de sistemas como uma sociedade artificial, de que cada indivíduo ou sistema tenha a possibilidade de apresentar um espaço emocional pessoal, com sutis diferenças nas relações entre os pares de emoções.

4. Discussão

A IA vem desenvolvendo de forma muito positiva seus modelos e métodos computacionais, chegando a resultados de muito interesse. A lógica *fuzzy* possibilita o tratamento de informações de maneira mais flexível. A interpolação de valores entre 0 e 1 permite a manipulação de dados ambíguos em ambiente computacional. A lógica paraconsistente expande ainda mais essas possibilidades, passando a incluir a contradição como uma propriedade tratável sob seu escopo. Porém, admitir-se que tais modelos explicam o funcionamento da mente é algo diferente, que requer maiores considerações.

A IA pode simular de maneira bem sucedida a mente, isto é, é possível conceber um sistema que gere o mesmo *output* que um humano perante um

determinado estímulo. Um sistema que passe no teste de Turing. Mas as colocações de Searle (1980) quanto ao aspecto semântico e à intencionalidade ainda permanecem válidas. Tais modelos, pelo seu aspecto pragmático, de fato funcionam, mas não esclarecem a natureza interna da mente, seus modos de operação, suas propriedades físicas e funcionais.

O modelo proposto por Gershenson para a modelagem das emoções revela um dos importantes problemas a serem resolvidos: a *arbitrariedade* existente entre a modelagem e aquilo que ela representa, que se traduz na veracidade das premissas utilizadas. No seu modelo há, por um lado, a colocação do par *Love/Hate* (Amor/Ódio) duas vezes no eixo *z*. Por outro lado, a emoção *Shame* (Vergonha), por exemplo, está representada no gráfico pelo ponto (0, 1, 1.5), o que significa que *Shame* é uma emoção negativa que está entre *Grief* (Dor, Pena) e *Sadness* (Tristeza). O que Gershenson está admitindo, implicitamente, é que há emoções primárias, mais básicas, como *Love/Hate* por exemplo, demarcadas no eixo *z*, e há emoções que são uma mescla destas emoções básicas, que é o caso de *Shame*. Já a colocação duas vezes do par *Love/Hate* para dar conta da representação de diferentes emoções revela ou uma inconsistência no modelo de representação, indicando que a representação tridimensional não é uma boa representação, ou que as emoções discriminadas no eixo *z* não são básicas. Emoções básicas funcionariam como

categorias, ou emoções prototípicas. Para tanto elas deveriam ser categorias motivadas, e não arbitrárias. A partir da identificação dessas categorias passaria a ser possível representá-las no eixo *z* e, posteriormente, passar a representar emoções mais complexas, como *Shame*, por exemplo. O problema, então, é que a relação entre premissas usadas na lógica e aquilo que elas representam não é motivada em seu modelo, e o fato dessas relações serem motivadas é fundamental para o estabelecimento de premissas verossíveis, o que termina por aproximar a forma de tratamento de problemas das máquinas à de humanos.

Mas, além disso, um entendimento não superficial das emoções a serem tratadas é fundamental, e aí parece haver uma certa superficialidade no seu modelo. *Indifference* (Indiferença), por exemplo, está representada pelo ponto (0.5, 0.5, 0). Ou seja, trata-se de um estado de *Love/Hate* que é meio positivo e meio negativo. No entanto *Indifference* não parece pertencer de modo exclusivo ao par *Love/Hate*. Poderia, ao contrário, pertencer a qualquer emoção do eixo *z*. Uma representação mais apropriada, portanto, poderia ser (0.5, 0.5, *z*), onde *z* representa qualquer emoção, o que introduz a possibilidade das emoções não serem somente representadas por pontos, mas também por segmentos de reta. Mas o estado de indiferença, se entendido como a não atribuição de um valor a algo, nem positivo nem negativo por não

ser nem bom nem mau dado o desinteresse implícito, seria um caso degenerado do gráfico no qual os eixos *x* e *y* são iguais a 0. Daí sua representação poder se transformar em (0, 0, *z*), isto é, sem qualquer emoção.

Já a emoção *Jealousy* (ciúme), entendida como o sentimento negativo que surge quando uma pessoa pensa que alguém está tentando tirar algo que ela julga lhe pertencer (um objeto ou uma pessoa), dificilmente coincide com a representação (1, 0.8, 0), ou seja, um sentimento negativo com 80% de dor e 20% de ódio. Se a pessoa está segura que aquilo que lhe pertence não pode ser tirado, por mais que haja uma tentativa a pessoa não sente ciúme. Ciúme, portanto, não é a reação emocional à *tentativa* de que algo lhe seja tirado, mas ao *medo*, a *dúvida* de que efetivamente a tentativa possa ter êxito. Trata-se, portanto, de um estado interno de *insegurança*, dificilmente representado por dor e ódio como componentes básicos.

A lógica multidimensional efetivamente amplia as possibilidades do cálculo lógico, mas não resolve o problema da arbitrariedade envolvida nas relações entre o modelo e aquilo que é representado pelo modelo[3] (veracidade das premissas). A identificação de emoções prototípicas, categorias básicas de emoções com base nas quais outras emoções podem ser mapeadas, resolveria por um lado o problema da representação no eixo *z*, já que essas emoções prototípicas estariam aí representadas de forma motivada, e não arbitrária.

Possivelmente ajudaria a definir melhor algumas emoções complexas, como é o caso de *Jealousy*, através dessas emoções mais básicas. Enfim, o problema da arbitrariedade entre as premissas adotadas pelo modelo e aquilo que ele representa permeia a IA, e esse parece ser um dos principais problemas a serem superados para que se diga que as suas realizações vão além de simulações da mente. E o caminho na direção do estabelecimento de relações motivadas parece ser uma direção frutífera no sentido de conectar as premissas com aquilo a que elas se referem, aproximando assim a IA da forma como procede a mente humana.

Abstract:
The multidimensional logic is a paraconsistent logic modelling suitable for the human features' simulations involving contradiction. This paper briefly present how AI (artificial intelligence) can make its logic procedures more flexible by multidimensional logic. The analysis of this logic modelling raise fundamental problems involving human mind's models, as the evaluation of truthfulness of logic premises. In this sense, alternative ways are pointed through the establishment of motivated rather than arbitrary relations between premises used in logic and something that are represented by them. We claim that motivated relations could be an interesting way toward a most satisfactory AI model of human mind, instead just a simulation of it.

Key-words: Artificial intelligence; paraconsistent logic, multidimentional logic.

Agradecimentos: Os autores são gratos ao Prof. Dr. Hércules Araújo Feitosa (Departamento de Matemática – FC – UNESP) pelo incetivo a esta

publicação, aos demais professores do programa de pós-graduação em ciência cognitiva e filosofia da mente (Departamento de Filosofia – FFC – UNESP), e também ao apoio financeiro fomentado pela FAPESP. Agradecemos especialmente a Carlos Gershenson pelo material que nos foi enviado.

Referências Bibliográficas:

DRESNER, E. Boolean algebra and natural language: a measurement theoretic approach. *Nordic Journal of Philosophical Logic*, v. 4, n. 2, p. 175–189, 2000.

FEITOSA, H. A. e PAULOVICH, L. *Um prelúdio à lógica* (apostila do curso Lógica Proposicional). Bauru, 2001.

FRAWLEY, W. *Vygotsky e a Ciência Cognitiva*. Porto Alegre: Artemed, 2000.

GERSHENSON, C. *Lógica multidimensional:* un modelo de lógica paraconsistente. In: Memorias del XI Congreso Nacional ANIEI, p. 132-141, 1998.

GERSHENSON, C. *Modelling emotions with multidimensional logic*. In: Proceedings 18th International Conference of the North American Fuzzy Information Processing Society (NAFIPS'99), p. 42-46, 1999.

YEN, J. *Fuzzy logic: a modern perspective*. In: IEEE Transactions on Knowledge and Data Engineering, v. 11, n. 1, 1998.

SEARLE, J. Minds, brains, and programs. *The behavioural and brain sciences*, v.3, p. 417-424, 1980.

WADLER, P. *Proofs are Programs: 19th Century Logic and 21st Century Computing*. 2000 ‹http://www.research.avayalabs.com/user/wadler/topics/history.html›. Acesso em:03 de dezembro de 2002.

ZADEH, L. A. Fuzzy sets. *Information and Control,* v. 8, p. 338-53, 1965.

ZADEH, L. A. *Toward a theory of fuzzy systems*. In: KALMAN, R. E. e DeCLARIS, N. (Eds). Aspects of network and system theory. New York: Holt, Rinehart and Winston, p. 469-490, 1971.

ZIEMKE, T. e SHARKEY, N. E. A stroll through the words of robots and animals: applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiótica* 134-1/4, p. 701-746, 2001.

[1] Muito se tem pesquisado atualmente sobre interações ambientais e sociais entre sistemas como comunidades de organismos artificiais. Nesse sentido confira, para uma visão geral, Ziemke e Sharkey (2001).
[2] Para uma visão geral sobre lógica *fuzzy*, confira Zadeh (1965 e 1971) e Yen (1998).
[3] Nesse sentido, exite uma interessante discussão sobre a relação representacional entre proposições e conectivos lógicos da linguagem natural em Dresner (2000).