



UNIVERSIDADE FEDERAL DA BAHIA

DISSERTAÇÃO DE MESTRADO

**Análise de Representações em um Experimento de Emergência
de Comunicação**

Emerson Silva de Oliveira

**Programa de Multi-institucional de Pós-Graduação em Ciência
da Computação da UFBA-UEFS**

Feira de Santana
25 de fevereiro de 2015

EMERSON SILVA DE OLIVEIRA

**ANÁLISE DE REPRESENTAÇÕES EM UM EXPERIMENTO DE
EMERGÊNCIA DE COMUNICAÇÃO**

Esta Dissertação de Mestrado foi apresentada ao Programa Multi-institucional de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia e Universidade Estadual de Feira de Santana, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Prof. Dr. Angelo Conrado Loula

Feira de Santana
25 de fevereiro de 2015

Ficha catalográfica - Sistema de Bibliotecas da UFBA

Oliveira, Emerson Silva de

Análise de Representações em um Experimento de Emergência de Comunicação/ Emerson Silva de Oliveira– 2015.

141p.: il.

Orientador: Prof. Dr. Angelo Conrado Loula.

Dissertação (Mestrado)– Universidade Federal da Bahia, Instituto de Matemática, Universidade Estadual de Feira de Santana, Feira de Santana, 2015.

1. Representação do conhecimento (Teoria da informação). 2. Redes neurais (Computação). 3. Simulação (Computadores digitais). 4. Semiótica. 5. Inteligência artificial. I. Loula, Angelo Conrado. II. Universidade Federal da Bahia. Instituto de Matemática. III. Universidade Estadual de Feira de Santana. IV Título.

CDD - 005.131

TERMO DE APROVAÇÃO

EMERSON SILVA DE OLIVEIRA

ANÁLISE DE REPRESENTAÇÕES EM UM EXPERIMENTO DE EMERGÊNCIA DE COMUNICAÇÃO

Esta Dissertação de Mestrado foi julgada adequada à obtenção do título de Mestre em Ciência da Computação e aprovada em sua forma final pelo Programa Multi-institucional de Pós-Graduação em Ciência da Computação da Universidade Federal da Bahia e Universidade Estadual de Feira de Santana.

Feira de Santana, 25 de fevereiro de 2015

Prof. Dr. Angelo Conrado Loula (orientador)
Universidade Estadual de Feira de Santana

Prof. Dr. Matheus Giovanni Pires
Universidade Estadual de Feira de Santana

Prof. Dr. Ricardo Ribeiro Gudwin
Universidade Estadual de Campinas

Aos meus pais, Salvador e Vercelina.

AGRADECIMENTOS

Várias pessoas contribuíram, de diversas maneiras, para que esse trabalho fosse realizado. Deixo aqui alguns agradecimentos.

Ao meu orientador Angelo Loula, por ter me apresentado essa área de pesquisa tão instigante e desafiadora, pela confiança depositada para realização desse trabalho, pelas discussões e contribuições decisivas ao longo dessa pesquisa.

Ao meu pai, à minha mãe, e ao meu irmão, que sempre incentivaram meus estudos, e nunca mediram esforços para me apoiar.

Ao SERPRO, que através do programa de incentivo à educação pós-graduada me concedeu liberação de horas de trabalho para que eu pudesse me dedicar ao mestrado. A participação nesse programa seria impossível sem as grandes contribuições de Regina Lúcia, Wilson e Izabel.

Ainda no SERPRO, aos meus colegas de setor, Cleverson, Danilo, Ednara, Eudes, Luciano, Thiago e Wilson, que além do incentivo, sempre estavam dispostos a mudar datas e horários de trabalhos para que minha rotina de estudos não fosse prejudicada.

Aos colegas de mestrado, Jairo, Matheus, Rafael e Thiago, pelo incentivo, cobranças, conversas e discussões ao longo de todo esse tempo.

À Luisa, que chegou em minha vida no decorrer desse trabalho e, com amor e compreensão, tornou esse caminho melhor.

Aos amigos que ajudaram direta ou indiretamente, e que sempre se fizeram presentes.

"For millions of years, mankind lived just like the animals. Then something happened which unleashed the power of our imagination. We learned to talk and we learned to listen."

—STEPHEN WILLIAM HAWKING

RESUMO

A noção de representação é um assunto bastante discutido na comunidade de Inteligência Artificial desde o seu surgimento, mas até hoje ainda não há consenso sobre questões básicas, como sua própria definição. Por outro lado, diversos trabalhos envolvendo emergência de comunicação em agentes artificiais, nos quais representações podem estar envolvidas, vêm sendo desenvolvidos, principalmente nas últimas duas décadas, motivados por avanços científicos e tecnológicos em diversas áreas de conhecimento (como linguística, biologia evolucionária e ciências cognitivas). Entretanto, pouco (ou quase nada) é analisado e discutido em relação aos possíveis processos de representação envolvidos nos sistemas de comunicação emergentes. Acreditamos que essa é uma questão importante, que pode ajudar a entender melhor os mecanismos envolvidos em processos cognitivos relacionados à linguagem e comunicação, e que pode contribuir nos avanços nas pesquisas da área. Nesse trabalho replicamos um experimento que aborda emergência de comunicação em uma população de agentes artificiais, e procuramos, baseados na teoria semiótica de Peirce, a existência de representações. O principal objetivo é investigar processos de representação através da análise da ativação de redes neurais que controlam criaturas artificiais em um experimento de emergência de comunicação, identificando a representação, seu referente e seu efeito nestas criaturas. Os resultados demonstram a emergência de representações simbólicas nos processos comunicativos e uma relação direta destas com o grau de adaptação da população.

Palavras-chave: Representação, Redes Neurais, Simulação Computacional, Emergência de Comunicação, Semiótica, Inteligência Artificial.

ABSTRACT

The notion of representation is a widely discussed subject in Artificial Intelligence community since its beginning, but until present days there's no agreement about basic questions, as its own definition. On the other hand, several works approaching the emergence of communication on artificial creatures, in which representations could be involved, have been developed, especially in the last two decades, motivated by scientific and technological progress in various knowledge areas (such as, linguistic, evolutionary biology and cognitive sciences). However, few (or almost nothing) are analyzed and discussed about possible processes of representation occurring in the emerging communication systems. We believe that this is an important question, which can contribute to a better understanding about the mechanisms involved in cognitive processes related to language and communication and, therefore, assist advances in researches on this area. In this work we replicate an experiment which deals with the emergence of communication in a population of artificial agents, and we seek, based on Peirce's semiotic theory, for the existence of representations. The main objective is to investigate representations processes by analyzing the activation of neural networks that control artificial creatures in an emergency of communication experiment, identifying the representation, its referent and its effect on these creatures. The results show the emergence of symbolic representation on the communicative processes and a direct relation between this representations and the adaptation level of the population.

Keywords: Representation, Neural Networks, Computer Simulation, Semiotic, Emergence of Communication, Artificial Intelligence.

SUMÁRIO

Capítulo 1—Introdução	1
1.1 Motivações	3
1.2 Objetivos	4
1.3 Organização da dissertação	5
Capítulo 2—Representação	7
2.1 IA Clássica	8
2.2 Conexionismo	10
2.3 Nova IA	12
2.4 Problema da Fundamentação do Símbolo	17
2.5 Mas afinal, o que é representação?	24
2.6 Semiótica de Charles S. Peirce	28
2.6.1 Ícone, Índice e Símbolo	30
Capítulo 3—Estudos de linguagem, comunicação, e simulações computacionais	35
3.1 Simulação computacional e estudo de linguagem e comunicação	36
3.1.1 Tipos de simulação	38
3.1.2 Simulações em emergência de comunicação	39
3.2 Trabalhos relacionados	43
Capítulo 4—Replicação do experimento	57
4.1 Escolha do experimento replicado	57
4.2 Descrição do experimento original	58
4.2.1 Motivações	59
4.2.2 Cenário do experimento	60
4.2.3 Análise dos resultados encontrados por Mirolli e Parisi	63
4.3 Experimento analisado sob a perspectiva da semiótica de Peirce	69
4.3.1 Comparação de ativações da camada intermediária da rede neural	71
4.3.2 Outras medidas	76
4.4 Replicação e modificações propostas no experimento original	78
4.4.1 Arquitetura modificada	78
4.4.2 Criaturas geneticamente semelhantes	80
4.4.3 Interação e seleção por grupo	80
4.5 Testes estatísticos	81

Capítulo 5—Resultados e análises dos processos de representação	83
5.1 Réplica do experimento	83
5.1.1 Relação signo-objeto	86
5.1.2 <i>Fitness</i> e relação signo-objeto	89
5.1.3 Medidas de correlação e variância	93
5.2 Falantes e ouvintes com similaridade genética	104
5.3 Seleção por grupo	110
5.4 Arquitetura sem representação	117
5.5 Considerações sobre os processos de representação nas simulações	120
Capítulo 6—Conclusão	125
6.1 Contribuições	128
6.2 Trabalhos Futuros	130

LISTA DE FIGURAS

2.1	Modelo de execução do argumento do quarto chinês. Fonte: alexandria.nu	18
2.2	Modelo da relação triádica do modelo semiótico de Peirce, nó borromeano, adaptado de (QUEIROZ, 2004).	29
3.1	Detalhes sobre robôs e cenário utilizados no experimento de (FLOREANO et al., 2007).	45
3.2	Modelo que ilustra as quatro configurações utilizadas nesse experimento de (FLOREANO et al., 2007). $r = 1$ indica que os indivíduos dos grupos são relacionados, e $r = 0$ indica que os indivíduos dos grupos não são relacionados.	46
3.3	Modelo da rede neural utilizada no experimento de (MAROCCO; NOLFI, 2007).	48
3.4	Esquema do experimento de (MAROCCO; NOLFI, 2007), com as duas áreas-alvo, e quatro agentes na arena por vez.	49
3.5	Detalhes da arena e do robô utilizados no experimento de (GREEFF; NOLFI, 2010).	50
3.6	Modelo da rede neural utilizada no experimento de (GREEFF; NOLFI, 2010).	50
3.7	Esquema do experimento em (MIROLI, 2012), com objetos de dois tipos caindo, e o agente podendo se mover lateralmente.	51
3.8	Modelo da rede neural utilizada no experimento de (MIROLI, 2012).	52
3.9	Modelo da arquitetura cognitiva utilizada no experimento de (LOULA; GUDWIN; QUEIROZ, 2013).	54
4.1	Modelo esquemático do experimento.	61
4.2	Arquitetura da rede neural que controla os agentes do experimento.	61
4.3	A geração dos novos indivíduos de cada geração é realizado em duas fases. Primeiro dois indivíduos são escolhidos para reprodução, e tem seus cromossomos cruzados, com duplo ponto de corte. Depois é aplicado o mecanismo de mutação, que pode alterar para um valor aleatório qualquer gene, como probabilidade de 0.1%.	64
4.4	Resultados alcançados por Mirolli e Parisi em (MIROLI; PARISI, 2008). As linhas cinza são os resultados do <i>fitness</i> médio da população, e as linhas pretas são os resultados de qualidade de comunicação média da população.	66
4.5	Arquitetura que evita o efeito da categorização visual na comunicação em (MIROLI; PARISI, 2008).	68

4.6	Ativações geradas a partir do recebimento de sinais comunicativos no grupo de neurônios da entrada comunicativa.	72
4.7	Ativações geradas no grupo de neurônios da entrada visual a partir da visualização de cogumelos.	72
4.8	Ilustração do cálculo de similaridade entre uma ativação originada por entrada comunicativa e ativações originadas por entrada visual, para o mesmo tipo de cogumelo.	74
4.9	Ilustração do cálculo de similaridade entre uma ativação produzida por entrada comunicativa e ativações originadas a partir entrada visual, para tipos de cogumelo diferentes.	74
4.10	Gráfico com os pontos gerados pelas ativações da camada intermediária da rede neural da criatura com maior <i>fitness</i> em uma geração.	77
4.11	Arquitetura modificada proposta, que pretende dificultar a utilização de signos.	79
5.1	Amostra representativa do <i>fitness</i> e qualidade da comunicação das 40 simulações realizadas com a replicação do experimento.	84
5.2	Amostra representativa da curva de variância genética da população, divididos por grupos de pesos. EC, NC, e SC se referem, respectivamente, aos grupos de pesos da Entrada Comunicativa, pesos Não-Comunicativos, e pesos da Saída Comunicativa.	85
5.3	Amostra representativa da curva do deslocamento dos centroides dos pesos entre duas gerações na população.	85
5.4	Amostra representativa dos sinais emitidos pela população.	86
5.5	Valores médios de R_p (medida de relação signo para o objeto presente) e R_a (medida de relação signo para o objeto ausente) durante uma simulação.	87
5.6	Valores médios da diferença entre R_p e R_a durante uma simulação.	89
5.7	Gráfico de dispersão do <i>fitness</i> e relação do signo com o objeto presente no ambiente (R_p), em uma execução do experimento.	91
5.8	Gráfico de dispersão do <i>fitness</i> e relação do signo com o objeto ausente do ambiente (R_a), em uma execução do experimento.	91
5.9	Gráfico de dispersão do <i>fitness</i> e diferença entre a medida de relação do signo com o objeto presente no ambiente e a relação do signo com o objeto ausente do ambiente ($R_p - R_a$), em uma execução do experimento.	92
5.10	Variância média da relação signo-objeto R_p (Var. Presente) e R_a (Var. Ausente).	94
5.11	Gráfico que apresenta <i>fitness</i> , diferença entre as relações signo-objeto com objeto presente e com ausente ($R_p - R_a$), e a médias das variâncias dessas medidas (Var). Está dividido em seis intervalos, os quais possuem características de relações diferentes entre essas três variáveis.	95
5.12	Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração do trecho 1.	97
5.13	Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração do trecho 3.	99

5.14	Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração do trecho 4.	101
5.15	Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração do trecho 5.	101
5.16	Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração do trecho 6.	103
5.17	Exemplo de curva de <i>fitness</i> obtidas nas execuções da simulação com agentes ouvinte e falante geneticamente similares.	105
5.18	Exemplos de duas execuções da simulação com interação entre agentes geneticamente similares, nos quais há grande diferença na média dos valores da diferença $R_p - R_a$	106
5.19	Gráficos que mostram a correlação praticamente inexistente entre as medidas de relação signo-objeto e o <i>fitness</i>	107
5.20	Exemplo de curva de variância das relações signo-objeto R_p e R_a em execução de simulação com interação entre agentes geneticamente similares.	108
5.21	Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração durante uma execução do experimento com interação entre agentes geneticamente similares.	109
5.22	Exemplo de curva de <i>fitness</i> obtidas nas execuções da simulação com interação e seleção por grupo.	111
5.23	Exemplo de curva de medidas de relação signo-objeto com objeto presente (R_p) e com objeto ausente (R_a), em uma execução do experimento com grupos.	113
5.24	Gráfico de dispersão do <i>fitness</i> e relação signo-objeto com o objeto presente em uma execução do experimento com interação e seleção por grupo. . .	113
5.25	Variância das medidas de relação do signo com o objeto presente e ausente, durante uma execução da simulação com interação e seleção por grupo. .	114
5.26	Gráficos que mostram a tendência de queda na variância genética da população.	116
5.27	Exemplo de curva de <i>fitness</i> em uma execução de simulação com a arquitetura da rede neural modificada.	119
5.28	Variâncias dos sinais emitidos com a configuração original e com a arquitetura modificada.	120

LISTA DE TABELAS

5.1	Tabela com dados calculados em intervalos distintos de gerações de uma execução.	96
-----	--	----

Capítulo

1

INTRODUÇÃO

A história dos computadores modernos e da ciência da computação é bastante recente, se comparada com outras áreas da ciência como a física e a biologia, por exemplo. Mesmo assim, nesse relativo curto espaço de tempo, tanto as máquinas quanto as técnicas utilizadas se desenvolveram de tal modo que os computadores se tornaram muito populares, e hoje estão inseridos de algum modo em praticamente todas as áreas de atuação humana: da produção industrial à pesquisa acadêmica, passando por comunicação e entretenimento. Tal popularidade se deve, em boa parte, à grande flexibilidade que os computadores possuem, uma vez que podem ser programados para executarem tarefas das mais diversas, e com restrições específicas a cada domínio no qual atuará. Entretanto, existem alguns domínios para os quais os computadores e as técnicas existentes parecem ainda não ser desenvolvidos o suficiente para dar um tratamento adequado. Esse é particularmente o caso da linguagem. Questões sobre como os computadores podem lidar com linguagem, como podem se comunicar de forma autônoma (com humanos ou outras máquinas), e se podem se comunicar através de linguagem natural como fazem os humanos, são questões que acompanham algumas áreas de conhecimento, incluindo a Inteligência Artificial (IA), praticamente desde o início da história dos computadores modernos, e para as quais ainda não existem respostas consolidadas, e.g. (BOBROW, 1964; HARNAD, 1990; ANDERSEN, 1997; REITER; DALE; FENG, 2000).

A Inteligência Artificial (IA) surge em meados da década de 1950, definindo um novo sub-campo de pesquisa da ciência da computação, envolvendo o estudo e o desenvolvimento de máquinas e programas inteligentes (MCCARTHY et al., 2006). É dessa época o famoso Teste de Turing. Nesse teste um juiz humano conversa em linguagem natural com um humano e com uma máquina, através de um terminal, sem saber previamente quem está conversando com ele. A máquina é criada com o propósito de ter comportamento idêntico ao humano e, caso o juiz não consiga distinguir entre o humano e a máquina, a máquina passa no teste (TURING, 1950). Apesar de algumas críticas sofridas pelo Teste de Turing (FRENCH, 2012), e da vasta discussão sobre o que seria realmente inteligência, o fato é que até os dias atuais ainda há muito interesse em criar máquinas e/ou programas que sejam capazes de passar por esse teste.

O Teste de Turing não diz respeito apenas à linguagem e comunicação, mas à inteligência humana como um todo. Essa é uma das características que marcaram as primeiras décadas da IA. Estavam lançando esforços em processos cognitivos de alto nível, como raciocínio, planejamento e linguagem, e seus objetivos e expectativas eram de que em pouco tempo fossem capazes de desenvolver máquinas e programas que replicassem completamente a inteligência humana. Exemplo desse grande otimismo é uma declaração de Simon, em 1965, afirmando que dentro de 20 anos as máquinas seriam capazes de realizar qualquer trabalho realizado por um ser humano (SIMON, 1965). Esse pensamento de replicação de inteligência de alto nível é uma das características da IA Clássica, que predominou como abordagem utilizada entre os primeiros pesquisadores da IA.

Outra característica marcante dos estudos sobre cognição, e que refletem diretamente nos estudos da IA, foi a adoção da chamada abordagem simbólica para modelar sistemas inteligentes. Segundo esse modelo, sistemas inteligentes seriam aqueles capazes de armazenar conhecimento através de representações simbólicas, e de manipular esses símbolos através de algum algoritmo (BEER, 2014). Pelo fato de ter adotado uma abordagem baseada em símbolos para tentar construir sistemas com inteligência de alto nível, a IA Clássica é também conhecida como IA Simbólica (STEELS, 2003a). A formulação da Hipótese dos Sistemas de Símbolos Físicos, em 1976 por Newell e Simon é considerado o maior marco dessa abordagem. Segundo essa hipótese, sistemas capazes de manipular símbolos e fazer inferências sobre eles teriam o necessário e suficiente para toda e qualquer ação inteligente (NEWELL; SIMON, 1976). Essa afirmação iguala manipulação sintática de representações simbólicas à inteligência. Tal formulação reflete bem o pensamento mais difundido na época, segundo o qual, símbolos dentro de máquinas que fossem capazes de manipulá-los seriam o caminho para a ação inteligente.

Várias críticas à abordagem utilizada pela IA Simbólica surgiram então, com destaque para duas específicas. Uma dessas críticas afirma que situações reais são tipicamente complexas, aleatórias e imediatas, e questionam a possibilidade dos sistemas simbólicos lidarem com essas situações, como afirmam Agre e Chapman em (AGRE; CHAPMAN, 1987). Outra aponta justamente para o fato de que os símbolos eram inseridos e interpretados por operadores humanos, sem nada significar para as próprias máquinas que os manipulavam. Esse último problema foi nomeado por Harnad como O Problema da Fundamentação do Símbolo (HARNAD, 1990), e questiona como estabelecer para a máquina uma relação entre o símbolo, e aquilo que ele representa.

Para tentar superar esses problemas, surgem novas abordagens, dentre elas o conexionismo e a chamada *Nouvelle AI* (Nova IA). A característica principal do conexionismo é a computação paralela realizada em nós de uma rede. Ela não nega a existência de representações, mas tem um entendimento diferente da IA Simbólica, afirmando que as representações estão nos padrões de conexões de uma rede, e não em cada nó separadamente (CHALMERS, 1992; SMOLENSKY, 1988). Na Nova IA, as principais ideias, em oposição a IA Clássica, consistem de que ações inteligentes mais complexas devem emergir de ações mais simples, em interação do agente artificial com o ambiente, inclusive com outros agentes, defendendo a ideia de um agente situado, incorporado e em interação com seu ambiente (WAGNER et al., 2003; BROOKS; STEIN, 1994a; BROOKS, 1991a; ZIEMKE, 1999).

Além disso, percebeu-se que o objetivo de replicar a inteligência humana de forma completa em um curto espaço de tempo era irreal, considerando-se o tamanho e a complexidade da tarefa, que foi sendo melhor compreendida ao longo dos anos. Por isso, atualmente pouco ou nada é proposto no sentido de replicar a inteligência ou o comportamento humano de forma completa, e os problemas relacionados são geralmente divididos em sub-problemas bem definidos (como representação do conhecimento, processamento de linguagem natural, aprendizado, planejamento, navegação, percepção e inteligência social), e tratados de maneira separada (BROOKS, 1995).

1.1 MOTIVAÇÕES

Vários experimentos têm sido realizados, tanto em robôs físicos quanto em robôs e agentes simulados, em todas essas sub-áreas relacionadas acima. Dentre esses experimentos, vêm ganhando destaque os experimentos que simulam a emergência de linguagem e comunicação entre agentes artificiais (STEELS, 2003b; NOLFI; MIROLI, 2010; WAGNER et al., 2003). As motivações para a realização desses experimentos tem diversas naturezas, as quais conseguimos dividir em dois grupos de motivações: (a) pretendem ajudar a entender como surgiu a linguagem e os primeiros sistemas de comunicação humana, e como nós lidamos com ela, um desafio científico bastante importante, desde que a linguagem e a forma como os humanos se comunicam é algo que nos distingue dos outros animais, além de ser algo fortemente relacionado ao conceito de inteligência (LIEBERMAN, 1998); (b) pretendem gerar conhecimento tecnológico suficiente para que possamos construir sistemas artificiais que possam lidar apropriadamente com linguagem (STEELS, 1997). Nesse caso, seria possível criar novas tecnologias que explorariam esse recurso das mais diversas formas, além de beneficiar várias tecnologias já existentes. Dentre estas tecnologias estão engenhos de busca, sistemas de armazenamento e recuperação de informação (como bancos de dados), tradutores automáticos, comunicação entre agentes artificiais e interações homem-máquina em linguagem natural.

Além disso, experimentos sintéticos vêm se consolidando como uma abordagem científica para que pesquisadores de várias áreas de conhecimento validem seus testes e teorias, podendo também ser utilizados como um laboratório experimental, no qual é possível modificar parâmetros e verificar as mudanças nos resultados (BEDAU, 1998; STEELS, 1997). Em particular, experimentos sintéticos podem oferecer contribuição para várias áreas interessadas na compreensão da linguagem, e estão sendo bastante utilizados, por exemplo, em estudos sobre a origem e evolução da linguagem, fornecendo novos métodos para que os pesquisadores possam realizar testes e experimentos que seriam impossíveis de serem realizados de maneira real (CANGELOSI; PARISI, 2002).

Porém, praticamente nenhum dos autores dos trabalhos sobre emergência de comunicação tem se dedicado ao estudo de um fenômeno importante, e que tem forte relação com comunicação: processos de representação. Na comunicação algum sinal é transmitido de um indivíduo para outro, e aquele que recebe o sinal deve interpretá-lo de algum modo, e possivelmente tomar alguma ação. Se o sinal transmitido consegue representar, em algum aspecto, alguma outra coisa, e o indivíduo que o recebe consegue interpretar isso, e estabelecer essa conexão, há um processo de representação ocorrendo (discutire-

mos mais essa definição na seção 2.6). Por exemplo, se alguém grita “fogo!” em uma sala fechada, e todos os presentes imediatamente deixam o local, podemos admitir que os receptores da informação conseguiram associar a palavra “fogo” com o elemento fogo e, a partir dessa interpretação tomaram uma ação, a de sair do local de perigo. Assim, houve a ocorrência de processos de representação nesse cenário. Dessa forma, acreditamos que a noção de representação é uma noção muito importante nesse contexto, que pode ajudar a entender melhor os mecanismos envolvidos em processos cognitivos relacionados à linguagem e comunicação, e que pode contribuir avanços nas pesquisas da área.

Por sua vez, os avanços científicos que as pesquisas sobre representações podem trazer, podem ajudar no desenvolvimento de tecnologias que facilitariam muitas tarefas importantes, para as quais algumas instituições importantes já se pronunciaram a respeito. Por exemplo, o DARPA, Agência de Projetos de Pesquisa Avançada de Defesa (*Defense Advanced Research Projects Agency*), órgão de pesquisa ligado ao departamento de defesa norte americano, vem lançando sucessivas chamadas e programas na tentativa de desenvolver tecnologias possivelmente relacionadas ao uso de representações por computadores. Em 2008 o DARPA lançou uma chamada intitulada: “Aprendizado Computacional de Representações Linguísticas do Mundo Físico”. Nessa chamada, podemos destacar o interesse principal por novas propostas para superar as dificuldades existentes em lidar com linguagem em computadores, já que, mesmo que algumas máquinas já tenham superado o poder de processamento do cérebro humano, ainda há um dificuldade muito grande para que computadores associem construções linguísticas com objetos, conceitos e ações no mundo físico (DARPA, 2008). Em 2011 foi lançada a chamada para o projeto BOLT, Tradutor de Idioma de Ampla Operação (*Broad Operational Language Translation*). O objetivo desse projeto é desenvolver uma tecnologia capaz de traduzir vários idiomas, recuperar informações a partir do material traduzido, e permitir a comunicação bilíngue do forma oral ou escrita. Um dos objetivos específicos diz que deve haver “pesquisa em aquisição da linguagem semântica profunda, usando visão robótica e informações táteis como entrada para a aprendizagem experiencial de objetos, ações e consequências” (tradução nossa) (DARPA, 2011). E em 2012 foi lançada a chamada para o projeto DEFT, Exploração e Filtragem Profundas de Texto (*Deep Exploration and Filtering of Text*), que tem o objetivo de desenvolver uma tecnologia para profunda compreensão de linguagem natural de forma automatizada, que deve procurar pelo significado de textos, podendo fazer uso de suas informações mais relevantes (DARPA, 2012).

1.2 OBJETIVOS

Como comentamos na seção anterior, na comunicação existe algo que é transmitido de um agente falante (produtor de algum tipo de sinal, não necessariamente fala humana) para um agente ouvinte (intérprete de algum tipo de sinal, não necessariamente sonoro), e aquilo que é recebido pelo ouvinte pode produzir algum efeito nele. É possível então que o sinal recebido pelo ouvinte seja interpretado como uma representação de uma outra coisa para este? E se de fato for, que tipo de representação seria? Essas questões, no entanto, são raramente abordadas nos trabalhos que apresentam experimentos computacionais que tratam de comunicação e linguagem. Assim, os processos representativos que podem

estar ocorrendo nos agentes, ou o que de fato seriam para o sistema os sinais emitidos pelos agentes, praticamente não são discutidos.

Entendemos que para avançar nas pesquisas sobre comunicação e linguagem com agentes artificiais e ajudar no avanço científico e tecnológico da área, é essencial ter uma melhor compreensão do que ocorre durante estes experimentos. Nesse sentido, entender os processos representativos que podem estar ocorrendo nesses experimentos pode ser um aspecto importante para esse objetivo. No presente trabalho, pretendemos estudar e analisar os processos de representação em um experimento de emergência de comunicação em agentes artificiais já realizado previamente por outro autor, nos propondo assim a contribuir e a dialogar com essa comunidade, podendo complementar as análises e descobertas já realizadas por outros pesquisadores, e desenvolver uma metodologia que possa ser também utilizada pela comunidade ao analisar seus experimentos.

Para realizar esse objetivo, dois pontos devem estar bem definidos. Um deles é em relação ao conceito de representação, e o outro diz respeito aos critérios adotados para seleção dos experimentos que serão replicados e analisados. A temática da representação é bastante recorrente na IA desde sua origem, gerando discussões importantes. Ainda, representação, como discutiremos no capítulo 2, não é um termo plenamente definido no domínio da computação, e seu significado pode variar a depender do autor e/ou do contexto no qual está sendo utilizado. Por outro lado, a semiótica, que utiliza o termo técnico signo para designar representação, procura definir e classificar os tipos de signo existentes (SEBEOK, 2001). Nesse trabalho o conceito de representação é central, e devemos defini-lo de modo claro e suficientemente abrangente. Aqui iremos utilizar como base para essa definição a teoria dos signos de Charles Sanders Peirce.

Para selecionar os experimentos que serão replicados e avaliados, devem abordar a emergência de comunicação em comunidades de agentes artificiais, de forma que nem todos os aspectos do sistema de comunicação estejam pré-definido, mas que haja emergência de comunicação, seja através de aprendizado, de evolução ou de alguma outra técnica utilizada. De fato, quanto menor a quantidade de pré-definições no experimento, melhor poderemos avaliar os processos envolvidos na emergência da comunicação.

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

- **Representação:** Nesse capítulo faremos uma discussão teórica sobre representação. Iniciaremos com uma retrospectiva do uso do termo e de seu lugar de destaque em algumas abordagens e momentos da IA. Após essa discussão apresentaremos pontos relevantes da teoria semiótica de C. S. Peirce para esse trabalho, procurando mostrar sua definição de signo, assim como sua classificação básica de ícone, índice e símbolo procurando deixar claro o que entenderemos por representação no restante da dissertação.
- **Estudos de linguagem, comunicação, e simulações computacionais:** Aqui mostraremos como simulações computacionais (experimentos sintéticos) estão sendo utilizados em pesquisas sobre comunicação e linguagem. Inicialmente mostraremos porque várias áreas do conhecimento estão utilizando experimentos sintéticos para auxiliar seus estudos. Depois de apresentar uma breve classificação desses experi-

mentos, introduziremos os experimentos que abordam a emergência da comunicação em agentes artificiais, apresentando suas principais características e técnicas utilizadas. Finalizamos o capítulo com os trabalhos relacionados à essa dissertação.

- **Replicação do experimento:** Nesse capítulo discutimos os critérios para a escolha do experimento a ser replicado, e descrevemos o experimento escolhido em detalhes. Em seguida, discutimos sobre como realizar as avaliações de modo que possamos responder às questões de pesquisa, propomos o cálculo de algumas medidas, e modificações sobre o experimento original, de forma que nos ajude a fazer a avaliação dos processos de representação, e as condições favoráveis e contrárias à ocorrência dessa natureza de processo.
- **Resultados:** Um conjunto de condições e configurações são avaliados, permitindo uma ampla discussão sobre as questões levantadas nesse trabalho. Esse capítulo apresenta então os resultados, as discussões, e possíveis conclusões a respeito das condições e configurações adotadas, e pretende estabelecer conexão entre a teoria apresentada e os resultados encontrados, procurando assim responder às principais questões levantadas no presente trabalho: se houve de fato algum tipo de representação no processo comunicativo analisado e qual seria esse tipo.
- **Conclusão:** Finalizamos a dissertação avaliando quais diálogos e contribuições nosso trabalho pode fomentar na comunidade de pesquisadores, e também para os avanços científicos e tecnológicos da área. E ainda serão apresentadas as expectativas de trabalhos futuros, tais como novos experimentos a serem realizados, revisão de metodologia e novas investigações a serem realizadas.

REPRESENTAÇÃO

Representação é um tópico de grande interesse para a ciência da computação, e em particular para a IA, desde seu início. Muitas abordagens e concepções de comportamento inteligente utilizados na IA e nas ciências cognitivas têm como um dos principais pontos de discussão aspectos relacionados à representação, seja sobre sua natureza, sobre a possibilidade e a forma com que programas e máquinas possam atribuir significado a representações, ou mesmo sobre sua real necessidade para o comportamento inteligente. Harvey (1992), por exemplo, afirma que o conceito de representação está no cerne da ciência da computação. No entanto, o uso do termo representação na computação é muitas vezes impreciso e não padronizado. É relativamente comum que um pesquisador utilize o termo se referindo a algo diferente do que refere um outro pesquisador (de outra área, ou até da mesma área) ao utilizar esse mesmo termo. Além disso, existem ainda variações e supostos sinônimos, tais como: símbolo, sinal, palavra, signo e token; que contribuem ainda mais para estabelecer a confusão de significados quando se fala em representação.

Beer (2014), falando sobre sistemas dinâmicos e cognição incorporada no contexto da IA, afirma que, infelizmente, apesar do papel fundamental que a noção de representação desempenha nas abordagens computacionais sobre sistemas inteligentes, não há consenso sobre sua definição. Esse problema é particularmente acentuado em discussões que abordam os temas de linguagem e/ou comunicação (CANGELOSI; GRECO; HAR-NAD, 2002). Uma discussão ampla sobre esse importante tema da computação e das ciências cognitivas se faz necessária para que se tenha melhor compreensão de como a representação é utilizada na IA e na computação em geral, atentando para os problemas da falta de padronização na utilização do termo, na dificuldade que os pesquisadores encontram ao tentar estabelecer esse padrão, e para os impactos negativos que essa indefinição gera para as pesquisas acadêmicas na área (GLENBERG; VEGA; GRAESSER, 2008; HARVEY, 1992; MIROLI, 2012). Porém, apesar da necessária discussão, há ainda um certo receio sobre os resultados práticos que uma tentativa de definição do termo teria. Apesar da demanda de definição precisa do termo “representação”, tal definição

pode “abrir a caixa de Pandora de divergências” (tradução nossa) (GLENBERG; VEGA; GRAESSER, 2008).

Essa discussão está também fundamentalmente relacionada aos objetivos deste trabalho. Por isso, neste capítulo faremos uma discussão sobre alguns aspectos da representação dentro do contexto da IA e das ciências cognitivas. Apresentaremos algumas das abordagens e perspectivas mais importantes utilizadas nessas áreas, mostrando a forma como essas diferentes abordagens tratam representações em agentes artificiais, e os principais desafios que os pesquisadores encontram nesses *frameworks*. Conhecer esses *frameworks* é importante, pois eles influenciam de várias maneiras a forma como se investiga determinado fenômeno, moldando inclusive a linguagem utilizada para descrevê-lo (BEER, 2014). Ao fazer um breve histórico, veremos mudanças radicais de paradigmas na forma de entender e lidar com representações. Começaremos apresentando a IA Clássica e sua “Hipótese dos Sistemas de Símbolos Físicos”. Novas abordagens sobre diversos temas relacionados a IA surgem para tentar dar um tratamento adequado aos problemas apontados na IA Clássica. Dentre eles tem destaque o Conexionismo, que apresenta uma outra visão para a cognição, e o que muitos chamam de *Nouvelle IA*, que apresenta uma visão que pretende romper de forma muito mais forte com os conceitos da IA Clássica, questionando inclusive a necessidade de representações para a síntese de sistemas inteligentes. Em seguida exporemos ainda o “Problema da Fundamentação do Símbolo”, uma das questões mais relevantes nessa área da IA. Após esse apanhado sobre a utilização e abordagens acerca do uso de representações, buscaremos uma definição precisa de representação na semiótica, mais especificamente, nos apoiaremos no trabalho do cientista Charles Sanders Peirce, e apresentaremos algumas definições e ideias pertinentes para esse trabalho defendendo sua utilização.

2.1 IA CLÁSSICA

Por volta de 1960 o pensamento dominante nas ciências cognitivas e nas pesquisas da IA foi o computacionalismo. Esse modelo tem a perspectiva de que o funcionamento da mente humana é muito parecida com o funcionamento de programas de computador, e esse modelo foi adotado como abordagem para o estudo e a síntese de sistemas inteligentes. Segundo esse modelo, os sistemas artificiais inteligentes podiam armazenar e codificar conhecimento através de representações simbólicas, e essas representações simbólicas poderiam ser manipuladas através de algum algoritmo, de forma que fossem produzidas as soluções esperadas do sistema. O foco desse modelo está na computação feita com as representações simbólicas. Nesse cenário, o substrato utilizado para a computação não é importante, tanto faz um cérebro animal ou um computador pessoal, o importante são as computações realizadas.

A primeira escola a utilizar esses princípios é hoje conhecida como IA Clássica. Por ter como pilar o “conhecimento”, que por sua vez é realizado através de “representações simbólicas”, essa abordagem também é chamada de IA Simbólica, ou modelo baseado em conhecimento (STEELS, 2003a). Essa abordagem nasce praticamente junto com a IA, e inicialmente há grande entusiasmo com esse novo campo de pesquisa. Exemplo dessa empolgação é a previsão feita por Simon em 1965, afirmando que dali a 20 anos os

cientistas seriam capazes de construir máquinas com habilidades para realizar qualquer tarefa desempenhada por humanos (SIMON, 1965).

O grande marco da IA Clássica é a Hipótese dos Sistemas de Símbolos Físicos, enunciada por Newell e Simon em 1976, quando fazem a categórica afirmação de que sistemas de símbolos físicos (representações simbólicas) teriam o necessário e suficiente para toda e qualquer ação inteligente (NEWELL; SIMON, 1976). Um sistema de símbolos físicos pode ser definido como “um conjunto de entidades, chamados de símbolos, que são padrões físicos que podem ocorrer como componentes de um outro tipo de entidade chamada de expressão”. Tal estrutura simbólica pode ser composta por vários símbolos, que podem estar relacionados entre si por algum padrão físico. E, além dessa estrutura, o sistema também tem um conjunto de processos que, operando sobre uma expressão, deriva outra expressão. Esses processos seriam unicamente baseados em regras sintáticas (considerando apenas a forma dos símbolos, não um possível significado semântico). Esse conjunto de representações simbólicas e esse conjunto de processos formariam o conhecimento de determinado agente artificial.

Segundo esses autores, as capacidades que o sistema teria de designação e interpretação das expressões, dariam a eles as propriedades de um sistema completo e fechado. Ou seja, um sistema constituído apenas de símbolos e que pudesse fazer inferências sobre eles, estaria apto a qualquer ação inteligente. Nesse mesmo trabalho, Newell e Simon afirmam ainda que os símbolos estão na raiz da ação inteligente (NEWELL; SIMON, 1976), enfatizando a grande importância que o símbolo teria para que fosse possível realizar a síntese de um sistema capaz de realizar ações humanas. A fim de que as manipulações sobre tais símbolos e expressões sejam significativas, é necessário que elas possam representar estados do mundo externo. Newell e Simon (1976) afirmam que esse relacionamento é uma propriedade básica dos sistemas de símbolos físicos, e chamam essa propriedade de “designação”. Apesar de não discutirem como isso seria possível, definem que uma expressão designa um objetivo “se, dada a expressão, o sistema pode afetar o objeto e/ou se comportar de maneira dependente do objeto”. Sistemas inteligentes seriam então formados por uma categoria de programas que operassem algoritmos computacionais sobre tais símbolos.

Utilizando ideias e conceitos oriundos da IA Simbólica, alguns produtos vieram a ter sucesso acadêmico, e mesmo comercial. É o caso dos sistemas especialistas, que ainda hoje são utilizados em algumas áreas. Esses sistemas utilizam grande quantidade de dados e de regras que são inseridos por programadores, a partir do conhecimento de vários especialistas humanos e, utilizando as regras fazem inferências sobre os dados inseridos no sistema, e chegam a algum resultado, que em geral serve como auxílio para um operador humano resolver determinado problema. O MYCIN (COPELAND, 2015), é um exemplo desse tipo de sistema que teve sucesso.

Entretanto, as críticas a essa abordagem não demoraram a surgir, e uma das principais é: como tais sistemas, que fazem apenas manipulação sintática de símbolos de forma isolada, poderiam conectá-los ao mundo, estabelecendo algum tipo de relação entre algo do ambiente externo pela representação simbólica manipulada? Críticos questionaram o fato de que todo o conhecimento (supostamente representado pelos símbolos) era inserido por um operador externo dentro do sistema, assim como os resultados obtidos também de-

veriam ser interpretados por um agente externo, e afirmaram que para o próprio sistema, os símbolos manipulados nada significavam (SEARLE, 1980; HARNAD, 1990; ZIEMKE, 1999; CANGELOSI; GRECO; HARNAD, 2002; VOGT, 2002). Essa é a base para um conhecido problema da IA, o Problema da Fundamentação do Símbolo (ZIEMKE, 1999). Discutiremos mais essa questão na seção 2.4. Além disso, mesmo nos sistemas que fizeram sucesso, como os sistemas especialistas, haviam problemas graves. Primeiro, o fato de que todo o “conhecimento” do sistema ter que ser inserido a priori por um operador externo, e que, em geral, esses sistemas precisam de muitos dados para funcionar, limita bastante o escopo de ação de cada sistema. O MYCIN, por exemplo, é especialista em diagnóstico de doenças causadas por bactérias.

Esses problemas apontam para limitações existentes na abordagem da IA clássica para se alcançar os objetivos almejados nas primeiras décadas de pesquisas. A necessidade de uma grande quantidade de dados, aliada a dificuldade (ou impossibilidade) de relacionar apropriadamente os símbolos internos ao sistema ao que estes se referem, fazem com que os sistemas construídos com base nos conceitos da IA Simbólica ter grande dificuldade de responder adequadamente a mudanças de contextos, e dessa forma limitando bastante seu campo de atuação. Essas limitações podem levar a erros em situações não previstas, já que inferências sobre determinada situação dependem de pressuposições adequadas sobre a situação, e essa adequação pode ser altamente dependente do contexto (EMMECHE, 1991). Se os dados que seriam o conhecimento armazenado não conseguem mapear todos os cenários possíveis (o que é algo extremamente difícil, se não impossível), em alguma(s) situação(ões) o sistema sofrerá alguma falha decorrente desse problema.

A partir do início dos anos 1980 ganham relevância algumas outras propostas de abordagens para a construção de sistemas inteligentes e, particularmente, para a forma como as representações devem ser entendidas e “inseridas” no sistema. Duas dessas propostas obtiveram maior relevância para a comunidade de pesquisadores da IA, que são: o Conexionismo e a chamada Nova IA. Enquanto a primeira mantém certa suposição de que cognição pode ser realizada através da manipulação de certas entidades (mas com manipulações e entidades distintas da IA Clássica, como veremos), a segunda vai de encontro a quase todas as suposições da IA Clássica, propõe a utilização de sistemas situados e incorporados, atuando de forma dinâmica no mundo, e põe em questão até mesmo a necessidade de algo como representações para a formações de criaturas e/ou sistemas inteligentes.

2.2 CONEXIONISMO

Em (PICCININI, 2009) Piccinini alerta que o termo “conexionismo” pode ser bastante ambíguo, e gerar confusão quanto ao seu entendimento. Um dos motivos para essa confusão é a variedade de noção de computação utilizada pelos autores. Para termos uma ideia, não há consenso entre os conexionistas se seu modelo de cognição é computacionista ou não; enquanto alguns defendem que sim, outros defendem que não. Outro fator apontado por Piccinini é a utilização do termo com diferentes propósitos. Segundo ele, o sentido original do conexionismo diz que o comportamento é explicado pela mudança de conexões entre estímulos e respostas, biologicamente intermediadas pelas modificações das

conexões entre os neurônios. Porém, atualmente o sentido do termo parece ser bastante diferente. Percebemos a defesa de outras características para a modelagem conexionista da cognição: a utilização do modelo de redes neurais, cujo funcionamento é caracterizado por padrões de ativação entre simples unidades de processamento interconectadas (THOMAS; MCCLELLAND, 2008); o paralelismo dos processamentos (RUMELHART et al., 1986); e o processamento realizado sobre entidades chamadas de unidades sub-simbólicas (CHALMERS, 1992).

A computação nos sistemas conexionistas é realizada em seus nós (unidades de processamento relativamente simples) e nas conexões entre esses nós. Um modelo bastante utilizado, e tão identificado com a escola, que chega a ser confundido com o próprio conexionismo são as redes neurais artificiais (as vezes chamada de rede conexionista (ELMAN, 1993)). Nelas, os nós seriam o equivalente aos neurônios da rede, e as conexões equivalentes ao peso entre os diferentes neurônios associados entre si. Nessa perspectiva, a cognição não é modelada como a aplicação de uma série de regras sintáticas sobre entidades chamadas de símbolos, mas está modelada essencialmente como computações sobre os nós e suas conexões.

Esse modelo, que realiza computação em vários nós e conexões, está diretamente relacionado com outra importante característica do conexionismo, sua natureza inerentemente paralela, visto que é possível realizar as computações em vários nós ao mesmo tempo (HINTON; MCCLELLAND; RUMELHART, 1986). Além dos aspectos de modelagem, essas computações se diferenciam da IA Clássica por enfatizar aspectos estatísticos e de aprendizado da cognição (CORDESCHI; FRIXIONE, 2007). Com essas características do modelo de cognição, alguns conexionistas adotam o termo representação distribuída para se referirem a forma de representação utilizada pela escola.

A unidade de representação utilizada, de forma diferente das entidades atômicas da IA Clássica, consiste em um conjunto (ou padrão) de atividades sobre determinado número de nós e conexões (CHALMERS, 1992; SMOLENSKY, 1988; HINTON; MCCLELLAND; RUMELHART, 1986). O padrão de conectividade, que é a maneira como as unidades são interligadas, isto é, como estão estabelecidas as conexões entre as unidades e seus respectivos pesos, constituem o conhecimento do sistema, e determina a forma como este vai responder a entradas externas (HINTON; MCCLELLAND; RUMELHART, 1986). Thomas e McClelland afirmam ainda que nesse modelo, o “conhecimento é armazenado na força das conexões entre as unidades” (THOMAS; MCCLELLAND, 2008).

Os conexionistas utilizam então essa noção de representação distribuída, enquanto na IA Simbólica é defendido que o nível de abstração que deve ser utilizado nas computações deve ser o símbolo. Os conexionistas afirmam que o nível simbólico é um nível muito alto de abstração, e que é preferível trabalhar com entidades de menor granularidade. Lembremos que nesse contexto, o símbolo seria algo como uma entidade (muitas vezes também chamada de *token*) dentro do computador, que é utilizada para referenciar algo externo à máquina. Não há, no entanto, consenso sobre a granularidade da entidade a ser utilizada. Alguns acham que deveria ser o nível do neurônio, visto que a maioria dos modelos conexionistas utilizam o modelo matemático do neurônio como unidade básica, como é o caso das redes neurais artificiais. Já outros acham que esse nível é mais baixo que o necessário, e que seria melhor trabalhar com um nível intermediário, menor que

o simbólico, porém, mais alto que o neurônio. De forma genérica, podemos afirmar que os conexionistas propunham um modelo de cognição no qual a estrutura básica não era simbólica, mas sub-simbólica.

Dessa maneira, enquanto a IA Simbólica manipula símbolos supostamente com valores semânticos, os conexionistas estariam manipulando estruturas com o nível mais baixo que o semântico. Nesse contexto, afirmar que uma entidade tem valor semântico significa que é possível associá-la diretamente com algo no mundo externo. Smolensky (1988) afirma que os sub-símbolos não são manipulados pelas regras da manipulação do símbolo, mas são submetidos a computações numéricas, de natureza não simbólica; e enquanto uma operação no paradigma da IA Simbólica pode consistir em uma única computação, no paradigma sub-simbólico tal operação tomaria o resultado de um grande número de operações numéricas (principalmente de natureza estatística) de menor granularidade. Os conexionistas esperam que a partir das regras e operações de baixo nível (com menor granularidade que a concepção simbólica), propriedades semânticas possam emergir, mesmo sem terem sido explicitamente programadas (CHALMERS, 1992).

Embora o conexionismo tente dar uma nova perspectiva à área das ciências cognitivas, trazendo um modelo de cognição diferente do dominante até então, segundo alguns autores, ela continua mantendo uma visão centralizada e focada apenas na mente (RUTKOWSKA, 1995; HARVEY, 1992). Essas características fazem o conexionismo e a IA Clássica compartilharem alguns fundamentos básicos. Por exemplo, apesar de terem visões diferentes sobre sua natureza, ambas as abordagens concordam sobre a existência, grande importância, e sobre algumas características das representações (FODOR; PYLYSHYN, 1988). Assim, apesar de não ser a unidade fundamental de processamento, como o é na IA Simbólica, as representações possuem bastante importância para o conexionismo. Por outro lado, alguns aspectos fundamentais do conexionismo, como a utilização de redes neurais em seus processamentos, se popularizou bastante, e vem sendo largamente utilizado por pesquisadores das mais diversas áreas, incluindo pesquisadores adeptos das ideias defendidas pela abordagem da Nova IA, que abordaremos a seguir.

2.3 NOVA IA

Rompendo com os principais paradigmas das duas abordagens descritas até aqui, a Nova IA (ou *Nouvelle AI*) começa a surgir em meados dos anos 1980, e ganha mais força no início dos anos 1990. Essa abordagem traz uma série de novas perspectivas sobre o entendimento da inteligência, e de como sistemas artificiais devem ser projetados para se comportar de forma inteligente. Ela transfere o foco essencialmente cognitivo, baseado somente no que ocorre “entre as orelhas”, e passa a lidar com aspectos e questões solenemente ignoradas nas abordagens com viés computacionalista da cognição. Por exemplo, como o fato de ter um corpo interfere na cognição? Ou como a interação com o ambiente e contexto no qual o agente está inserido influenciam seu comportamento? A representação é realmente necessária para o comportamento inteligente? Esses são alguns dos principais aspectos dessa abordagem: a incorporação (o agente deve ter um corpo), a situacionalidade (o agente deve estar situado em um ambiente e interagir com este), e a rejeição à necessidade de representações (pelo menos da forma como ela foi definida

na IA Clássica, como veremos ainda nesta seção) (BROOKS, 1995). Por conta desses aspectos, em muitos textos encontramos nomenclaturas como incorporada (*embodied*), situada (*situated*), baseada em comportamento (*behavior-based*), dinâmica (*dynamics*) ou atuativa (*enactive*), para se referir (como um todo ou em algum aspecto) à abordagem que trataremos aqui pelo nome de Nova IA.

A necessidade de ter um corpo é oposta a toda abordagem computacionalista. Sistemas da IA Clássica são muitas vezes chamados por seus críticos de “cérebros-na-caixa”, em referência a completa ausência de importância dada a experiências corpóreas. Na Nova IA, os agentes devem ser capazes de sensoriar diretamente o mundo, agir nele, ser parte de sua dinâmica, e “sentir” os resultados de sua interação de forma imediata. A situacionalidade diz respeito aos agentes lidarem diretamente com o mundo, e não com descrições abstratas deste. O agente deve ter seu comportamento influenciado pelo “aqui” e “agora” (BROOKS, 1991a). Dessa forma, os agentes seriam capazes de ter suas próprias experiências e sensações (BROOKS, 1991b). Nessa mesma direção, Emmeche, em (EMMECHE, 2007), afirma que é necessário dar uma maior atenção ao fato de que “um robô pode viver em um mundo ‘perceptivo’ que difere daquele que o construtor do robô pode ver de forma imediata”. Esse parece ser um caminho para que os agentes tenham maior autonomia, tomem decisões e se comportem não de acordo com interpretações de agentes externos (e.g. um operador), mas sim de acordo com sua percepção de mundo.

Rodney Brooks foi um dos primeiros defensores dessa abordagem. Ele propõe a Hipótese da Fundamentação Física (Physical Grounding Hypothesis), sob a qual está baseada a Nova IA, em contraste à Hipótese dos Sistemas de Símbolos Físicos, na qual está fundamentada a IA Clássica. Segundo essa hipótese, “para construir um sistema inteligente é necessário que suas representações estejam fundamentadas no mundo físico”, o que faria a necessidade da representação simbólica utilizada pela IA Clássica desaparecer (BROOKS, 1990). Isso porque, de acordo com (BROOKS, 1991b), “O mundo é seu melhor modelo”, e, em lugar das representações simbólicas tradicionais, é suficiente sensoriar-lo com a frequência apropriada. Tais sistemas devem ser dotados de sensores e atuadores, e devem estabelecer ciclos de percepção e ação decorrentes da sua interação com o mundo físico. Outro princípio importante dessa hipótese é que os sistemas devem ser construídos de forma *bottom-up*, iniciando com ações bastante simples e de baixo nível, e gradualmente dando passos em direção ao aumento da complexidade das ações, com os níveis mais altos dependendo dos níveis mais baixos.

Brooks utiliza um argumento baseado na biologia evolutiva para defender a precedência de sistemas baseados na Hipótese de Fundamentação Física sobre sistemas que armazenam e manipulam representações. Antes que surgissem criaturas capazes de utilizar representações, como na comunicação através da linguagem por exemplo, habilidade adquirida recentemente em termos evolutivos, a natureza levou muito mais tempo para desenvolver as habilidades necessárias para sobrevivência e reprodução. Tais habilidades seriam essencialmente a locomoção em um ambiente dinâmico, e a capacidade de também sensoriar e agir apropriadamente no ambiente. Segundo Brooks, foi nessa parte da inteligência que a evolução concentrou maior parte do seu tempo, e é a base para o que ele e outros pesquisadores chamam de “inteligência verdadeira”. Essas características forneceriam o suporte sem o qual outros “tipos de inteligência” de nível mais alto (como

a utilização de símbolos) não poderiam surgir. Dessa forma, um sistema sem fundamentação física não estaria preparado para a utilização fundamentada de representações simbólicas, uma vez que a fundamentação física iria prover as condições necessárias para que as representações sejam de fato úteis para tais sistemas (BROOKS, 1991b).

Sendo então aspectos relevantes dessa proposta, o foco na ação, na noção de corpo do agente artificial e na sua interação com o ambiente, que deixa de lado ambições como a rápida construção de sistemas que imitem a inteligência humana e, por conseguinte retirando a ênfase de aspectos como raciocínio e planejamento, é de fato natural que surjam pontos de vista céticos em relação à necessidade de representação para a inteligência, como Brooks em (BROOKS, 1991b), por exemplo. O próprio nome do artigo de Brooks, “Inteligência sem Representação” (*Intelligence Without Representation*), já demonstra boa parte da visão do autor. Nesse trabalho, Brooks, seguindo a abordagem da Nova IA, pretende promover a inteligência de determinado agente robótico de forma incremental, e sem a utilização de representações internas. Ele argumenta que elementos como a mobilidade, visão e a habilidade de sobreviver em um ambiente dinâmico são as bases necessárias para o desenvolvimento da “inteligência verdadeira”, logo, representação não é posto como um elemento importante para essa inteligência, podendo até mesmo ser descartada, pois representação seria “a unidade errada de abstração na construção da maior parte dos sistemas inteligentes”.

Brooks refuta o uso da representação, no sentido utilizado na IA Clássica: estruturas centralizadas, as quais é atribuída alguma semântica (em geral, se referindo a algo no mundo). No entanto, ele aceita que possam existir nos sistemas da IA Clássica alguma forma de representação implícita, codificada, por exemplo, na topologia das conexões entre processos. Porém, em trabalhos posteriores, como em (BROOKS; STEIN, 1994a), o próprio Brooks mostra um ponto de vista menos radical em relação às abstrações fornecidas pelas representações. Ele continua afirmando que abordagens baseadas na Hipótese dos Sistemas de Símbolos Físicos não poderão explicar a inteligência de forma completa; porém, reconhece que é possível, e até pode trazer alguns benefícios, a utilização de algum tipo abstrações por parte desses sistemas. Essas abstrações seriam representações de informações fundamentadas nas “sensações” do agente, ao longo de sua história.

Emmeche (2007) e Steels (2003a) resumem as características dessa abordagem com uma lista de seus principais fundamentos:

- **Fundamentação física:** diz respeito à necessidade da existência de um corpo para o agente inteligente;
- **Interação dinâmica com o ambiente:** diz respeito à necessidade do agente (que possui um corpo), estar posicionado e poder interagir com seu ambiente (o que inclui a interação com outros agentes);
- **Comportamento emergente:** enfatiza que o comportamento do agente não deve ser (pelo menos não completamente) definido previamente pelos projetistas, mas deve emergir durante sua história. É priorizada a estratégia *bottom-up*, na qual comportamentos e mecanismos mais complexos devem surgir apoiados em mecanismos e comportamentos mais simples;

- **Adaptabilidade e evolução:** a emergência de determinado comportamento na Nova IA, muitas vezes é realizado através de técnicas como aprendizado (como no treinamento de redes neurais) e computação evolutiva (como algoritmos genéticos). Em geral, o aprendizado está voltado para a adaptabilidade do agente. Ou seja, o agente precisa se adaptar a determinadas situações que ocorrem ao longo de sua história. Por outro lado, a computação evolutiva atua a nível populacional, ao longo das gerações, permitindo o surgimento e a disseminação de determinadas características na população, assim como o desaparecimento de outras;
- **Acoplamento mais direto entre percepção e ação:** em lugar de sensoriar o ambiente, traduzir os dados sensoriais para outro formato (geralmente inteligível para um operador humano), realizar operações sobre esses dados, planejar as ações e só então executá-las, a Nova IA prioriza maior simplicidade nessa ligação entre percepção e ação;
- **Distribuição e descentralização:** o modelo de arquitetura proposta para os agentes inteligentes nessa abordagem é basicamente composta por módulos distribuídos, que eram responsáveis por algum aspecto do agente, e não existe o papel de um controlador centralizado das ações;
- **Mecanismos intrínsecos para lidar com limitações de recursos e conhecimento incompleto:** uma das limitações dos sistemas construídos com o modelo da IA Clássica é a dificuldade de lidar com mudanças de contexto, devido ao seu limitado conhecimento armazenado; os sistemas da Nova IA não se baseiam em acúmulo de conhecimento, mas em mecanismos capazes de lidar com problemas e situações comuns à maioria dos seres vivos, como a limitação de recursos e a falta de conhecimento. A adaptabilidade e a evolução são alguns desses mecanismos.

Vários pesquisadores seguiram essa abordagem, por exemplo, (CLARK, 1995; SUN, 2000; BROOKS; MAES, 1994b; VOGT, 2002; CANGELOSI; GRECO; HARNAD, 2002; STEELS, 1994; ROY, 2005). Um deles, Luc Steels, escreveu um trabalho intitulado “Inteligência com Representação” (Intelligence with Representation) (STEELS, 2003a), em uma clara alusão ao trabalho publicado anteriormente por Brooks (“Inteligência sem Representação” (BROOKS, 1991b)). Nesse trabalho, Steels admite vários méritos da nova abordagem, mas também aponta para alguns problemas e limitações que a ausência completa de abstrações pode causar. É verdade que conceitos como a incorporação, adaptação, emergência e situacionalidade contribuíram muito para a construção de sistemas que lidam com ambientes dinâmicos e em tempo real (PFEIFER; SCHEIER, 1999), mas como esses sistemas responderiam aos problemas que a abordagem baseada no conhecimento se propunham a resolver? Segundo ele, estes e outros aspectos diretamente ligados à noção de representação ainda carecem de respostas (e de resultados) convincentes da Nova IA.

Steels enxerga no discurso contrário a representações de muitos pesquisadores dessa abordagem (como o próprio Brooks), um dos motivos para o relativo pouco progresso nesse sentido, visto que desencoraja outros pesquisadores a seguirem esse caminho. Para

defender a utilização de representações, Steels pontua que a primeira coisa a perceber é que o sentido atribuído ao termo representação mudou ao longo do tempo, e que seu significado à época do surgimento da Nova IA (em torno de 1990), não é mais o mesmo utilizado na época em que escreveu seu trabalho (em 2003) (STEELS, 2003a).

Por volta de 1990 representações seriam vistas como entidades utilizadas internamente pelos agentes como substitutos para objetos no mundo; na prática, apenas uma forma de mapear o mundo externo em abstrações sem valor semântico para o sistema. Já em 2003 Steels defende que as representações (do mundo externo) não são mais vistas como apenas “entidades matemáticas abstratas” utilizadas por operadores externos para se referir a objetos no mundo, e ganharam uma conotação interna ao agente, que devem ser entendidas pelas qualidades de ajudarem na organização de atividades, na comunicação e na função de memória do agente (memória aqui utilizada não no sentido de um depósito no qual se armazenam coisas para serem acessadas no futuro, mas no sentido de um processo ativo, no qual a categorização e recategorização de estruturas, eventos e objetos, que o tornam significativos em relação a determinado objetivo). Alguns experimentos foram realizados por Steels e outros pesquisadores, como em (STEELS, 2001; STEELS; KAPLAN, 2002; STEELS; VOGT, 1997), na tentativa de avançar nessa direção, enfatizando a importância de que, mesmo seguindo a abordagem da Nova IA, algum tipo de abstração será necessária se quisermos que os agentes desenvolvam a habilidade para realizar tarefas que exigem cognição de alto nível, como linguagem e comunicação, por exemplo.

Por outro lado, Ziemke e Sharkey (2001) concordam que a abordagem da Nova IA, com os conceitos de incorporação, situacionalidade e desenvolvimento incremental das funcionalidades e comportamentos, oferece aos pesquisadores uma alternativa em relação às abordagens computacionalistas para o estudo do comportamento inteligente e, em específico, o uso de representações. Essa abordagem, ao contrário do conexionismo, traria avanços reais no sentido de prover aos agentes autônomos a capacidade de realizarem processos semióticos. Seguindo a definição de Morris, eles definem processo semiótico como “um processo no qual alguma coisa é um signo para algum organismo” (MORRIS, 1946) (discutiremos mais sobre semiose e signo na seção 2.6). Uma das razões é que ao inserir o agente artificial fisicamente no ambiente (e não apenas utilizar abstrações, como na IA Clássica), e ao proporcionar a experiência corpórea, dando a ele a possibilidade de ter respostas imediatas para suas ações, e em termos de suas próprias sensações (seus próprios sensores capturam essas respostas), é dado um passo na direção da intencionalidade¹, e de dar atenção ao “mundo perceptivo” do agente, como pontuou Emmeche (EMMECHE, 2007). Além disso, um número cada vez maior de pesquisadores vem reconhecendo as limitações que métodos puramente lógicos enfrentam para criar máquinas inteligentes. Na tentativa de fugir destas limitações, muitos pesquisadores são atraídos por métodos e princípios baseados na biologia, como é o caso das redes neurais, e da estratégia evolu-

¹Termo utilizado por Searle (1980) para se referir a uma “característica de certos estados mentais pelos quais eles são direcionados para ou sobre objetos e estados de coisas no mundo”. Uma característica que, segundo Searle, falta aos sistemas artificiais, na abordagem clássica da IA. Tal relação só existia na mente de um operador. Esse é um dos principais conceitos do Problema da Fundamentação do Símbolo, discutido com mais detalhes na seção 2.4

tiva. Essas técnicas, dentre outras coisas, pretendem dotar o agente de adaptabilidade e flexibilidade, características importantes na Nova IA, e contrastantes com a abordagem clássica de ter armazenadas respostas para cada situação concebível (EMMECHE, 2007).

Ao longo das últimas décadas, essa abordagem se tornou a mais popular e dominante, e muitos pesquisadores da IA, e em específico em trabalhos relacionados a emergência de linguagem e comunicação, a vem utilizando em seus trabalhos, e.g. (CANGELOSI; PARISI, 1998; CANGELOSI, 1999; NOLFI; FLOREANO, 2002; VOGT, 2002; STEELS, 2003b; MIROLI; PARISI, 2005; ROY; REITER, 2005; CANGELOSI; RIGA, 2006; FLOREANO et al., 2007; MAROCCO; NOLFI, 2007; MIROLI; PARISI, 2008; AMPATZIS et al., 2008; YONG; MIIKKULAINEN, 2009; GREEFF; NOLFI, 2010; NOLFI; MIROLI, 2010; TUCI et al., 2011; AGMON; BEER, 2013). Para uma revisão ver (WAGNER et al., 2003). Nem todos os trabalhos irão apresentar todas as características apresentadas aqui. As definições apresentadas tendem a criar uma classificação que abrange muitos trabalhos, e é natural que, a depender dos objetivos e restrições de cada pesquisa, seja dada mais ênfase à determinado aspecto do modelo, e que outros aspectos não sejam considerados significativos, sendo deixados de lado. Mas as ideias centrais, como comportamento emergente e interação com ambiente, em alguma medida, estão presentes em grande parte desses trabalhos (se não em todos).

2.4 PROBLEMA DA FUNDAMENTAÇÃO DO SÍMBOLO

Em 1980 o filósofo John Searle publicou um artigo que ficou bastante conhecido e cujas ideias apresentadas são muito discutidas até hoje: *Minds, Brains, and Programs* (Mentes, Cérebros, e Programas) (SEARLE, 1980). Nesse artigo, Searle faz severas críticas ao computacionalismo e apresenta o famoso Argumento do Quarto Chinês, experimento mental que sintetiza a maior crítica à IA Clássica. O Argumento do Quarto Chinês proposto por Searle pode ser resumido dessa forma: suponha que uma pessoa que não possui nenhum conhecimento do idioma chinês seja colocada em um quarto, e receba perguntas escritas em chinês, que devem também ser respondidas em chinês. No quarto há um manual, contendo regras de correlação entre perguntas e respostas em chinês. Então, para responder corretamente, a pessoa teria que identificar a pergunta recebida no manual apenas através de sua forma (do desenho das letras), e da mesma forma selecionar a resposta. Procedendo dessa maneira, a pessoa no quarto aparentaria, aos olhos de um observador externo, entender chinês perfeitamente. A figura 2.1 ilustra como ocorreria a dinâmica desse experimento.

Para Searle, porém, essa pessoa não “entende” nada das perguntas ou das respostas, pelo menos não no mesmo sentido de como um chinês entende o próprio idioma, ou da mesma forma que esta pessoa entenderia se as perguntas fossem feitas em seu idioma nativo, por exemplo. Searle argumenta também que em ambos os casos (tanto no experimento utilizando o chinês como no experimento utilizando o idioma nativo) a pessoa passaria no teste de Turing, mas que isso não é suficiente para afirmar que em ambos os casos o sistema entendeu as sentenças utilizadas da mesma maneira. De forma análoga, um sistema computacional que processa símbolos baseados apenas em regras de manipulação, incorreria nesse mesmo tipo de problema: mesmo que aparente agir como



Figura 2.1 Modelo de execução do argumento do quarto chinês. Fonte: alexandria.nu²

se estivesse entendendo determinadas estruturas simbólicas, na verdade estão apenas seguindo as regras sintáticas de manipulação de tais estruturas, e para o próprio sistema, nada significariam .

O experimento de Searle evidencia as principais críticas que a abordagem dos sistemas simbólicos recebeu: como uma coisa pode representar outra? Os símbolos são manipulados internamente, e podem ser interpretados por um observador externo, mas para o próprio sistema os símbolos manipulados nada representam. Esse problema de conectar símbolos ao que estes representariam de fato, foi estendido e refinado por Harnad (ZIEMKE, 1999), que o denominou de Problema da Fundamentação do Símbolo (HARNAD, 1990): Como a interpretação semântica de um sistema de símbolos formais pode ser realizada de forma intrínseca ao sistema, de modo a não depender da interpretação parasita de observadores externos? Como pode o significado de um símbolo, manipulado com base apenas no seu formato, ser fundamentado em outra coisa que não apenas outros símbolos também sem significado? Em outras palavras, para o próprio sistema computacional o símbolo manipulado não possui nenhum vínculo com aquilo a que se refere, mas apenas observadores externos atribuem significado aos seus símbolos e comportamentos; além disso, a noção de sistema fechado proposto por Newell e Simon (NEWELL; SIMON, 1976) deve ser repensada, visto que a relação de símbolos apenas com outros símbolos parece tornar impossível a tarefa de atribuir a eles algum significado para o sistema. Além de usar o exemplo do Argumento do Quarto Chinês elaborado por Searle, ele apresenta a analogia do dicionário circular. Harnad afirma que a tarefa enfrentada pelo modelo puramente simbólico da mente é equivalente aprender chinês como primeira língua, tendo como ferramenta apenas um dicionário Chinês-Chinês: como um sistema que apenas manipula símbolos e os relaciona com outros símbolos pode sair desse ciclo

²Disponível em: http://www.alexandria.nu/ai/blog/attachments/00000050_ChineseRoom.jpg

e relacionar os símbolos a outra coisa que não um outro símbolo sem significado para o próprio sistema?

Mas o que realmente seria “fundamentar” algo como um símbolo ou representação? Glenberg pontua que ao utilizar o termo “fundamentação” é necessário ser claro e definido de forma precisa (GLENBERG; VEGA; GRAESSER, 2008). Ele próprio adota a definição de fundamentação fornecida por Roy, que define fundamentação como o processo através do qual um agente (artificial ou humano) relaciona estruturas mentais com objetos físicos externos (ROY, 2005). Já Cangelosi define a fundamentação de símbolos como “o processo através do qual símbolos (como as palavras) são relacionados à representação semântica do mundo do próprio agente” (tradução nossa) (CANGELOSI, 2011). Essas definições apontam para o fato de que fundamentar representações simbólicas não pode ser descrito como sendo, por exemplo, inserir sensores em determinado agente artificial, de forma que esse agente possa identificar e nomear objetos do ambiente de acordo com as percepções fornecidas pelos sensores. Isso seria equivalente a conectar símbolos de forma puramente sintática, uma vez que as entradas provenientes dos sensores sejam desprovidas de semântica para o sistema, já que as questões sobre como tal conexão pode ser estabelecida não foram tratadas. Um agente com tais características não teria símbolos fundamentados, mas seria simplesmente um “rotulador de objetos”.

É notável a quantidade de pesquisadores que se dedicaram a esse tema ao longo das últimas décadas, e.g. (SEARLE, 1980; HARNAD, 1990; ZIEMKE, 1999; FROESE; ZIEMKE, 2009; PESCHL, 1994; ROY, 2005), utilizando várias estratégias em várias abordagens diferentes (TADDEO; FLORIDI, 2005). Tal interesse demonstra que o Problema da Fundamentação do Símbolo é uma questão bastante importante para várias áreas da IA, da filosofia, da semiótica, e das ciências cognitivas, e, em particular para áreas relacionadas à representação do conhecimento, comunicação e linguagem. Floridi (2004), por exemplo, afirma que o Problema da Fundamentação do Símbolo é uma das questões mais importantes para a filosofia da informação, e Cangelosi (2011) por sua vez, afirma que nas últimas décadas houve um grande avanço na formulação de novos modelos de evolução de comunicação e aquisição de linguagem na robótica, e que esses modelos estão diretamente relacionados à questão da fundamentação do símbolo, desde que os agentes (robóticos) são treinados para usar símbolos que em geral se referem a entidades ou estados no “próprio mundo do robô”. Essa declaração faz uma relação direta entre o desenvolvimento das pesquisas sobre o Problema da Fundamentação do Símbolo e áreas como a emergência de comunicação entre agentes artificiais.

Alguns conexionistas defenderam que sistemas construídos utilizando princípios do conexionismo não deveriam ser encarados como os sistemas simbólicos em relação ao Problema da Fundamentação do Símbolo. Chalmers, por exemplo, argumenta que, enquanto nos sistemas simbólicos a unidade de valor semântico (chamada por ele de representação) correspondia ao próprio *token* computacional, nos sistemas conexionistas essa sobreposição não existia, sendo a unidade de processamento computacional e a unidade de valor semântico entidades bastante distintas (CHALMERS, 1992). Dessa forma, ele entende que quando se fala em fundamentação de símbolos, a questão é essencialmente sobre representações, e o fato de que os *tokens* computacionais estejam fundamentados não necessariamente implica a representação esteja fundamentada, e propõe se falar

explicitamente em problema da fundamentação da representação. Ainda segundo Chalmers, essa distinção faria do axioma de Searle que diz que “sintática não é suficiente para semântica” (SEARLE, 1984), ter uma outra interpretação para o conexionismo, e se tornar falsa. O raciocínio utilizado é que esse axioma claramente traz um argumento de cunho linguístico, e seria verdadeiro ao ser aplicado a uma sentença, pois nela, as unidades semânticas coincidem com as unidades manipuladas pelas regras sintáticas (como nos sistemas simbólicos). Porém, seria falso ao ser aplicado em sistemas cuja unidade de aplicação das regras sintáticas não coincidem com as unidades de valor semântico (representações). Como analogia ele utiliza o caso do cérebro humano, que em nível mais baixo segue as regras da lei da física, e em nível mais alto apresenta um comportamento extremamente flexível e intencional. Para Chalmers, a sentença correta seria: “sintaxe, em um certo nível, nunca é suficiente para conteúdo semântico no mesmo nível”. Entretanto, ele se limita a expor o ponto de vista de que, contrariando a argumentação de Searle, é possível construir sistemas capazes de superar o Problema da Fundamentação do Símbolo utilizando uma abordagem computacional, desde que o nível da unidade computacional utilizada tenha uma granularidade menor do que as unidades de valor semântico, mas não há nenhuma proposição prática de um método de como conseguir tal feito.

Taddeo e Floridi (2005) fazem uma revisão de alguns trabalhos que buscavam atacar o Problema da Fundamentação do Símbolo entre 1990 (quando ele foi formalmente formulado por Harnad) e 2005 (ano em que publicaram o trabalho). Para analisar o sucesso desses trabalhos em resolver o problema, eles se baseiam em alguns princípios, denotados pelo termo condição Z (zero comprometimento semântico), que visam verificar (a) se há algum recurso semântico previamente instalado no agente; (b) se há alguma forma de interpretação parasítica, (feita por um agente externo, por exemplo); e (c) se o agente tem seus próprios meios para ser capaz de fundamentar símbolos (como por exemplo, sensores, atuadores e capacidade de aprender). Essas exigências deixam clara a perspectiva tomada pelos autores de que o próprio agente, em interação com seu ambiente, e utilizando seus próprios recursos, deve ser capaz de ser plenamente autônomo, tentando reduzir ao máximo a influência e dependência do agente artificial de um projetista externo. Foram analisados ao todo oito trabalhos e, segundo a análise realizada pelos autores, nenhum deles atende à condição Z, pois estão, de alguma forma, semanticamente comprometidos.

Segundo Taddeo e Floridi (2005), a estratégia dos representacionalistas consiste basicamente em fazer emergir essas representações nos agentes a partir da manipulação de seus dados perceptuais. Os agentes deveriam ser providos de sensores, capazes de capturar características relevantes dos dados percebidos, abstrair esse conteúdo e classificá-los como sendo de alguma representação de categoria e/ou conceito, e então utilizar essas representações para fundamentar o símbolo. Os representacionalistas consideram as representações conceituais e categóricas como o significado dos símbolos utilizados pelos agentes.

Um dos trabalhos representacionalistas analisados foi o de Harnad (HARNAD, 1990). Nesse mesmo trabalho no qual define o Problema da Fundamentação do Símbolo, Harnad utiliza essa abordagem e sugere um modelo híbrido, mesclando técnicas da abordagem simbólica e da abordagem conexionista. Harnad sugere que o símbolo pode ser funda-

mentado através de sua conexão com os dados perceptivos que denotam. Tal conexão seria estabelecida em três passos: iconização, discriminação e identificação. A iconização seria o processo de transformar os sinais perceptivos (oriundos dos sensores do agente), relativos a uma determinada entidade, em representações icônicas; essas representações icônicas, por sua vez, seriam sinais internos equivalentes à exposição do agente aos dados perceptivos daquele objeto. Discriminação seria o processo de decidir se duas entradas são iguais ou não e, caso não sejam, o quão diferentes são. Já a identificação seria algo semelhante a rotular, com um nome único (um símbolo, na visão clássica), determinada classe de entradas, colocando-as como equivalente ou invariante em algum aspecto. Os dois primeiros estágios utilizam redes neurais e representações sub-simbólicas, e tornam possível que no terceiro estágio essa representação seja associada com um nome símbolo. Como o modelo sub-simbólico não é apropriado para produzir e manipular representações simbólicas, o sistema que Harnad propôs também possui um sistema simbólico, que manipula sintaticamente os símbolos e promove sua fundamentação. Esse método, porém, não consegue satisfazer a condição Z, uma vez que as redes neurais utilizadas para fazer a classificação categórica das representações são invariavelmente embutidas de carga semântica colocadas pelo projetista, seja no processo de treinamento ou na inserção implícita de certos parâmetros.

As outras duas estratégias representacionistas analisadas se baseiam diretamente no trabalho de Harnad, e também incorrem em problemas em atender a condição Z. Mayo propõe uma organização funcional das representações, de modo que os símbolos sejam formados para ajudar a resolver um problema ou tarefa específica em um determinado ambiente (MAYO, 2003). A proposta fere a condição Z, pois dota previamente os agentes de critério funcional, além de previamente programar a tarefa que deve ser realizada. Taddeo e Floridi dizem que o agente possui previamente uma semântica “funcional”, o que invalida essa proposta. Já Sun propõe a arquitetura Clarion (SUN; PETERSON, 1998), que também utiliza redes neurais, e propõe uma estratégia evolutiva, através de algoritmos genéticos para evoluir populações de agentes, que dessa forma daria aos agentes a possibilidade, por exemplo, de emergir suas próprias necessidades e comprometimentos (SUN, 2000). Porém, de forma semelhante ao treinamento das redes neurais, o algoritmo evolutivo também é projetado e inserido no sistema por um projetista externo, que através da função *fitness* irá definir quais agentes serão selecionados para a próxima geração. E o projetista (algum deus ex machina), se baseia em julgamento de características que devem ser priorizadas e aprimoradas, ficando os próprios agentes alheios a essa definição.

Já os semi-representacionistas são diferenciados dos representacionistas essencialmente no uso das representações, onde adotariam uma visão mais próxima da Nova IA. Um dos trabalhos analisados por Taddeo e Floridi (2005) foi o de Vogt (VOGT, 2002), que adota uma estratégia de utilizar símbolos, mas de forma incorporada no agente; tenta unir a proposta de Harnad (HARNAD, 1990) com a robótica situada e incorporada de Brooks (BROOKS, 1991b, 1990), e transforma o Problema da Fundamentação do Símbolo no Problema da Fundamentação do Símbolo Físico. Ele diz que os símbolos devem ser o “acoplamento entre as ativações sensório-motoras do agente e seu ambiente”, e para auxiliá-lo, utiliza a definição de signo e símbolo de Peirce (que define o signo como um processo triádico, envolvendo representamen, objeto e interpretante) (PEIRCE,

1931-1935, 1958) para criar sua definição de símbolo semiótico, distinto da interpretação tradicional do termo símbolo, alterando composição e terminologia da definição peirceana. A proposta para resolução está baseada em jogos de linguagem entre robôs, que pretende estabelecer um vocabulário comum entre uma população deles. Porém, Taddeo e Floridi argumentam que essa definição só será verdadeira se o símbolo é interpretado por um agente que já possui proficiência semântica naquele vocabulário, pois de outra forma o processo semiótico defendido não ocorre, e não se pode atribuir aos agentes essa capacidade sem desvirtuar a questão.

Já os não-representacionistas tem suas principais bases de apoio nos trabalhos de Brooks (BROOKS, 1991b, 1990), nos quais, como já vimos, critica fortemente o conceito clássico de representação, e mesmo questiona sua necessidade para o comportamento inteligente. Porém, segundo Taddeo e Floridi, caso se pretenda desenvolver qualquer tipo de linguagem, mesmo uma forma muito simples, ou alguma capacidade cognitiva mais avançada do que “simplesmente” lidar com o ambiente, tal sistema deverá ser capaz de manipular símbolos, e então o Problema da Fundamentação do Símbolo surge. Um dos trabalhos analisados é o de Varshavskaya (VARSHAVSKAYA, 2002). Nesse trabalho Varshavskaya argumenta que a aquisição de semântica por agentes artificiais pode ser modelada a partir do desenvolvimento de capacidades linguísticas em crianças. Varshavskaya, porém, defende uma visão pragmática da linguagem, segunda a qual, a linguagem seria apenas um outro meio de interagir com outros agentes e com o ambiente, para satisfazer suas necessidades e atingir objetivos. Ainda segundo essa visão, essa semântica não precisaria de representação. No entanto, o experimento apresentado também não atende a condição Z. Uma das causas é que o aprendizado do agente tem intervenção externa (contrariando a exigência do agente de ter os recursos necessários para sua própria aprendizagem). Uma outra limitação apontada nesse trabalho é que, sem o uso de representações não é possível ao agente associar símbolos a categorias abstratas.

Conforme os critérios da condição Z, nenhum trabalho conseguiu resolver o Problema da Fundamentação do Símbolo, segundo Taddeo e Floridi (TADDEO; FLORIDI, 2005). Eles concluem que para resolver o problema, é preciso que não haja nenhuma pressuposição ou interferência externa nas capacidades de representação, categorização, comunicação e no uso da estratégia evolutiva. Tais exigências podem ser bastante complexas e difíceis de cumprir, e os próprios autores deixam claro que a questão da possibilidade de confeccionar tal sistema é uma questão a ser analisada de forma independente.

O Problema da Fundamentação do Símbolo visto dessa forma se torna bastante semelhante ao *problema de fundamentação*, definido por Ziemke como o problema de conectar um agente artificial com seu ambiente de forma causal, de tal modo que o comportamento, os mecanismos e as representações de tal agente sejam intrínsecas e significativas para ele próprio, e não necessite de alguma intervenção ou observador externo (ZIEMKE, 1999). Essa definição de Ziemke é mais genérica que as outras duas propostas por Glenberg (GLENBERG; VEGA; GRAESSER, 2008) e Roy (ROY, 2005), respectivamente, e expõe a perspectiva de que o problema da fundamentação não se limita aos símbolos, mas é um problema mais geral, e o problema da fundamentação do símbolo pode ser visto como um sub-problema. No exemplo do agente artificial que seria um mero rotulador de objetos, caso o agente conseguisse tratar das questões relativas a estabelecer a conexão

entre o rótulo e o objeto, fundamentando assim o rótulo utilizado, esse agente não sabe o que está fazendo, ou para quê utilizar os rótulos produzidos (não teria intencionalidade). Nas palavras de Ziemke, esses rótulos “não estão intrínsecos em nenhum contexto que permitiria/requereria ao agente fazer alguma utilização significativa desses rótulos” (tradução nossa). Da mesma forma, não estão intrínsecos ao agente, por exemplo, os sensores que ele utiliza, visto que foram projetados e inseridos por um projetista externo. Toda a arquitetura do agente estaria portanto sujeita a restrições impostas por projetistas externos, e não seria intrínseca a ele mesmo, nem surgida de necessidades próprias.

Tal observação pode levar a questionamentos muito sérios a respeito da viabilidade de criar um agente artificial totalmente fundamentado, visto que, segundo a visão defendida por Ziemke, qualquer imposição externa “necessariamente impõe restrições estruturais ao sistema” e este deixa de estar fundamentado. Porém, como aponta Loula, ao tratarmos de agentes artificiais estamos tratando de artefatos pensados, projetados e implementados por indivíduos (necessariamente externos ao agente artificial), pois de outra forma não seriam agentes artificiais. Essa contradição de interesses pode vir a negar a possibilidade de avanços nas pesquisas da área. Mesmo que não seja possível a fundamentação total, avançar na fronteira de prover cada vez mais autonomia e fundamentação a um agente artificial pode certamente ajudar a diminuir suas limitações (LOULA, 2011).

Contrariando as expectativas por conta de toda essa problemática, Steels publica um artigo cujo título “O Problema da Fundamentação do Símbolo está Resolvido. Qual será o próximo?”, sugere justamente que tais dificuldades já foram superadas (STEELS, 2008). Um dos pontos abordados nesse trabalho é a necessidade de clareza na utilização de termos como símbolo e representação. Ele discute sobre os conceitos desses termos, que têm sido fonte de muita confusão e discordâncias entre pesquisadores, especialmente entre cientistas da computação (e pesquisadores da IA, de maneira específica) e pesquisadores de outras áreas, como filósofos, cientistas cognitivos, linguistas, psicólogos e neurocientistas, o que tem dificultado o avanço das pesquisas na IA como um todo, e especificamente para o tema da fundamentação do símbolo. Representação, no sentido e uso tradicional, seria algo que está em lugar de outro, e o símbolo seria um tipo de representação. Segundo Steels, para resolver o problema da fundamentação do símbolo é preciso compreender como se originam e são escolhidos os significados utilizados para as representações, e como expressões de diferentes significados podem se combinar para gerar novos significados em uma representação combinacional. Em sua proposta, utiliza jogos de linguagem para mostrar que os agentes devem ser capazes de criar categorias e de relacionar dados perceptivos a estas (de maneira semelhante aos métodos classificados por Taddeo e Floridi (2005) como representacionalistas e alguns semi-representacionalistas), de forma que um símbolo esteja relacionado a uma categoria. Também devem ser capazes de negociar esse símbolo com outros agentes, de forma que um vocabulário comum possa surgir e uma comunicação efetiva possa ocorrer. Essa estratégia, no entanto, parece não trazer novidades significativas no tratamento da questão. Esses símbolos parecem ter as mesmas características dos rótulos citados anteriormente. Podemos mesmo reproduzir novamente as mesmas palavras de Ziemke para dizer que esses símbolos “não estão intrínsecos em nenhum contexto que permitiria/requereria ao agente fazer alguma utilização significativa desses rótulos” (ZIEMKE, 1999). De fato, esses símbolos não

possuem nenhum valor funcional real para os agentes, e apenas o utilizam para nomear outros objetos (ZIEMKE; SHARKEY, 2001). Ele finaliza pontuando que muito progresso já foi feito, mas que é preciso haver mais experimentos com agentes robóticos para que tenhamos melhor entendimento de certos aspectos da conceitualização e simbolização.

Em trabalho mais recente, Cangelosi concorda com Steels na afirmação de que se precisa saber mais sobre vários aspectos das representações simbólicas, e também que bastante progresso foi realizado no tema nas últimas décadas (CANGELOSI, 2011). Mas, segundo Cangelosi, não em todas as frentes do problema. Conectar símbolos ao que se referem, e fazer uma população de agentes compartilharem um mesmo vocabulário foram aspectos bastante trabalhados, nos quais já se tem certa maturidade e domínio. O aspecto negligenciado sobre o problema da fundamentação seria a transição da comunicação baseada em outras modalidades de representação (para mais detalhes sobre essas modalidades veja seção 2.6) para um sistema simbólico completo, e segundo ele é esse o principal problema a ser atacado. De fato, é uma questão de bastante interesse para várias áreas (como linguística e neurociência), como os humanos passam da interpretação e comunicação baseada em outras modalidades de representações para o domínio e uso dos símbolos. Essa transição é gradual (contínua), ou há uma mudança brusca? Cangelosi faz esses e outros questionamentos, e convida vários pesquisadores para ajudarem a respondê-los, procurando estabelecer um consenso do atual estado do problema.

2.5 MAS AFINAL, O QUE É REPRESENTAÇÃO?

Para a IA Clássica, as representações eram realizadas pelos chamados símbolos físicos (ou *tokens* computacionais), que teriam uma correspondência com algo no mundo real, e um papel fundamental para o raciocínio inteligente, pois é sobre eles que operam as computações. Para os conexionistas, uma representação se mostrava muito mais como um padrão nas conexões, cuja emergência se daria a partir de computações em entidades com granularidade menor. Já alguns adeptos da Nova IA mostram resistência quanto à necessidade da existência de representações para o comportamento inteligente, muitas vezes afirmando que ela não é realmente necessária (GELDER, 1998). Mas essas são realmente noções válidas? O quê exatamente seria “representar”? E mais, representando o quê e para quem?

Steels em (STEELS, 2008) faz uma análise interessante sobre o uso dos termos símbolo e representação na ciência da computação (de maneira particular na IA), e a confusão causada pela terminologia no diálogo não só com outras áreas, como filosofia e linguística, mas também dentro da própria IA. No final dos anos 1950, quando a IA estava surgindo então, os cientistas da computação usavam o termo símbolo no contexto de linguagem de programação, no qual era basicamente uma abstração computacional bastante útil e disponível em algumas linguagens de programação, como LISP e Prolog. A maior parte da confusão teria se dado então quando cientistas de outras áreas (filósofos, linguistas, psicólogos, neurocientistas, etc.), que não têm conhecimento sobre linguagens de programação liam artigos produzidos pelos pesquisadores da IA, e interpretavam o termo símbolo da mesma forma que interpretariam em artigos de suas áreas (geralmente como algo de algum valor semântico). Essa confusão foi aumentando, causando o que seria pro-

vavelmente a maior confusão terminológica da história da ciência. Dessa forma, parte da rejeição à IA Clássica se daria em razão dessa confusão entre terminologias. Algo semelhante teria acontecido também com o termo representação, que na mesma época (final dos anos 1950), quando a programação de alto nível iniciou seu desenvolvimento, era usada pelos cientistas da computação para designar estruturas de dados que carregavam informações para algum processo computacional, adotando, dentre tantos aspectos que as ciências humanas atribuíam à representação, apenas o aspecto de modelar algo do mundo de forma que possibilite o uso por parte de um programa de computador. Dessa forma, processamento de informação começou a ser entendido como projetar representações, e controlar os processos de criação e manipulação destas. E isso, outra vez causou muita confusão. O próprio Brooks em (BROOKS, 1991b), já nos anos 1990, quando defende que o comportamento inteligente não se fundamenta na representação, e que robôs senso-reando o mundo em tempo real poderiam adquirir melhor comportamento sem precisar de representações, estaria se referindo a essa noção.

Porém, existem outros pontos de vista sobre essa confusão de termos, e geralmente não são convergentes. Chalmers, por exemplo, diz que o termo símbolo é utilizado para se referir a um *token* computacional, ao qual não faz sentido atribuir valor semântico; já o termo representação deve ser utilizado para se referir a entidades que podem possuir valor semântico (CHALMERS, 1992). Por sua vez, Emmeche afirma que o conceito de representação como algo estável e explícito é alvo de muitas críticas, e leva muitos pesquisadores a se colocarem como anti-representacionistas no que diz respeito a cognição (EMMECHE, 2007). Essa visão e posicionamento, no entanto, seria uma maneira inadequada de pontuar que a visão da IA Clássica em relação aos termos símbolo e representação é demasiadamente simplista.

Podemos perceber que achar uma resposta para as perguntas feitas no primeiro parágrafo dessa seção, e encontrar algum consenso na comunidade, não é algo trivial, visto que a confusão teve início praticamente junto com a própria ciência da computação. Vários autores falam sobre essa indefinição, e dos problemas que ela pode causar às pesquisas. Glenberg, Vega e Graesser (2008) dizem que alguns pesquisadores até preferem não utilizar o termo representação, por que pode acabar causando confusão e atrapalhar mais do que ajudar. Bickhard e Terveen (1996) abordam o que eles afirmam ser uma questão fundamental para a IA e para as Ciências Cognitivas: a natureza da representação. Apesar de ser uma das principais bases para essas áreas, essa questão seria, na melhor das hipóteses, reconhecida parcialmente, e as abordagens mais utilizadas seriam incoerentes, em grande parte por pressuporem que toda representação constitui uma relação de correspondência com algo no mundo. A simples correspondência não pode ser suficiente para que algo seja entendido como representação. Cangelosi, Greco e Harnad (2002) veem a definição de símbolo (ou representação simbólica) como uma questão muito debatida, mas ainda sem consenso. Enquanto é relativamente fácil definir símbolo em um sistema computacional simbólico, a mesma coisa não pode ser dita se essa definição tiver que abranger também aspectos como linguagem e comunicação, se tornando então uma tarefa bem mais complexa. Davis, Shrobe e Szolovits afirmam que apesar de ser um dos temas mais fundamentais da IA, a questão sobre o que é representação do conhecimento, em geral, é deixada de lado e quase nunca respondida diretamente. Como dito no

início desse capítulo, este é um tema bastante controverso, que pode estar atrapalhando o avanço de pesquisas que de algum modo se relacionam com o tema representação. Alguns artigos apresentam as variedades de representações, enquanto outros falam das propriedades que devem existir em uma representação, mas fogem da pergunta central (DAVIS; SHROBE; SZOLOVITS, 1993). Dietrich e Markman se dizem surpreendidos pelo fato de que uma noção tão importante para as ciências cognitivas ainda não tenha uma teoria consolidada, e questionam: “como pode uma noção tão importante ser tão pobremente entendida?”. Porém, reconhecem que essa é uma questão bastante complexa, e que as ciências cognitivas ainda são um ramo relativamente novo do conhecimento humano, e provavelmente ainda não teve o tempo suficiente para construir e consolidar um entendimento mais profundo sobre representações (DIETRICH; MARKMAN, 2003). Beer e Williams (2014) enfatizam que o conceito de representação desempenha um papel fundamental na perspectiva clássica de processamento de informação na cognição, mas que a falta de uma definição aceita universalmente provoca dificuldades nas pesquisas.

Há, como citado, alguns pesquisadores que se abstêm de tecer definições diretas sobre representações (DAVIS; SHROBE; SZOLOVITS, 1993). Por exemplo, McCarty e Hayes chegam a afirmar que para um sistema ser capaz de se comportar de maneira inteligente no mundo, tal sistema deve ter uma representação real desse mundo, e fazem uma discussão sobre quais características devem existir em uma representação para que esta seja adequada em aspectos como metafísica, epistemologia e heurística, mas não há uma discussão nem uma definição sobre o que de fato seria uma representação (MCCARTHY; HAYES, 1968).

Por outro lado, existe também certo número de trabalhos que tentam fugir dessa confusão de termos e definir representação de alguma forma, se não de uma forma geral, pelo menos no âmbito do trabalho apresentado. A maioria faz isso atribuindo às representações, dentre outras coisas, a propriedade de ser, em alguma medida, substituto de outra coisa. Haselager (2005) antes de iniciar a discussão sobre o representacionalismo nas ciências cognitivas descreve uma definição do conceito de representação adotado. Para Haselager as representações têm duas características principais: “se colocar no lugar de algo”, e o sistema que a utiliza deve utilizá-la para guiar seu comportamento. Segundo ele, mesmo a ciência cognitiva clássica e o conexionismo, que discordam quanto ao efetivo formato das representações, concordam que a representação (a) carrega um conteúdo e (b) causam um comportamento. Davis também afirma que, dentre outras características, de forma mais fundamental representação é um substituto para outra coisa (que é substituída) (DAVIS; SHROBE; SZOLOVITS, 1993). Steels também diz que o sentido tradicional de representação é o de estar no lugar de algo, de forma que o algo representado possa, em algum aspecto, estar presente novamente; afirma ainda que não há nada de especial da natureza de um objeto que o torne apto para ser utilizado como representação, mas que um objeto passa a ser utilizado como representação simplesmente por que indivíduos decidem utilizá-los como tal (STEELES, 2008). Harvey (1992) assume a interpretação de que representações e referências simbólicas são a mesma coisa, e que estes podem substituir objetos no mundo real, em um sentido abstrato e absoluto. O sentido absoluto vem da perspectiva de que tanto os objetos quanto as representações existem de maneira independente do observador e, dessa forma, praticamente ignorando

qualquer papel que este possa exercer na ação representativa. Peschl tenta se basear em conceitos da neurociência para construir uma concepção de representação que seria caracterizada por uma relação/covariância estável entre o ambiente e algo no corpo do indivíduo, levando a um comportamento funcionalmente adaptado (PESCHL, 1994).

Brooks e Stein (1994a) apresentam uma analogia que diz que um circuito combinacional pode perfeitamente ser analisado em termos de lógica booleana, apesar de que o que existe de fato no circuito não são zeros e um's, mas diferenças de tensão elétrica. Da mesma forma, segundo eles, o fato de que determinados comportamentos de alguns agentes poderem ser analisados em termos de abstrações simbólicas, não implica a existência de entidades que correspondam a símbolos, dentro desses agentes. Dando um exemplo concreto, eles falam sobre o humano, e afirmam que apesar do seu comportamento poder ser analisado em termos de abstrações simbólicas, isso não muda o fato de que “tudo no cérebro é atividade neural” (BROOKS; STEIN, 1994a). Uma questão que surge dessa analogia é: apesar de determinado agente poder ser analisado em termos de representações, realmente existem representações nesse agente? A resposta para essa questão depende das limitações impostas na definição de tal entidade representativa. Pode-se, por exemplo, considerar que determinado padrão de atividade neural é uma representação, como fazem, por exemplo, os conexionistas. Mirolli (2012), utilizou uma definição semelhante a essa. Nesse trabalho, ele se propõe a verificar a existência de representação em um agente implementado a partir da abordagem dinâmica. Ele definiu representação como um estado interno correlacionado com algum aspecto externo e que seja funcional para o agente, ou como ele próprio enfatiza: “o estado interno deve contribuir para o comportamento do agente de uma maneira que seja adaptável em relação ao que ele se correlaciona”. Seguindo essa definição, Mirolli (2012) realizou dois experimentos (variando basicamente seu desenho) em busca de padrões de ativação de uma camada interna da rede neural que controla o agente, que obedecesse à definição sobre representação estabelecida, vendo nesses padrões a possibilidade de serem utilizados como representações.

A partir das discussões apresentadas, podemos notar que a noção de representação na IA não tem sido estável, mas tem se modificado bastante ao longo do tempo. No entanto, é fato que discussões sobre os mais diversos aspectos da representação sempre se fizeram presentes na comunidade. Parece ser uma noção que os pesquisadores, tanto da IA quanto de outras ciências cognitivas, precisarão tentar entender melhor e discutir o máximo possível para conseguir mais êxito em suas pesquisas. Como o termo representação é frequentemente utilizada de forma confusa e sem o devido cuidado, (infelizmente) podendo variar de sentido significativamente, é razoável (e desejável) que os pesquisadores deixem explícito em seus trabalhos o sentido com qual essa palavra está sendo empregada (HARVEY, 1992), pois o uso indiscriminado torna a utilização do termo cada vez mais vago e fracamente definido, podendo criar dificuldades para o avanço de pesquisas em determinadas áreas (HARVEY, 1992; MIROLLI, 2012). Em concordância com essa premissa, e com o objetivo de colaborar com um melhor desenvolvimento das pesquisas, especificamente na área da IA e ciências cognitivas, e para os fins a que esse trabalho se propõe, é necessário definir uma concepção clara e precisa sobre o que é representação (ou qual o significado desse termo neste trabalho), para que possamos procurar

suas possíveis ocorrências em agentes artificiais. É o que faremos na próxima seção.

2.6 SEMIÓTICA DE CHARLES S. PEIRCE

Ziemke e Sharkey lamentam o fato de haver pouca interação entre disciplinas que têm interesse em comum em representações, e apontam o fato de que a semiótica, campo de conhecimento dedicado ao estudo dos signos, ter tido até então pouca relevância na IA e nas ciências cognitivas de forma geral (ZIEMKE; SHARKEY, 2001). Esse parece ser um caminho interessante para encontrarmos uma definição de representação robusta e genérica o suficiente para ser entendida e utilizada por todos os pesquisadores que têm interesse no tema, e é esse o método que tentaremos aplicar nesse trabalho. Para essa tarefa, utilizaremos a teoria do signo, ou semiótica, de Charles S. Peirce como *framework* teórico para estudar o fenômeno da representação em sistemas cognitivos.

Em trabalhos nos quais simulações computacionais tratam da emergência de comunicação, há sempre algo que é comunicado de um agente para outro, e à isso (que é comunicado) são dados vários nomes: sinal, símbolo, som, palavra, expressão ou emissão. Na maioria desses trabalhos, o que é comunicado também parece ter capacidade de representação. Para tais entidades estávamos nos referindo até aqui pelo termo ‘representação’, para enfatizar essa propriedade e para utilizar uma expressão mais familiar à comunidade de inteligência artificial. Contudo, nessa seção utilizaremos o termo ‘signo’, como um termo técnico em uma revisão teórica. Isso significa que as representações que buscamos nesse trabalho são equivalentes aos signos definidos por Peirce.

Peirce teve atuação em várias áreas de conhecimento, e se destacou principalmente como filósofo e lógico, sendo considerado fundador da semiótica moderna (QUEIROZ; LOULA; GUDWIN, 2007), e tendo deixado um trabalho muito abrangente nessa disciplina. Ele define semiótica como “a doutrina da natureza essencial e fundamental de todas as variedades de possíveis semioses” (PEIRCE, 1935, §484). Para definir semiose, Sebeok utiliza o seguinte raciocínio: “do surgimento da civilização até a época presente, tem sido sempre reconhecido na cultura ocidental - pelo menos implicitamente - que há uma conexão intrínseca entre corpo, mente e cultura, e que o processo que interliga essas três dimensões da existência humana é a semiose, a produção e interpretação de signos” (tradução nossa) (SEBEOK, 2001). Assim, ao estudar “todas as variedades” de semioses, a semiótica desenvolvida por Peirce não faz restrições sobre a base material nos quais os processos de produção e interpretação de signos podem ocorrer, conseqüentemente não restringindo a natureza das possíveis criaturas envolvidas nesses processos (QUEIROZ, 2004). Além disso, parte do seu trabalho se dedica a “condições necessárias e suficientes da representação” (HOUSER, 1997). Essa grande abrangência aliada ao seu rigor formal, faz com que sua teoria influencie vários autores em diversas áreas de pesquisa, como por exemplo em (QUEIROZ; LOULA; GUDWIN, 2007; DEACON, 1998; HABERMAS, 1995; NOBLE; DAVIDSON, 1997; EMMECHE, 1991), e torna a aplicação dessa teoria confiável e adequada aos objetivos do presente trabalho, que necessita de uma definição de representação que seja precisa e genérica o suficiente para ser utilizada em agentes artificiais. Outros pesquisadores que também estudam a comunicação entre agentes artificiais e/ou o problema da fundamentação do símbolo, e que possuem considerável destaque na área,

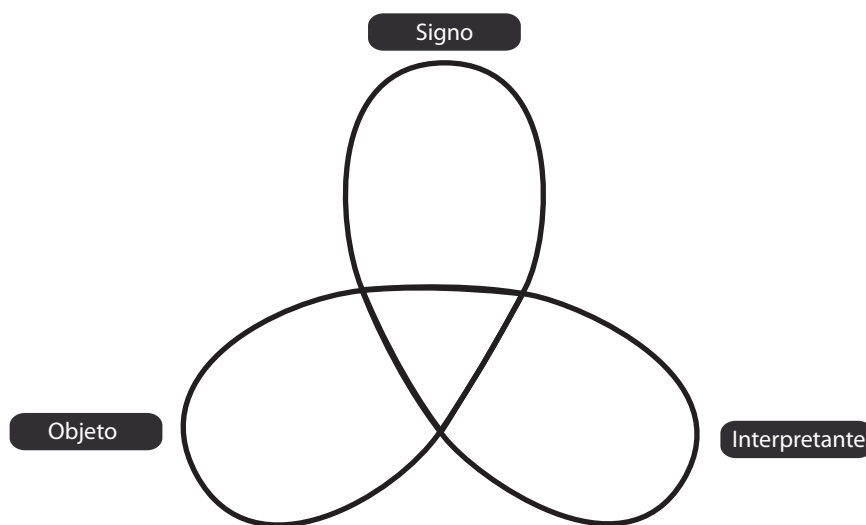


Figura 2.2 Modelo da relação triádica do modelo semiótico de Peirce, nó borromeo, adaptado de (QUEIROZ, 2004).

também utilizaram a teoria de Peirce para fundamentar seus estudos, como por exemplo (CANGELOSI; GRECO; HARNAD, 2002; CANGELOSI, 2001; VOGT, 2002; STEELS, 2008; ROY, 2005; LOULA, 2011) e nós em trabalho recente (OLIVEIRA; LOULA, 2014).

Para Peirce, o signo é essencialmente uma relação, um processo, que envolve de forma necessária e suficiente três elementos: signo, objeto e interpretante. A figura 2.2 ilustra o modelo dessa relação triádica. Nessa relação, o signo representa alguma coisa (seu objeto) em algum aspecto, para alguém (um intérprete), gerando neste um efeito interpretante (PEIRCE, 1935, §2.228).

Devemos notar que o termo signo é utilizado por Peirce para se referir à relação como um todo, envolvendo todos os elementos da tríade, mas também é utilizado para se referir especificamente a um desses elementos. Nesse último caso, trata-se do elemento significante, ou seja, o elemento que pode significar seu objeto em algum aspecto, como por exemplo uma fumaça como indício de fogo. Utilizado nesse sentido, o signo é também chamado de *representamen*, *representation* e *ground* em alguns dos seus textos. Nesse trabalho vamos sempre nos referir a ambos como signo e, se acharmos que o contexto não deixa claro a distinção de forma implícita, o faremos explicitamente. O objeto pode ser pensado como algo referenciado pelo signo, como por exemplo o fogo sinalizado pela fumaça. E interpretante é o efeito gerado no intérprete pela interpretação do signo, o qual pode ser compreendido como tradução (LISZKA, 1996; SAVAN; CIRCLE, 1988). É importante também entender que interpretante, não é um conceito ou algo necessariamente pré-estabelecido na mente do intérprete, mas um efeito produzido no intérprete, um produto da tradução, “é algo criado na mente do intérprete” (PEIRCE, 1958, §8.179). Esse efeito pode ser por exemplo, uma sensação, a tomada de uma ação, ou outro signo (LOULA, 2011).

Para Peirce, só existe processo de representação no momento em que o signo opera,

estabelecendo a relação com seu objeto e gerando um efeito interpretante em algum intérprete (PEIRCE, 1935, §1.540), de forma que faz mais sentido dizer que algo atua como um signo em determinado contexto, do que determinar algo como sendo intrinsecamente um signo. Assim, não existe algo como uma propriedade que torne algo um signo, mas basta que ele seja interpretado como tal.

Como já definimos anteriormente, o signo não representa seu objeto como um todo, mas sim em alguns aspectos. Da mesma maneira, nem todos os aspectos do objeto são importantes para que ele seja representado por um signo. Para Peirce, o objeto determina o signo. Essa determinação não deve ser entendida no sentido causal, mas na medida em que o objeto impõe certas restrições ou condições ao signo, e este deve atendê-las para que possa representar tal objeto. Por exemplo, considere uma relação onde o signo é a fumaça e o objeto é o fogo. Nesse contexto, alguns aspectos do objeto, como temperatura e quantidade do fogo não são relevantes, assim como também não o são alguns aspectos do signo como a cor, cheiro e a densidade da fumaça. O que importa é a existência do fogo, e a presença de alguma fumaça que possa indicar esse fato para alguém. Por outro lado, se a fumaça falha em indicar a presença do fogo, então ela não poderá ser seu signo. Note que esse exemplo diz respeito apenas a indicação da presença do fogo pela fumaça. A temperatura do fogo, assim como cheiro, cor e densidade da fumaça podem agir como outros signos e indicar algumas propriedades do próprio fogo, como o material que está sendo consumido.

De forma similar, o interpretante é determinado pelo signo, da mesma maneira, não de forma causal. Esta determinação ocorre pelo uso de determinadas características que o signo utiliza para referenciar seu objeto de forma que gera e molda o entendimento do intérprete. Utilizando o mesmo exemplo da fumaça e do fogo, podemos afirmar que a maneira pela qual a fumaça gera um interpretante para seu objeto (fogo), é convergindo sua atenção para a conexão física entre fumaça e fogo, seja por meio visual (sua conexão física com o fogo) ou olfativo (ao percebermos o cheiro da fumaça, sabemos que provavelmente em um raio de distância relativamente próximo há, ou houve a pouco tempo, algum fogo). E existem propriedades na fumaça que não tem qualquer relação direta com o fogo, como é o caso da sua direção (que pode ser relacionada com a direção do vento e pressão atmosférica). Logo, essa propriedade não pode ser utilizada como signo do fogo, visto que ela não molda o entendimento do intérprete de forma adequada.

No âmbito da comunicação entre agentes, o signo é também o mediador da relação entre o falante e o intérprete. O signo produzido pelo falante gera no intérprete um efeito (interpretante): “é conveniente falar como se o signo tivesse sido originado com um falante e determinado seu interpretante na mente de um intérprete” (PEIRCE, 1967). A definição de signo, em conjunto com essa ideia de comunicação, pode ajudar a compreender e a elaborar experimentos em que há emergência de comunicação entre agentes autônomos, a partir de perspectivas ainda não exploradas.

2.6.1 Ícone, Índice e Símbolo

Na teoria sógnica de Peirce, ao longo do tempo, foram identificadas várias classes de signo. Entretanto, podemos afirmar que, a partir da natureza da relação entre signo e

objeto, três classes mais fundamentais podem ser estabelecidas. São elas: Ícone, Índice e Símbolo.

Ícones são signos que compartilham alguma característica ou qualidade similar com o objeto que representam, ou seja, há algum grau de semelhança entre signo e objeto. Exemplos de ícones são pinturas, desenhos e esculturas. Apesar da conotação poder sugerir, ícones não estão restritos a serem entidades visuais. Uma versão de uma música pode trazer a mente a música original, um imitador fazendo imitações de características marcantes de uma pessoa famosa, trazem também de imediato a lembrança da pessoa imitada. Para Peirce,

Um ícone é um signo que se refere ao Objeto que denota apenas em virtude de seus caracteres próprios, caracteres que ele igualmente possui, quer um tal Objeto exista ou não. É certo que, a menos que realmente exista tal Objeto, o Ícone não atua como signo. Qualquer coisa, seja uma qualidade, um existente individual ou uma lei, é Ícone de qualquer coisa, na medida em que for semelhante a essa coisa e utilizado como um signo seu. (PEIRCE, 1935, §2.247)

Já o índice se relaciona com seu objeto através de uma proximidade física, espaço-temporal ou de uma relação causal, constituindo uma relação de caráter associativo. Também é característica do índice ser, de algum modo, afetado por seu objeto. Por exemplo, marcas de pneus de carro em uma estrada fazem supor que, não apenas os pneus passaram por ali, mas um automóvel, e mais ainda, um automóvel com ocupante(s). Ou seja, apesar das marcas serem claramente identificadas como tendo sido feitas por pneus, podemos inferir outros objetos e contextos associados. Outro exemplo de índice é quando ouvimos o som de um trovão, e somos levados a imaginar que logo virá uma chuva ou uma tempestade, gerando uma relação de proximidade tanto de espaço (a chuva está próxima para que consigamos ouvir o trovão) quanto de tempo (em poucos minutos irá chover). Uma seta indicando um quebra-molas é um exemplo de índice que se utiliza da proximidade física com o objeto que representa. Segundo Peirce o índice “direciona a atenção para seu objeto por compulsão cega” (PEIRCE, 1935, §2.305). Sendo característica do índice direcionar a atenção do indivíduo para seu objeto, fica implícito que de forma diferente do ícone, o índice depende da presença do seu objeto nas proximidades para que possa representá-lo.

O símbolo, por sua vez, ocorre quando o signo e o seu objeto não possuem nem qualidades semelhantes, nem proximidade física, espaço-temporal ou relação causal, sendo então essa representação de caráter convencional ou habitual. Mas se o símbolo não se relaciona nem por semelhanças (signos icônicos), nem de forma física, espaço-temporal ou causal (signos indiciais), de que maneira um signo simbólico pode ser relacionado com um objeto? Segundo Peirce, o símbolo se relaciona com seu objeto através de “uma convenção, um hábito ou uma disposição natural” (PEIRCE, 1935, §8.335). Por exemplo, quando um falante usa um substantivo qualquer para se comunicar, seja de forma verbal ou escrita, o intérprete deste signo comunicado só terá a capacidade de interpretá-lo corretamente se já houver aprendido a convenção utilizada para aquele símbolo. Se for um estrangeiro, por exemplo, precisará aprender o idioma; se for alguém que não

saiba ler, deverá ser alfabetizada para entender a forma escrita. Assim, se o sujeito não estabelecer, em caráter convencional ou habitual, a relação do símbolo com seu objeto, este simplesmente não existe, já que este signo não possui nenhuma característica que lembre seu objeto, nem estabelece com este uma relação causal, física ou espaço-temporal. É necessária então uma disposição natural ou convencional do sujeito intérprete para que essa relação aconteça. Peirce ainda diz que

Um símbolo é um [signo] que realiza sua função sem a consideração de qualquer similaridade ou analogia com seu objeto e igualmente sem a consideração de qualquer conexão factual com ele, mas somente e simplesmente porque irá ser interpretado como sendo um [signo]. (PEIRCE, 1935, §5.73)

Palavras são os exemplos mais comuns de símbolos, mas não se limitam a isso. Placas e sinais de trânsito, em geral são elementos simbólicos (e não icônicos, como muitos poderiam pensar), visto que é preciso ter aprendido sua simbologia para interpretar, por exemplo um sinal vermelho ou uma placa que indique alfândega.

Liszka (1996) apresenta ainda outra característica dos símbolos, que aparecem em alguns textos de Peirce: sua característica de ser geral, ou seja, o objeto ao qual o símbolo se refere, normalmente, é de tipo geral, e não individual. Por exemplo, a palavra "árvore" é um símbolo que se refere não a uma instância de árvore individual, mas a a classe de objetos que são árvore, de maneira geral. Esse tipo mais comum de símbolo é classificado como símbolo genuíno. Se um símbolo se refere a um objeto individual (como "lua", por exemplo), é classificado como símbolo singular. E ainda, se o objeto do símbolo é uma característica ou qualidade, então é classificado como símbolo abstrato.

As três classes de signo (ícone, índice e símbolo) não são mutuamente exclusivas, mas podem perfeitamente ser interrelacionadas. Para Peirce "Um Símbolo é uma lei, ou uma regularidade do futuro indefinido. [...] Mas uma lei necessariamente é governada, ou 'é incorporada em' indivíduos, e prescreve algumas de suas qualidades. Consequentemente, um constituinte de um Símbolo pode ser um Índice, e um constituinte pode ser um Ícone" (PEIRCE, 1935, §2.293). Índices e ícones podem ser elementos constituintes do símbolo, o precedendo.

Como só ocorre o processo semiótico no momento em que o signo opera, não é possível saber antes do processo acontecer, se determinada entidade é um ícone, índice ou símbolo, mas apenas no momento no qual ele é interpretado como sendo um signo de alguma dessas classes. Uma palavra, algo comumente associado ao símbolo por exemplo, pode atuar também como um índice. Se alguém grita alto a palavra "Fogo!", ocorrem duas coisas, tanto a interpretação da palavra "fogo" (conforme convenção), como a atenção das pessoas no entorno é, pelo menos inicialmente, voltada para o vocalizador. Nenhuma pressuposição pode ser feita sobre a natureza do signo, até que seja analisado o efeito que ele causa no intérprete.

O tema representação tem gerado muitas discussões na ciência da computação, e em particular na área da IA. Muitas abordagens diferentes dessa área de estudo já levantaram essa questão, cada uma com um entendimento diferente, e sem consenso. Cada escola tem um entendimento diferente do que pode ser representação, e atribuindo mais ou menos importância à sua influência para o comportamento inteligente. Termos como

símbolo, representação e representação simbólica são utilizados muitas vezes sem que fique claro para o leitor o que eles realmente significam. Ao utilizar a semiótica de Peirce, pretendemos evitar o problema da indefinição e a confusão que pode ser gerada pelo uso do termo e oferecer uma definição genérica, coerente e embasada em um arcabouço teórico bastante robusto. Uma definição genérica significa que a teoria de Peirce não possui comprometimentos com domínios específicos de conhecimento, como linguística, filosofia, psicologia e ciência da computação. Essa característica é importante, uma vez que, como vimos na seção 2.5, alguns autores apontam que o uso de termos como símbolo e representação foram introduzidos na ciência da computação com significado diferente dos significados com que eram utilizados em outras áreas, iniciando (ou aumentando) assim toda a confusão. Agora, quando o termo símbolo for utilizado, sabemos que estamos nos referindo a um tipo específico de representação (aquele para o qual o intérprete teve que estabelecer a relação com seu objeto por convenção ou hábito), e que representação é um processo que envolve necessariamente três elementos, independentemente de seu contexto.

ESTUDOS DE LINGUAGEM, COMUNICAÇÃO, E SIMULAÇÕES COMPUTACIONAIS

A linguagem humana é uma característica tão importante para nossa espécie que, tanto do ponto de vista científico quanto tecnológico, existe um grande interesse em conhecê-la profundamente em seus vários aspectos. Importantes aspectos científicos estão relacionados à origem e evolução de linguagem, por exemplo, do ponto de vista evolucionário, como nossa espécie passou de um estado pré-linguístico para o efetivo uso da linguagem (GONG; SHUAI, 2013). Do ponto de vista tecnológico, conseguir que máquinas utilizem linguagem de maneira eficiente pode representar avanços consideráveis, especialmente em áreas como: robótica, interações homem-máquina e máquina-máquina, extratores de informações textuais, engenhos de busca, sistemas de recomendação, web semântica, banco de dados, sistemas de defesa contra fraude, tradutores automáticos entre idiomas, dentre outras. E esses avanços tecnológicos dependem, em grande parte, do entendimento das questões científicas (como e por que utilizamos a linguagem de forma tão eficiente e o problema da fundamentação do símbolo são exemplos dessa fronteira na qual o avanço tecnológico depende do avanço científico).

Por outro lado, estudos sobre linguagem, em geral, são bastante complexos, exigindo muitas vezes conhecimento multidisciplinar. As maiores dificuldades são justamente os fatos de que: o comportamento linguístico não deixa fósseis (e dificilmente deixa outras evidências), e o seu surgimento é um fenômeno que não se repete. Biologia evolucionária, antropologia, arqueologia, paleontologia, linguística, ciências cognitivas e neurociência são algumas das disciplinas envolvidas nesses estudos. Cada disciplina envolvida possui sua própria metodologia e prática (que muitas vezes são difíceis de conciliar). Por exemplo, alguns pesquisadores da área genética tentam identificar e analisar genes que tenham relação com a capacidade linguística dos humanos, como (FISHER et al., 1998); arqueólogos e antropólogos tentam relacionar o surgimento de vestígios simbólicos com o aparecimento da capacidade linguística, como (D'ERRICO; HENSHILWOOD; NILSSON, 2001). Porém, praticamente nenhuma pesquisa possui dados empíricos consistentes (do ponto de vista estatístico/científico) que possam de fato confirmar determinada teoria

(CANGELOSI; PARISI, 2002). O resultado é uma grande quantidade de teorias, mas poucas conclusões.

3.1 SIMULAÇÃO COMPUTACIONAL E ESTUDO DE LINGUAGEM E COMUNICAÇÃO

Nesse cenário, a simulação computacional surge como mais uma importante ferramenta, que pode contribuir bastante para as pesquisas dessa área, e experimentos sintéticos vêm ganhando espaço nos últimos anos entre a comunidade interessada no melhor entendimento de mecanismos utilizados para a comunicação linguística, e em como ela se desenvolve (STEELS, 1997), principalmente depois do trabalho de Hurford (1989) (GONG; SHUAI, 2013). A simulação é uma maneira de expressar, testar e avaliar teorias e hipóteses através de programa de computador, uma vez que o programa deve ser ajustado de forma a fornecer como resultado previsões (empíricas) de comportamento de determinado sistema, baseado nessas teorias e hipóteses (CANGELOSI; PARISI, 2002). Uma simulação baseada em um modelo teórico não pode, em princípio, ser utilizada para provar a validade de determinada teoria empiricamente, mas pode fornecer conclusões acerca de sua coerência e da probabilidade de se obter determinado efeito, assumindo-se um conjunto de pressuposições (JÄGER et al., 2009).

A metodologia de síntese de experimentos é uma alternativa à abordagem puramente analítica. A metodologia analítica é largamente adotada na ciência, e nela geralmente é adotada uma abordagem do tipo *top-down*, na qual o todo vai sendo subdividido, e suas partes sendo analisadas de forma recursiva, até se chegar em uma granularidade que seja suficiente para se analisar o fenômeno pretendido, ou alcançar a impossibilidade de continuar as divisões. Por exemplo, na linguística, uma determinada linguagem é analisada e, para isso, é dividida em termos sintáticos e léxicos; o léxico por sua vez é decomposto em morfemas e fonemas, e assim por diante (CANGELOSI; PARISI, 2002). A abordagem sintética, que pode ser adotada com as simulações computacionais, por sua vez, utiliza a máxima de “construir para explicar”, adotando uma abordagem *bottom-up*. O pesquisador decide, dentre outras coisas, quais são as condições iniciais, as regras que devem ser aplicadas no sistema, o ambiente da simulação. O que ocorre então é a emergência de comportamentos e entidades mais complexas (de nível mais alto) que surgem em decorrência da interação entre elementos do sistema e a aplicação das regras e condições previamente definidas (CANGELOSI; PARISI, 2002). Esse comportamento do sistema é bastante semelhante ao que Camazine et al. definem como sistema auto-organizado, no qual o comportamento global do sistema surge a partir de inúmeras interações entre os componentes de menor nível (CAMAZINE et al., 2001). A estratégia sintética, através da simulação computacional pode lidar com essas características e ser utilizada para estudar linguagem de forma mais apropriada e eficiente (GONG; SHUAI, 2013).

A simulação computacional pode ainda ajudar no diálogo entre as várias disciplinas envolvidas no estudo da linguagem, na medida em que proporciona um meio comum para que teorias, hipóteses, regras e cenários oriundos de disciplinas diversas possam ser colocadas juntas em uma mesma simulação. Esse tipo de relacionamento já acontece há algum tempo, por exemplo, entre áreas como ciências cognitivas e vida artificial (CANGELOSI;

PARISI, 2002). Além disso, devido a sua natureza, um grande número de fatores, dos mais diversos tipos, podem influenciar a linguagem (esse é um dos motivos responsáveis pelo grande número de disciplinas envolvidas em seu estudo) e, nessas circunstâncias, nas quais grande quantidade de processos devem ser considerados simultaneamente, as ferramentas computacionais podem ser de grande ajuda, “até mesmo necessárias”, visto que nossa capacidade intuitiva sobre esses processos é bastante limitada (CHRISTIANSEN; KIRBY, 2003).

Alguns outros benefícios podem ser extraídos da utilização das simulações computacionais. Alguns são explícitos, decorrentes diretamente dos resultados das simulações, enquanto outros podem ser considerados implícitos, por imporem condições e restrições técnicas e metodológicas aos pesquisadores. Por exemplo, para ser possível colocar uma teoria em forma de um programa de computador, ela deve estar bem definida. Muitas vezes as teorias sobre questões relacionadas à origem da linguagem são definidas em termos muito vagos e genéricos, o que impede especificações para previsões empíricas sobre esses termos (LYON; NEHANIV; CANGELOSI, 2007). Mas, para que o programa funcione de forma apropriada, determinada teoria deve ser declarada de forma “explícita, detalhada, consistente e completa” (CANGELOSI; PARISI, 2002), e suas suposições e condições iniciais também devem ser claras e explícitas (JÄGER et al., 2009). Nesse ponto, as simulações computacionais podem ajudar aos pesquisadores também na melhoria de suas teorias, na medida em que obriga a estarem descritas de maneira clara e precisa. Além disso, as simulações podem ser úteis inclusive quando os resultados não estão de acordo com a previsão ou com a realidade (mesmo quando a teoria é inserida de forma apropriada no programa), pois podem sugerir uma revisão nas teorias estudadas (GONG; SHUAI, 2013).

Mais do que simplesmente executar/testar determinada teoria/hipótese, as simulações computacionais podem servir como um laboratório virtual, permitindo que os pesquisadores executem seus experimentos em ambientes controlados e possibilitando que modifiquem parâmetros específicos e verifiquem as consequências dessas modificações nos resultados (CANGELOSI; PARISI, 2002). Os laboratórios virtuais auxiliam ainda em verificações dentro de um determinado espaço de hipóteses, que não poderiam ser feitas por meio empírico ou em laboratórios convencionais, pois estes provavelmente irão cobrir apenas uma porção limitada do espaço de busca (e.g. número limitado de condições iniciais, de indivíduos ou populações). Esses recursos oferecem aos pesquisadores um grande auxílio, com ferramentas nas quais se pode explorar grande quantidade de condições, e certamente podem ajudá-los não a voltar no tempo, mas a elaborar teorias cada vez mais robustas e precisas do processo de surgimento e evolução da linguagem (GONG; SHUAI, 2013). Além disso, esses laboratórios são ferramentas de natureza quantitativa, e essa característica também pode representar um avanço nas pesquisas sobre origem e evolução da linguagem, e para o modo como a entendemos, visto que é uma característica das ciências mais maduras (como as ciências naturais), expressarem teorias, dados e resultados de maneira quantitativa (CANGELOSI; PARISI, 2002).

No entanto, é preciso ser cuidadoso para que o uso da simulação computacional para o estudo da linguagem não seja superestimado. Ela deve sempre ser utilizada em conjunto com outras ferramentas, como *frameworks* teóricos, modelos matemáticos, utilizar dados

oriundos de áreas como linguística, antropologia, semiótica, arqueologia e biologia evolutiva; e oferecer meios para sugerir novos entendimentos, perspectivas e experimentos para os pesquisadores de todas as áreas envolvidas (CHRISTIANSEN; KIRBY, 2003). Devemos ainda ter em mente que simulações são realizadas com base em modelos, e modelos tendem a ser uma simplificação da realidade, não contemplando ou abstraindo determinadas características. Modelos tendem também a serem específicos, na medida em que tendem a focar em determinadas questões e características e deixarem outras, que poderiam impactar diretamente nos resultados da simulação, de lado (GONG; SHUAI, 2013). Portanto, é preciso ponderar essas nuances, e construir um modelo que aborde todas as características e questões necessárias para determinada pesquisa, apoiado em teorias e hipóteses construídas com auxílio de outras disciplinas, balanceando entre simplicidade (para não construir modelos triviais) e complexidade (para não construir modelos muito difíceis de serem colocados em um programa de computador, e cujos resultados gerados sejam impossíveis de serem analisados na prática), para construir simulações realmente relevantes (GREEFF; NOLFI, 2010).

3.1.1 Tipos de simulação

Como já citado neste capítulo, é uma característica dos modelos computacionais focarem em questões e hipóteses específicas, dando certo viés às simulações criadas. Assim, é de se esperar que surja algum tipo de classificação para estes experimentos de forma a caracterizar, de maneira geral, as principais questões e *frameworks* teóricos utilizados para realizar essas simulações. Jäger et al. (2009) afirmam não conhecer uma classificação largamente aceita na comunidade, e fazem então uma proposta de algumas classificações, das quais discutiremos três que são mais relevantes dentro do contexto desse trabalho. Essa classificação levou em conta o foco dado aos aspectos de transmissão cultural, comunicativos e cognitivos, e biológicos, que correspondem, respectivamente, aos modelos de aprendizagem iterativa, jogos de linguagem e evolução genética. Essas são classificações bastante genéricas e abrangentes e, por consequência, pode-se ainda subdividi-las e derivar várias outras. Também nem sempre será claro a classificação de determinado trabalho, desde que não há uma separação tão clara entre a especificidade de cada trabalho, podendo um mesmo trabalho mesclar vários fatores e técnicas utilizadas em todas essas classificações consideradas aqui. Porém, é bastante razoável que a maioria dos trabalhos de simulação computacional na área de estudo da origem da linguagem e comunicação possa ser classificada nesses termos, ou que possa ser explicada a partir desses paradigmas.

O paradigma do aprendizado iterativo enfatiza a questão do relacionamento entre restrições e tendências biológicas e propriedades universais da estrutura linguística. O foco está na ideia de que um desafio importante para a linguagem é ter que ser repetidamente repassada (transmitida) entre os indivíduos de uma população, ao longo das gerações, sendo esse processo suscetível a vários tipos de erros e interferências. Então, o resultado é um sistema adaptativo, no qual a linguagem evolui culturalmente, de forma que, apesar dos erros e interferências que podem ocorrer, é transmitida com bastante fidelidade entre as gerações, sem que o sistema de comunicação perca suas características principais. Na maioria das simulações, a principal simplificação é que a transmissão do comportamento

linguístico se dá em apenas em nível individual (um indivíduo "ensinando" outro), e entre gerações (ou seja, não há influência entre indivíduos que estejam em uma mesma geração).

Como o próprio nome já diz, o modelo dos jogos de linguagem utiliza jogos na interação comunicativa entre dois indivíduos pertencentes a uma população, com objetivo de cooperarem em alguma tarefa. Esses indivíduos podem ser simulados ou estarem incorporados (serem robôs) e situados em algum ambiente compartilhado por ambos. Em cada interação entre dois indivíduos (em geral, escolhidos de forma aleatória), há um falante e um ouvinte, que deverão interagir para resolver determinado problema. Para isso é preciso tomar ações como expandir o significado de algum signo ou inferir significado para outros (o que requer capacidade cognitiva). Por exemplo, se um agente deseja expressar um conceito para o qual ainda não houve nenhum signo relacionado, este cria um e o adiciona na memória (LEKVAM; GAMBÄCK; BUNGUM, 2014). Dessa forma, é esperado que a população venha a convergir sobre as convenções de comunicação utilizadas. A questão principal abordada é como a linguagem surge e evolui através de estratégias como invenção e adoção de termos e atribuições de significados por parte dos indivíduos, em uma população de agentes incorporados. A exemplo do aprendizado iterativo, a interação nos jogos de linguagem ocorre em nível individual, entre pares (pelo menos a cada interação, em iterações diferentes é possível que um indivíduo faça par e interaja com indivíduos diferentes).

O modelo de evolução genética utiliza princípios biológicos para modelar a transmissão da linguagem e comportamento comunicativo através de código genético. As simulações utilizam alguma estratégia da computação evolutiva para evoluir a população de agentes. De forma geral, os agentes são modelados de maneira que seu comportamento, incluindo características de linguagem, estão codificados em parâmetros que simulam um código genético. Esses agentes interagem e, baseado no seu sucesso comunicativo, cada agente obtém (geralmente de forma indireta) um valor de *fitness*, que reflete sua possibilidade de gerar filhos para formar a próxima geração (seu sucesso reprodutivo). Portanto, quanto maior o sucesso comunicativo, mais chances o indivíduo terá de gerar descendentes. Ao focar na contribuição do aspecto da seleção natural para o surgimento e evolução da linguagem, esse modelo acaba simplificando vários aspectos da genética evolutiva, como por exemplo o cálculo do *fitness*, no qual ignora diversos outros fatores que contribuem para o sucesso reprodutivo de cada indivíduo.

3.1.2 Simulações em emergência de comunicação

Apesar de cada uma das abordagens descritas na seção 3.1.1 enfatizarem questões específicas, podemos perceber que a característica da comunicação está presente em todas elas. Apesar de existir alguma distinção entre comunicação e linguagem em alguns trabalhos, essa diferenciação nem sempre é significativa, pois existe uma grande conexão e sobreposição de interesses nas pesquisas sobre linguagem e comunicação (CANGELOSI; PARISI, 2002). Além disso, utilizando princípios e técnicas relacionadas com a nova IA, cada vez mais as simulações tendem a trabalhar com formas de comunicação diferente da linguagem humana, e a utilizar a linguagem dos próprios agentes (ZIEMKE; SHARKEY, 2001), o que torna essa diferença ainda menor no contexto dessas simulações.

Sob influência das ideias da Nova IA, há uma tendência cada vez maior de tratar linguagem e comunicação sob uma perspectiva mais ampla de cognição, no sentido de não mais estudar esses fenômenos de forma isolada, ou a partir de descrições puramente sintáticas. Essa influência vem desde o início da década de 1990, como podemos confirmar por a afirmação de Brooks (1994):

“Tendências recentes na inteligência artificial, ciência cognitiva, neurociência, psicologia, linguística e sociologia estão convergindo em uma abordagem anti-objetivista e incorporada para a cognição abstrata, onde abordagens tradicionais nesses campos defendem uma realidade objetivamente especificada - cérebro na caixa, independente das restrições e interações corporais - essa nova abordagem insiste que inteligência não pode estar separada da experiência subjetiva de um corpo” (BROOKS; STEIN, 1994a).

Mais recentemente, com críticas que seguem a mesma linha de raciocínio, Roy argumenta contra as abordagens acadêmicas que tendem a estudar linguagem e comunicação como fenômenos isolados, e regidos puramente por regras lógicas e sintáticas:

“Há uma tendência no mundo acadêmico de estudar linguagem de forma isolada, como um sistema formal com regras para sentenças bem formadas; ou em focar em como a linguagem se relaciona com notações formais, como a lógica simbólica. Mas a linguagem não evolui como um sistema isolado ou como uma forma de comunicação lógico-simbólica; presumivelmente ela evolui como um mecanismo de troca de informação com o mundo, em última análise promovendo um meio para transmissão cultural entre as gerações” (ROY, 2005).

Nesse contexto, existem vários trabalhos que utilizam simulações que abordam a emergência de algum tipo de comunicação (e linguagem) em uma comunidade de agentes artificiais, buscando responder a uma variedade de questões, desde condições mínimas para o surgimento de linguagem até a razão pela qual um sistema de comunicação confiável pode emergir. O estudo de emergência de comunicação em populações de agentes nasceu praticamente junto com as perspectivas da Nova IA (MIROLI; NOLFI, 2010), e essa comunidade vem crescendo bastante nas últimas duas décadas, e.g. (CANGELOSI; PARISI, 1998; CANGELOSI, 1999; NOLFI; FLOREANO, 2002; VOGT, 2002; STEELS, 2003b; MIROLI; PARISI, 2005; ROY; REITER, 2005; CANGELOSI; RIGA, 2006; FLOREANO et al., 2007; MAROCCO; NOLFI, 2007; MIROLI; PARISI, 2008; AMPATZIS et al., 2008; YONG; MIIKKULAINEN, 2009; GREEFF; NOLFI, 2010; NOLFI; MIROLI, 2010; TUCI et al., 2011), para uma revisão ver (WAGNER et al., 2003). Dessa forma, apesar de simularem um mesmo fenômeno, as questões investigadas podem variar bastante e, por consequência, variam as técnicas e modelos aplicados em cada experimento. Essas variações podem acontecer por uma derivação da forma com que os modelos comentados acima são somados na simulação, colocando-se mais elementos de um ou de outro; mas de maneira geral, o modelo de evolução genética é muito utilizado (seja em conjunto com a aprendizagem iterativa ou com jogos de linguagem).

Seguindo as ideias e técnicas da Nova IA, muitos defendem que os agentes do experimento estejam incorporados e situados em um ambiente real, interagindo e se adaptando a este (VOGT, 2000). Apesar de ser uma ideia amplamente aceita na comunidade, em muitos experimentos é preciso trabalhar com grandes populações de agentes robóticos que devem evoluir durante várias gerações, e a exigência de utilizar agentes físicos pode tornar tal experimento impraticável. Isso ocorre tanto da perspectiva de disponibilidade de recursos de hardware, quanto do espaço físico necessário para a simulação, e também do tempo necessário para as interações e sucessão de gerações. Essa grande quantidade de recursos (financeiro, espacial e temporal) que as simulações com agentes físicos exigem, faz com que os experimentos com agentes simulados sejam também bastante comuns nessa comunidade. Alguns pesquisadores, como Floreano et al (2007), obtiveram bons resultados ao tentarem mitigar os efeitos (da limitação financeira e temporal) da simulação dos agentes simulando a evolução deles no computador, e depois embarcando o resultado em agentes físicos (FLOREANO et al., 2007); segundo esse trabalho, o comportamento dos agentes simulados e dos agentes físicos com o software do agente embarcado são bastante parecidos, o que ajuda a avaliar que, entre prós e contras, os experimentos com agentes simulados podem ser ferramentas igualmente úteis. A divisão entre esses tipos de simulação, entretanto, não é específica dentro dos experimentos com emergência de comunicação, mas de praticamente todos os experimentos com agentes artificiais na IA.

Para implementação de simulações com modelo de evolução genética, várias técnicas da computação evolutiva são utilizadas, dentre elas, os algoritmos evolutivos são os mais utilizados (MIROLI; NOLFI, 2010). A computação evolutiva abrange um conjunto de algoritmos inspirados na teoria evolutiva de Darwin, como algoritmos genéticos e programação evolutiva. Em linhas gerais, em experimentos que utilizam computação evolutiva, acontece o seguinte fluxo de eventos:

- uma população inicial com N indivíduos é definida;
- esses indivíduos possuem um cromossomo, que codifica certas características do indivíduo (geralmente somente os relevantes para os fins do experimento realizado);
- eles interagem, entre si e com o ambiente (real ou simulado), geralmente com o objetivo de cooperar em uma tarefa, ou de competir por algum recurso, durante algum tempo ou iterações;
- os indivíduos da população são avaliados por uma função de *fitness*, que avalia o grau de adaptação dos indivíduos às tarefas que devem ser executadas;
- os melhores indivíduos são selecionados e geram a população da próxima geração, passando para seus descendentes alguma parcela de seus cromossomos;

Esse ciclo se repete até que determinado resultado seja alcançado ou que a simulação chegue a um número máximo de gerações. Para mais detalhes, veja (ZUBEN, 2000).

Esse fluxo, por sua vez, pode ser implementado de várias maneiras, e nas simulações sobre emergência de comunicação com agentes artificiais existe grande variedade e também pontos em comum entre os experimentos. É bastante comum o uso de redes neurais para

o controle dos agentes (JÄGER et al., 2009) e, nesse cenário, os cromossomos codificam os pesos da rede neural que controla o agente, e a população inicial é criada com cromossomos gerados aleatoriamente (para uma introdução ao uso de redes neurais *feedforward* em conjunto com algoritmos genéticos, veja (MONTANA; DAVIS, 1989)). O tamanho da população pode variar muito, indo de alguns poucos indivíduos nas simulações com robôs (algo em torno de quatro), até populações com cem ou até mesmo mil agentes virtuais. As interações entre os agentes podem ser de várias naturezas, desde sinalizações explícitas para indicação de comida e veneno (CANGELOSI; PARISI, 1998; FLOREANO et al., 2007; MIROLI; PARISI, 2008) até a comunicação que inclui sinais implícitos para a realização de ações coordenadas entre os agentes (GREEFF; NOLFI, 2010; MAROCCO; NOLFI, 2007). Greeff e Nolfi (2010), por exemplo, argumentam que a tarefa a ser executada deve ser modesta o suficiente para que agentes artificiais relativamente simples (sem capacidade de cognição de alto nível) possam executá-la e também para que as análises dos resultados possam ser feitas apropriadamente, mas, por outro lado, deve ter complexidade suficiente para que a emergência de comunicação possa ocorrer. Para atender a essas condições os pesquisadores geralmente propõem tarefas que exigem colaboração ou tarefas competitivas.

Para escolher quais indivíduos darão origem à próxima geração, em geral, a avaliação é feita com base no sucesso dos indivíduos em cumprirem uma determinada tarefa, e não com base no sucesso comunicativo, visto que essa é a característica que deve emergir das iterações, durante as gerações. A escolha desses indivíduos normalmente segue uma de duas estratégias de seleção mais comuns, que seriam seleção por indivíduo e seleção por grupo. É comum o uso da seleção por indivíduos em experimentos nos quais a tarefa a ser cumprida pelos agentes é de natureza competitiva, e a seleção por grupo é normalmente utilizada quando há tarefas de natureza cooperativa. O trabalho de Floreano et al. (2007) é um exemplo bastante ilustrativo da utilização dessas estratégias. Nele são utilizados os dois tipos de seleção citados, e ainda outras duas estratégias que mesclam características de ambas, e com isso avalia qual cenário é mais propício para o aparecimento e evolução de um sistema de comunicação confiável entre os agentes (FLOREANO et al., 2007).

Apesar da grande variedade de características, como regra geral nesses experimentos tem-se o compromisso de que os agentes tenham o mínimo possível de comportamentos pré-determinados, principalmente no que diz respeito ao comportamento comunicativo do agente. Em geral, inicialmente os agentes possuem capacidades de emitir ou receber sinais através de seus sensores e atuadores (físicos ou simulados), mas com (pelo menos) algum aspecto do sistema de comunicação não definido, e, a partir dessas características e das regras impostas, podem evoluir o sistema durante o experimento. Ou seja, a partir de uma configuração inicial e das regras formuladas no modelo do experimento, espera-se a emergência de algum tipo de linguagem e comunicação (não pré-determinados) entre os agentes. Essa preocupação com o mínimo de interferência do projetista no experimento é derivado tanto das ideias da Nova IA quanto dos métodos sugeridos para a resolução do problema da fundamentação do símbolo, e prioriza que quanto menor a interferência do projetista, mais significativos podem ser os resultados das simulações.

3.2 TRABALHOS RELACIONADOS

Na comunicação existe algo que é transmitido de um agente emissor para um agente receptor. Aquilo que é recebido pelo agente receptor pode produzir algum efeito nele, uma vez que pode se constituir em uma representação para este. Nas simulações que abordam a emergência de comunicação encontramos vários tipos de análises e suposições, mas, até onde sabemos, muito pouco se abordou sobre as representações e processos semióticos que possam ter ocorrido nos agentes durante essas simulações, uma das motivações para este trabalho. Desse modo, procuraremos nesta seção descrever alguns trabalhos que possam ajudar a entender como o tema representação vem sendo (ou não vem sendo) tratado nesses experimentos, e como podemos contribuir para a pesquisa. Não é intenção ser uma apresentação exaustiva de todos os trabalhos (dentro do escopo de emergência de comunicação) que abordam o tema representação de alguma maneira, mas a intenção é ter uma amostra representativa desses trabalhos. Os critérios para a escolha dos trabalhos foram: proximidade com o propósito deste trabalho, relevância na comunidade (pelos resultados do trabalho e/ou pela reputação dos autores), expressividade quanto à demonstração de características e técnicas comuns a esses experimentos, e estarem entre os mais recentes. Importante lembrar também que, assim como no capítulo 2, ao detalhar sobre cada trabalho, é respeitada a nomenclatura utilizada pelo próprio autor. Essa ressalva é importante principalmente quanto à utilização de termos como símbolo e representação.

Cangelosi (1999) apresenta um experimento que pretende simular os estágios iniciais da evolução da sintaxe, com a combinação de dois símbolos. Em particular, ele estuda a evolução da regra de combinação entre dois símbolos que reflete uma estrutura do tipo “ação-objeto”. O surgimento espontâneo desse tipo de estrutura já foi observado em alguns tipos de primatas, como chimpanzés, em experimentos realizados por biólogos. Ele propõe então uma relação entre a modelagem da evolução de linguagens com uso de sintaxe, sua referência a estudos sobre a linguagem de animais e o problema de aquisição do símbolo. O modelo proposto no experimento tem como alvo a evolução de formas primitivas de uso da sintaxe na linguagem. Para esse estudo da evolução da comunicação foram utilizadas redes neurais evoluídas com um algoritmo genético. O experimento consiste em 80 agentes em um ambiente com 1200 cogumelos. Os cogumelos são divididos em 6 categorias, metade comestível e a outra metade venenosa. Cada agente é controlado por uma rede neural de três camadas. Essa rede neural recebe como entrada um vetor de 18 bits contendo as características do cogumelo e deve decidir se determinado cogumelo é comestível ou não. Ao final de cada iteração os 20 melhores organismos são selecionados para dar origem à próxima geração. Uma das entradas é o nome do cogumelo (composto por 2 símbolos), representada por 8 neurônios. E também representada por 8 neurônios, uma das saídas é a codificação do nome do cogumelo. Com o passar das gerações, a saída simbólica de uma geração anterior passou não só a influenciar o comportamento dos agentes da geração seguinte, como também fazer com que estes retransmissem a mesma identificação para as categorias dos cogumelos. Os resultados mostraram que nas simulações em que uma boa comunicação foi desenvolvida, prevalecia a combinação de símbolos (em 85% das populações). Já nas simulações em que a

comunicação foi considerada ruim, o percentual de populações que usaram combinação de símbolos foi de aproximadamente 50%. Como boa comunicação, o autor quer dizer que a população de agentes foi capaz de criar e utilizar adequadamente a quantidade de símbolos necessários para designar todas as categorias funcionais de cogumelo. Esse fato apresenta a forte tendência da evolução da combinação de símbolos em linguagens que permitem a boa comunicação. De fato, a evolução da sintaxe nos organismos foi bastante similar ao apresentado no experimento com chimpanzés, pois nos dois casos ocorreu uma generalização do símbolo que expressa a ação. O autor ainda argumenta que as redes neurais são uma boa escolha para modelar o problema da fundamentação do símbolo em modelos cognitivos. Há nesse artigo uma discussão interessante sobre a linguagem humana, afirmando que o aprendizado de palavras (símbolos) é caracterizado por uma dupla associação: uma que relaciona a palavra diretamente com algum objeto ou evento no mundo (associação semântica), e outra que associa um símbolo com outro símbolo (associação sintática). Esse fato pode indicar uma tendência evolutiva pelo desenvolvimento de regras sintáticas, que faria crescer de forma exponencial as referências para cada novo símbolo adicionado, devido às possibilidades de combinações entre símbolos. Porém, apesar de parecer haver surgido uma estrutura do tipo “ação-objeto”, não foi discutido o que esses chamados símbolos representariam de fato para os organismos, nem de forma separada nem operando em conjunto.

Wischnann e Pasemann (2006), em um contexto de comportamento minimamente cognitivo, descrito por eles como sendo um comportamento sensório-motor independente do metabolismo, geralmente resultado do acoplamento entre percepção e ação, utilizam redes neurais recorrentes para investigar a emergência de comunicação e cooperação na evolução de um sistema constituído de vários agentes situados. Essa rede neural é descrita como um sistema dinâmico parametrizado, cujos parâmetros são os neurônios, a topologia, os pesos das ligações entre os neurônios, e as entradas externas da rede (os estados do(s) sensor(es) do robô). Justamente por parametrizar também a topologia da rede neural, o trabalho tem como foco diferencial a proposta de análise detalhada da relação entre a comunicação e as propriedades dinâmicas da rede, bem como com mudanças estruturais durante o processo de evolução nas simulações (que utiliza algoritmos genéticos). Em cada iteração, cada robô deve explorar o ambiente a procura de fontes de alimento, evitando colisões com obstáculos e com os outros agentes também presentes no ambiente, e as chances de reprodução são calculadas com base no *fitness* médio do grupo de robôs que participaram da iteração, favorecendo assim o surgimento de um comportamento cooperativo entre os agentes. Durante as execuções do experimento, percebeu-se o surgimento de duas mudanças estruturais completamente diferentes para solucionar a emergência de comunicação, mas com resultados similares para um observador externo. As duas tiveram bom desempenho, mas uma delas (que utilizou picos de sinais sonoros) se mostrou um pouco mais robusta. Os autores argumentam que tais soluções dificilmente seriam encontradas caso houvesse “muita” informação conhecida acerca da função de *fitness* e da topologia das redes neurais recorrentes durante o processo evolutivo, e defendem fortemente a abordagem situada e com agentes com comportamento minimamente cognitivo, para que possa evoluir de forma dinâmica e se adaptar por meio de comportamentos mais complexos (se, e na medida em que essa necessidade for surgindo). Porém, mesmo

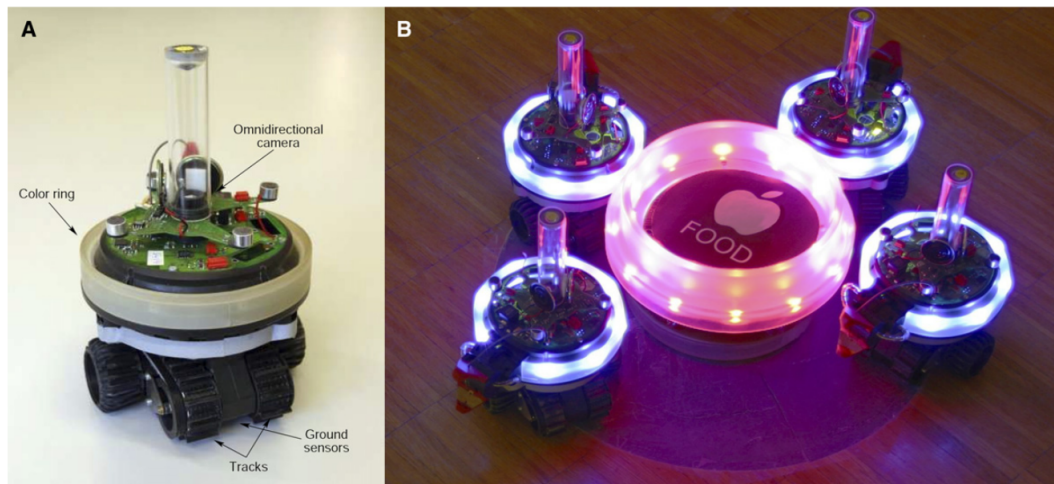


Figura 3.1 Detalhes sobre robôs e cenário utilizados no experimento de (FLOREANO et al., 2007).

fazendo uma análise das mudanças estruturais das redes neurais dos agentes relacionadas a mudanças do ambiente e da tarefa a ser realizada, e verificando a emergência de dois sistemas de comunicação diferentes, não é feita uma discussão sobre o que representariam os sinais emitidos nos sistemas de comunicação que surgiram, nem uma possível relação com as mudanças estruturais analisadas.

Floreano et al. (2007) realizaram experimentos que estudaram as condições de emergência e desenvolvimento de comunicação em uma comunidade de robôs. Os experimentos foram inspirados nos complexos sistemas de comunicação que sociedades animais desenvolveram ao longo de várias gerações, e proposto na tentativa de produzir evidências que permitam o aprofundamento do estudo sobre a origem e desenvolvimento dos sistemas de comunicação. Os experimentos foram realizados utilizando-se agentes que são simulações de robôs reais. Eles são controlados por uma rede neural de duas camadas, cujos pesos entre as conexões são determinados por cromossomos artificiais. A camada de entrada recebe as informações sobre a emissão de luz azul ou vermelha ao seu redor, e sobre a existência de uma fonte de comida ou veneno que esteja a uma distância distinguível. A camada de saída controla dois motores laterais, responsáveis por sua locomoção, e a emissão ou não de luz pelo agente. Cada experimento tem a duração de 500 gerações (durante as quais os cromossomos artificiais serão evoluídos) e utiliza 100 arenas com 10 agentes cada, totalizando 1000 indivíduos. Na arena há uma fonte de comida e uma fonte de veneno, e o objetivo dos agentes é permanecer o máximo de tempo possível próximo à fonte de comida. Ambos, comida e veneno, emitem luz da mesma cor (vermelha), que pode ser vista de qualquer ponto da arena, mas só podem ser discriminados a uma distância bastante pequena. A comunicação entre os agentes acontece apenas por meio da emissão de luz de cor azul (que também poderá ser vista pelos outros agentes de qualquer ponto da arena), permitindo a indicação da localização de uma das fontes. A figura 3.1 mostra alguns detalhes desse cenário e do robô utilizado. Os robôs e o ambiente são simulados em

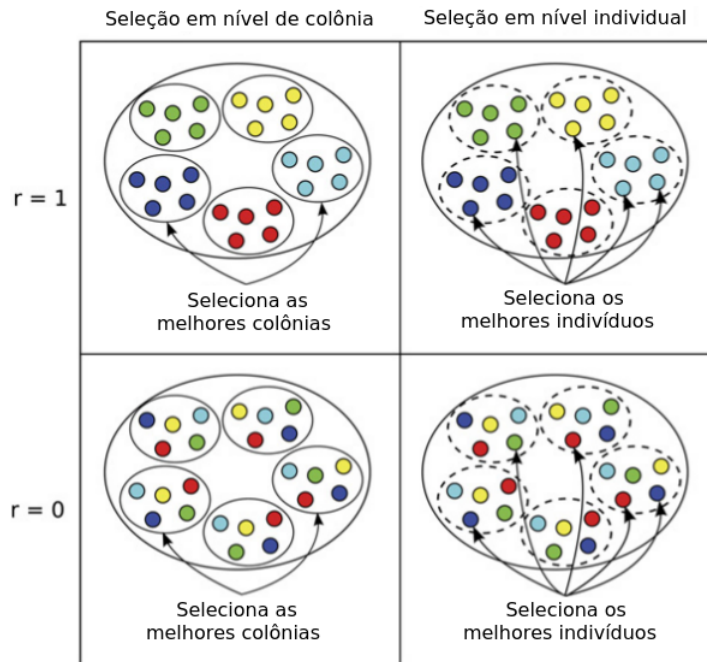


Figura 3.2 Modelo que ilustra as quatro configurações utilizadas nesse experimento de (FLOREANO et al., 2007). $r = 1$ indica que os indivíduos dos grupos são relacionados, e $r = 0$ indica que os indivíduos dos grupos não são relacionados.

computador, e após o final das 500 gerações, as redes neurais que resultaram da simulação são carregadas nos robôs físicos que apresentam desempenho bastante semelhante às simulações, tornando essa solução híbrida (juntando incorporação e virtualização) uma alternativa bastante interessante e viável no escopo desse experimento.

Como ilustra a figura 3.2, o experimento foi realizado utilizando-se duas técnicas de seleção e duas técnicas de parentesco, formando um total de quatro grupos evolutivos:

- **Seleção em nível de grupo com indivíduos relacionados:** nessa configuração, para formar a próxima geração, são selecionados todos os indivíduos dos grupos com melhor performance e, a partir desses indivíduos, são gerados novos grupos, cada grupo com indivíduos geneticamente iguais entre si.
- **Seleção em nível de grupo com indivíduos não relacionados:** nessa configuração, para formar a próxima geração, são selecionados os melhores grupos e, a partir dos indivíduos desse grupo são gerados novos grupos, nos quais os indivíduos são geneticamente diferentes.
- **Seleção em nível individual com indivíduos relacionados:** nessa configuração, para formar a próxima geração, os melhores indivíduos são selecionados, independente do grupo ao qual pertencem e, a partir desses indivíduos, são gerados novos grupos, cada grupo com indivíduos geneticamente iguais entre si.

- **Seleção em nível individual com indivíduos não relacionados:** nessa configuração, para formar a próxima geração, os melhores indivíduos são selecionados, independente do grupo ao qual pertencem e, a partir desses indivíduos, são gerados novos grupos, cada grupo com indivíduos geneticamente diferentes.

A cada geração, independente do nível de seleção, são escolhidos 200 indivíduos que darão origem aos 1000 indivíduos da próxima geração, através de técnicas de cruzamento e mutação. Nesses cenários, basicamente três tipos de sistemas de comunicação aparecem: um onde os agentes tendem a emitir luz azul ao encontrar comida e, conseqüentemente, são também atraídos pela luz azul; outro onde prevalece a estratégia enganosa, na qual os agentes tendem a emitir sinal de luz azul longe da fonte de comida e, conseqüentemente, sendo repelidos por essa sinalização; e por fim um sistema curioso, no qual, apesar de ser adotada a estratégia enganosa, os agentes continuam sendo atraídos pela sinalização dos outros agentes. Esse último sistema, que mostrou-se pouco eficaz, emergiu no grupo de experimento com seleção em nível individual com indivíduos não relacionados. Há uma discussão sobre esse comportamento, que indica uma possível causa nos primeiros estágios da seleção. Já os outros dois sistemas acontecem nas outras três configurações de grupo, com mais eficiência no grupo com seleção em nível de grupo e com indivíduos relacionados entre si. A estratégia de indicar a fonte de comida foi a mais utilizada e a que rendeu melhor desempenho aos agentes, apesar de intrinsecamente trazer questões de competição, como por exemplo pelo espaço em volta da comida, que não comporta os 10 agentes ao mesmo tempo. Apesar disso, nota-se que a tendência é que um sistema de comunicação confiável apareça, principalmente quando os indivíduos são relacionados entre si.

Apesar de apresentar vários dados dos experimentos, em nenhum momento se discute representação, ou que tipo de representação seriam as luzes para os agentes. Foi analisada a correlação entre nível de seleção e de relação entre os indivíduos, e as respostas destes (movimentação e acionamento da luz azul) aos estímulos externos (luz vermelha das fontes de veneno e comida, e luz azul dos outros agentes), mas não há uma discussão de como a utilização da comunicação auxilia os agentes no seu sucesso evolutivo.

Marocco e Nolfi (2007) fazem um experimento com robôs simulados, em que estes devem cooperar para a execução de uma tarefa. O principal objetivo é identificar como uma população de agentes equipados com sistemas de sensoriamento e motores incorporados pode desenvolver um sistema de comunicação fundamentada. Foram realizados dois experimentos, com pequena variação da rede neural utilizada, em que apenas uma tinha camada intermediária com dois neurônios, que são também ligados entre si. A figura 3.3 mostra o modelo das redes neurais utilizadas no experimento. Essa rede tem a peculiaridade de que o agente pode “ouvir” o que disse na iteração anterior, mantendo assim uma memória. O experimento é realizado em uma arena que possui duas áreas-alvo no chão, e com quatro agentes atuando por vez. A tarefa então consiste em que os robôs se dividam igualmente entre essas áreas, ou seja, que fiquem exatamente dois robôs em cada uma das áreas-alvo. A figura 3.4 mostra o esquema do experimento, com as áreas-alvo e os agentes. Os robôs podem emitir diferentes sinais e se locomover, além de poder capturar informações do ambiente, incluindo os sinais emitidos por outros agen-

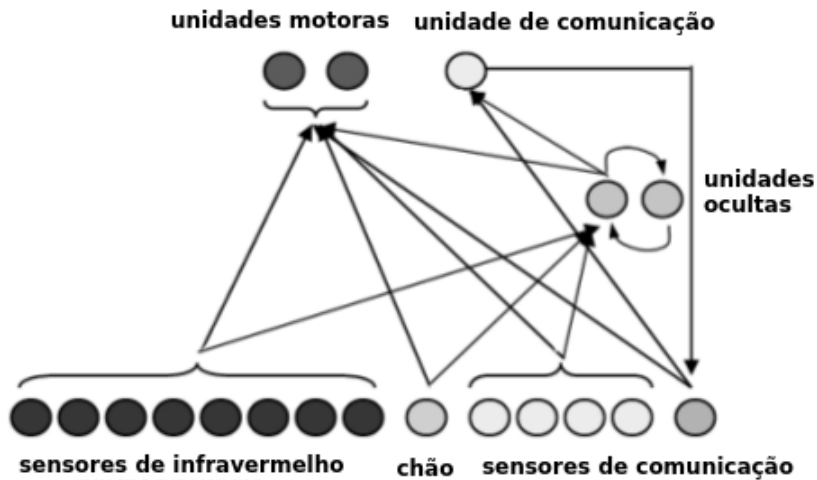


Figura 3.3 Modelo da rede neural utilizada no experimento de (MAROCCO; NOLFI, 2007).

tes. A tarefa proposta admite várias soluções, e os melhores robôs (que darão origem à próxima geração) são escolhidos com base na sua habilidade de resolver o problema, e não com base na qualidade da comunicação alcançada, deixando assim o caminho livre para que as circunstâncias de uso moldem o desenvolvimento da comunicação. Os resultados do experimento mostram que um sistema efetivo de comunicação emergiu entre os agentes. Para desenvolver esse sistema, os agentes se basearam em informações como localização e número de agentes dentro de uma área-alvo, dentre outras. Essas nuances foram sendo descobertas pelos agentes durante o experimento através de suas experiências sensorio-motoras diferentes, expressando características importantes para a execução da tarefa proposta. Foram desenvolvidas pelos agentes tanto habilidades sociais, de comunicação (como modalidades de comunicação: unidirecional, bi-direcional, e até mesmo entre 3 agentes), quanto individuais, (como uma melhor exploração do ambiente, passagem de obstáculos, etc.), o que pode sugerir uma interdependência entre essas duas características. Há uma extensa discussão acerca dos sistemas de comunicação emergidos durante as simulações, assim como em quais circunstâncias determinados sinais eram emitidos. Entretanto, pouco se fala sobre a fundamentação da comunicação, e não há nenhuma discussão sobre o que os sinais surgidos durante a emergência da comunicação representam para o sistema, ou se de fato uma comunicação através de representações simbólicas emergiu.

Greff e Nolfi (2010), investigaram as condições nas quais uma população de agentes incorporados evoluem para apresentar habilidades de cooperação e coordenação. Investigam nesse contexto, como sistemas inicialmente não comunicativos podem desenvolver habilidade comunicativa, se e em que medida o sistema de comunicação fica mais complexo no curso do processo evolutivo, e também como algumas características de tal sistema comunicativo variam evolucionariamente. O cenário do experimento foi desenvolvido para ser simples o suficiente para ser sistematicamente analisado, mas ao mesmo tempo “com todos os elementos que são necessários” para investigar importantes aspectos

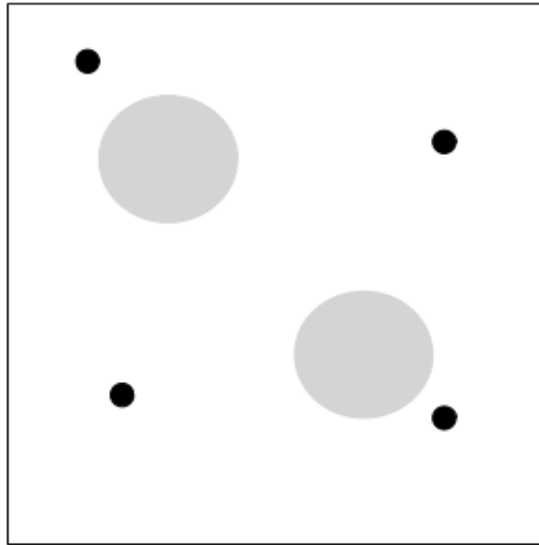


Figura 3.4 Esquema do experimento de (MAROCCO; NOLFI, 2007), com as duas áreas-alvo, e quatro agentes na arena por vez.

tos da emergência de comunicação, como: as condições que fazem emergir comunicação em uma população inicialmente sem essa habilidade; e como o ‘significado’ dos sinais se originam e evoluem. O experimento consiste em ter dois agentes em uma arena que possui duas áreas-alvo, e a tarefa dos agentes é de trocar de área-alvo com o máximo de frequência possível. Os agentes possuem sensores de infravermelho ao redor do corpo, sensores para identificar cada área alvo, uma câmera de vídeo frontal, um receptor e um atuador para comunicação sem fio, e dois motores para suas rodas. A figura 3.5 mostra a arena do experimento e alguns detalhes do robô utilizado. Os agentes são controlados por uma rede neural de três camadas, sendo uma camada interna, com conexões recorrentes, como mostra a figura 3.6. Como mecanismo de evolução, é utilizado um algoritmo evolutivo com estratégia seletiva elitista, com escolha dos 20% melhores indivíduos para reprodução. Os resultados indicam que um sistema de comunicação confiável surgiu, com o agente sendo capaz de produzir e perceber sinais diferentes (incluindo sinais implícitos) de forma regular, e de reagir apropriadamente para cada sinal recebido. Nesse trabalho há uma análise detalhada do comportamento motor e comunicativo do agente, o repertório lexical desenvolvido pelos agentes, bastante discussão sobre o surgimento e evolução das habilidades comunicativas, e das várias estratégias desenvolvidas para a execução da tarefa. Porém, não há discussão sobre o que esses sinais representariam para os agentes ou como os robôs os interpretam.

Mirolli (2012), não trata especificamente da emergência de comunicação, mas a sua proposta de análise da emergência de representação em um agente artificial se aproxima da discussão que propomos realizar, sendo então bastante relevante para essa pesquisa. Mirolli faz alguns experimentos com base nas ideias de Randal Beer sobre comportamento minimamente cognitivo em termos dinâmicos (BEER, 2003), e tenta complementar seus resultados investigando se, quando, e em quais condições agentes minimamente cogniti-

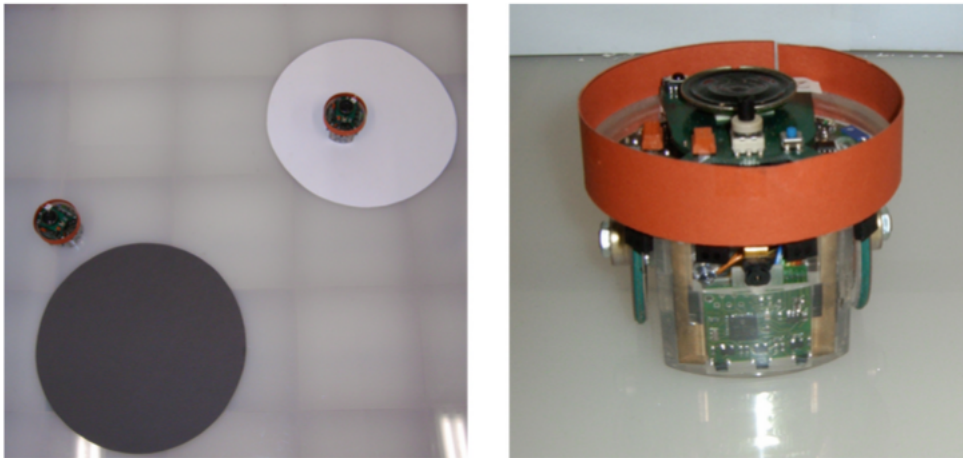


Figura 3.5 Detalhes da arena e do robô utilizados no experimento de (GREEFF; NOLFI, 2010).

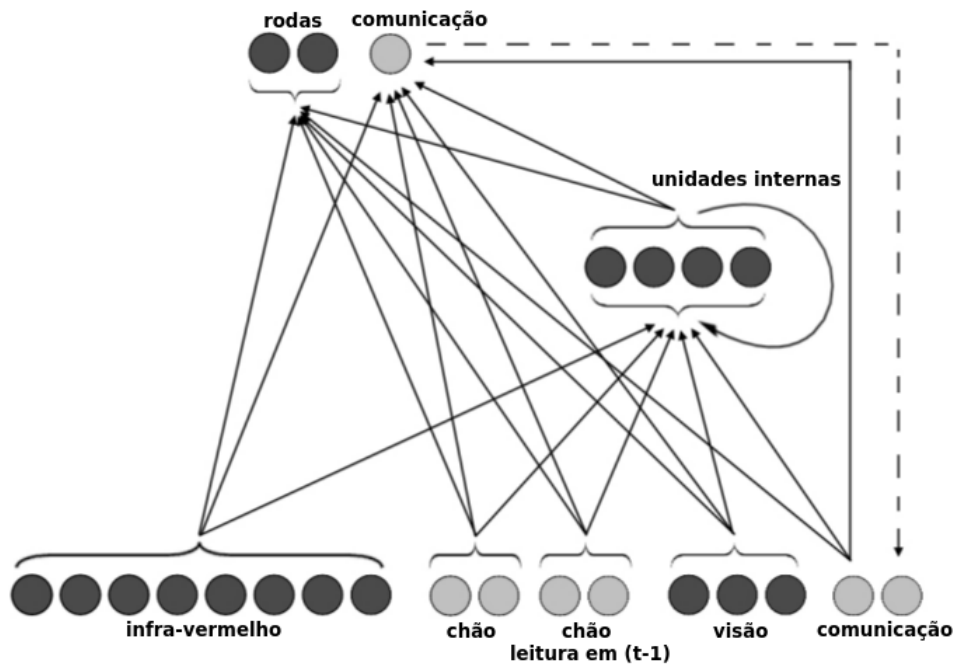


Figura 3.6 Modelo da rede neural utilizada no experimento de (GREEFF; NOLFI, 2010).

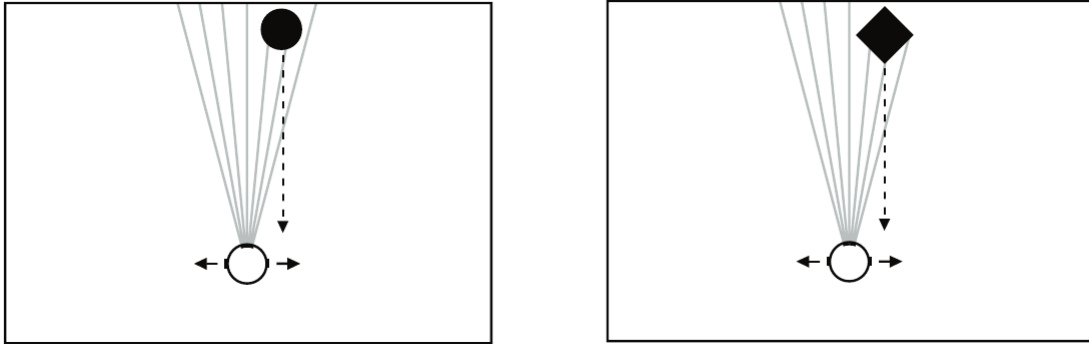


Figura 3.7 Esquema do experimento em (MIROLI, 2012), com objetos de dois tipos caindo, e o agente podendo se mover lateralmente.

vos utilizam representação. Mirolli argumenta que apesar de ser uma das propostas do trabalho de Beer, este não avalia de forma adequada a presença de representação nos agentes de sua simulação, e que isso pode ocorrer devido ao fato de que muitos críticos e céticos do representacionalismo (o próprio Beer, em (BEER, 2014), se declara entre os cético) consideram a ideia de representação interna muito vaga e mal definida para que se faça uma investigação pragmática. Mirolli afirma então que uma investigação mais profunda revela que um conceito definitivo ainda precisa ser obtido, mas também revela um conjunto de consensos a respeito da ideia de representação, o qual pode dar suporte à busca por representações no trabalho proposto em (BEER, 2003), e o é que ele se propõe a fazer.

O experimento simulado consiste em deixar um agente de formato circular em uma arena, e esse agente pode se mover apenas horizontalmente. A uma determinada distância, verticalmente, é colocado um objeto, que pode ser um círculo ou um diamante, que cai a uma determinada velocidade. A figura 3.7 ilustra essa dinâmica. O agente então deve evitar os diamantes e "pegar" os círculos. Quanto maior a distância horizontal entre os centros do objeto e do agente quando a distância vertical for zero, melhor, caso o objeto seja o diamante, e pior, caso o objeto em questão seja um círculo. A população é de cem agentes, controlados por uma rede neural do tipo *feedforward* com 3 camadas (para mais detalhes sobre redes neurais do tipo *feedforward*, veja (BEBIS; GEORGIPOULOS, 1994)). A figura 3.8 mostra o modelo da rede neural utilizada nesse experimento. A camada de entrada tem sete neurônios, e cada um recebe uma informação referente a distância entre o agente e o final de um raio, que sai do agente e tem tamanho máximo de 220 unidades. Esses sete raios estão distribuídos numa abrangência de ângulo de $\pi/6$. A camada intermediária tem 3 neurônios e é completamente recorrente, e os 2 neurônios da camada de saída controlam a movimentação (horizontal) do agente. Foram realizadas cinquenta simulações, nas quais os cem agentes são evoluídos com algoritmos genéticos durante duzentas e cinquenta gerações. Para produzir a próxima geração são escolhidos os vinte melhores indivíduos da geração anterior, fazendo-se cinco cópias de cada, com taxa de mutação de 3% (com exceção do melhor indivíduo, que é mantido sem modificações).

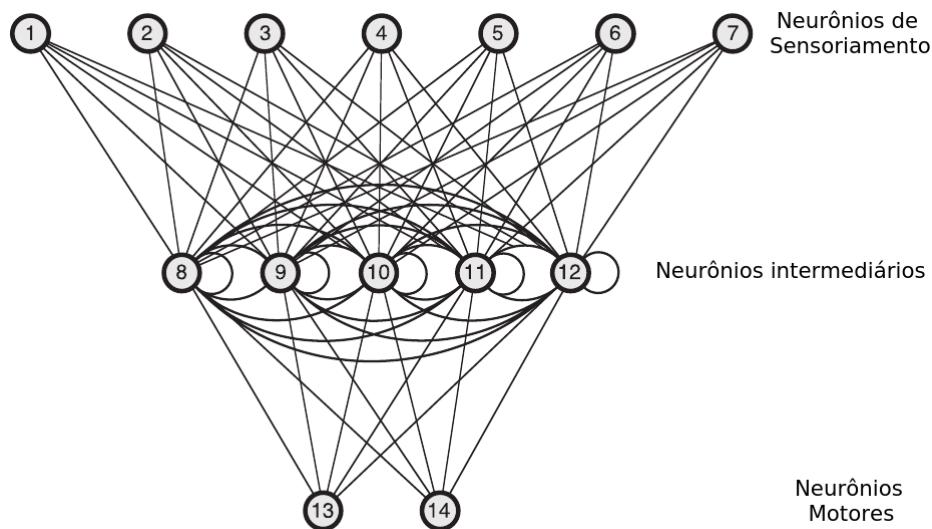


Figura 3.8 Modelo da rede neural utilizada no experimento de (MIROLI, 2012).

Com o experimento recriado, a busca por representação nos agentes se baseia em dois princípios básicos definidos pelo autor. A representação deve ser algo que está: (a) correlacionado com alguma característica do ambiente relevante para o agente, e (b) que contribua para o comportamento do agente de forma que seja funcional (adaptativa) em relação à característica a qual está correlacionada. A correlação de (a) deve ser entre o ambiente e algum estado interno do agente, desde que os neurônios da camada de entrada e da camada de saída constituem a interface do agente, assumiu-se o estado interno do agente como o estado das ativações da camada interna da rede. Inicialmente foi possível verificar a ocorrência de (a), ou seja, determinado padrão de ativação ocorre sempre que determinado objeto está presente no ambiente (e pode ser captado pelos sensores do agente), porém, (b) não ocorre. Logo, esse estado interno, apesar de se correlacionar com características do ambiente, não pode ser considerado uma representação, segundo a definição estabelecida.

Nesse ponto foi analisada a real necessidade de nesse experimento, o agente possuir algum tipo de representação interna. Para verificar tal necessidade, Mirolli retirou a camada intermediária, ligando diretamente os 7 neurônios da camada de sensoriamento aos 2 neurônios da camada motora, tirando dessa forma qualquer possibilidade de estado interno. Os resultados semelhantes à simulação anterior atestam que a tarefa proposta, apesar de apropriada para a ciência cognitiva no que diz respeito à categorização, não é apropriada para verificar a existência de representação. Mirolli faz então uma mudança na tarefa proposta. Nessa nova proposta os agentes precisam possuir alguma memória, pois devem ir o máximo para esquerda quando houver um círculo, e o máximo para a direita quando houver um diamante. Isso requer que eles continuem agindo adequadamente mesmo quando o objeto não puder mais ser percebido por eles (quando a distância entre eles for maior do que as 220 unidades do raio). Para ter certeza de que essa tarefa neces-

sitaria de uma representação interna, inicialmente foi simulada com uma rede também sem camada intermediária, e os resultados ruins (com *fitness* final praticamente com a metade do valor das simulações anteriores) atestam que essa é uma tarefa que necessita de representação interna para ser realizada com sucesso, e que pode ser utilizada para investigar a existência de representação. Voltando a utilizar a rede neural original (com a camada intermediária completamente recorrente), Mirolli avalia mais uma vez a presença de uma representação interna, e dessa vez acha algo interessante. Primeiramente, ao contrário dos resultados com redes de duas camadas, os resultados são bons, similares aos encontrados no primeiro experimento. Depois ao forçar o valor da camada intermediária para determinado objeto, antes deste estar presente no ambiente, o agente se comporta exatamente como se aquele objeto estivesse no ambiente e, por consequência, consegue um *fitness* ainda mais elevado (pois antes mesmo do objeto estar presente o agente já inicia a ação apropriada, sem precisar esperar). Esse fato completa os princípios básicos utilizados por Mirolli para estabelecer a presença de representação no agente.

O método utilizado por Mirolli aqui, tem significativas semelhanças com o que propomos nesse trabalho. Ele avalia um trabalho proposto por um pesquisador da comunidade envolvida no estudo do processo de representação em agentes artificiais, propõe uma definição para representação, replica o experimento do trabalho avaliado, faz então as buscas por representação de acordo com a definição formulada previamente e altera o experimento original a fim de testar hipóteses. Apesar dessas semelhanças, existem várias diferenças (e.g. pretendemos avaliar o processo de representação dentro do contexto de emergência de comunicação), sendo a mais fundamental dada não pelo método, mas pela definição de representação utilizada por Mirolli.

Em (LOULA; GUDWIN; QUEIROZ, 2013), é enfatizado o fato de que na comunicação há tanto a produção (pelo emissor) quanto a interpretação (pelo intérprete) de representações, e que apesar da metodologia de modelagem e simulação computacional de fenômenos relacionados a linguagem e comunicação (como é o caso das simulações sobre emergência de comunicação) ter se consolidado nos últimos anos, há, em geral pouca discussão sobre esses processos representativos presentes nesses experimentos. Utilizando conceitos da teoria de Peirce, realizam então uma análise em um experimento realizado por eles, em busca de representações indiciais e simbólicas. O experimento foi organizado da seguinte forma: em um ambiente de 50 por 10 posições, são colocados um recurso e cinco agentes escolhidos de forma aleatória. Existem dois tipos de recursos, um com efeito positivo e outro com efeitos negativos para os agentes que os consumirem. Os agentes ganham pontos se ficarem próximos à fonte positiva, e perdem pontos se ficarem próximos ou, mesmo distantes, não se protegerem da fonte negativa. Os agentes, que são potenciais emissores e interpretadores de representações, são controlados por rede neural, que podem ter arquiteturas diferentes (veja figura 3.9). A arquitetura 1 seria utilizada para interpretação indexical, uma vez que o interpretante é levado a identificar seu objeto como algo espaço-temporalmente relacionado com o signo. Já a arquitetura 2 tem a adição de uma camada (RD2), que habilitaria os agentes a realizarem interpretações simbólicas, visto que apenas através dela seria possível estabelecer associação entre signo e objeto, já que essa associação deve ser realizada através de hábito ou lei. Nessa rede neural, as camadas internas (tanto visual quanto auditiva) só podem ativar um neurônio

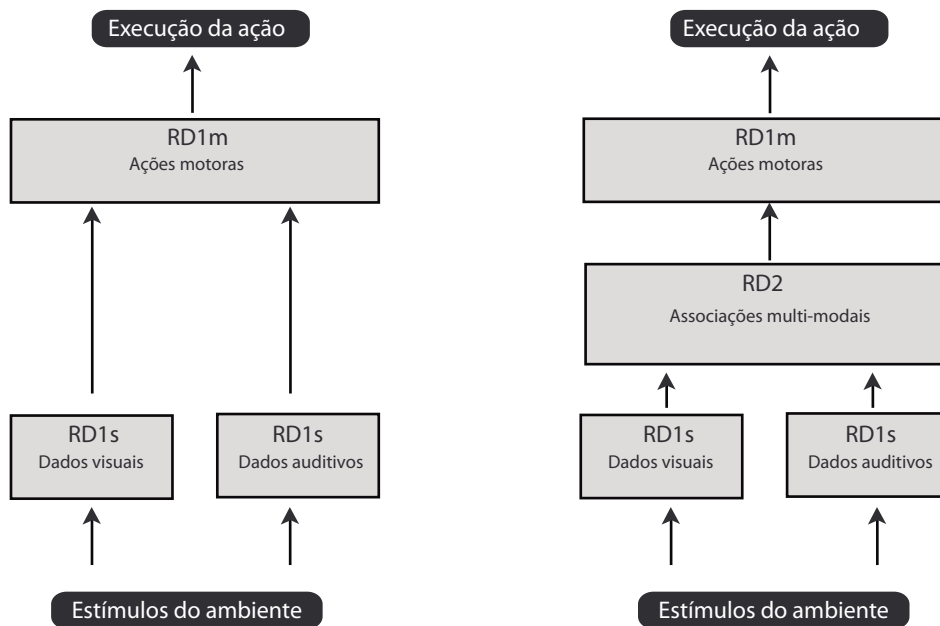


Figura 3.9 Modelo da arquitetura cognitiva utilizada no experimento de (LOULA; GUDWIN; QUEIROZ, 2013).

por vez, seguindo a estratégia conhecida como “o vencedor leva tudo”, na qual o neurônio com maior ativação é disparado, enquanto os outros permanecem inativos. Os agentes podem perceber “visualmente” os recursos a 2 posições de distância, enquanto podem “ouvir” sinais emitidos a até 25 posições de distância. Os sinais emitidos pelos agentes já eram parcialmente pré-determinados, visto que de uma sequência de três bits os dois primeiros seriam ou '10' ou '01', o outro sendo gerado aleatoriamente. Ao todo no experimento existe uma população de quinhentos agentes, que são divididos em cem grupos de cinco, e um indivíduo de cada um dos 20 melhores grupos é escolhido para gerar a próxima população, através de *crossover* e mutação. Caso haja divergência entre arquiteturas dos agentes no *crossover*, a arquitetura do tipo 1 prevalece, e as mutações podem modificar a arquitetura da rede neural, com mais chances de mudança do tipo 1 para o tipo 2 do que do contrário.

Nessas condições, foram realizadas duas simulações com condições um pouco diferentes, e que trouxe resultados interessantes. O primeiro cenário é justamente a configuração já mencionada, e ocorre a emergência de uma comunicação indicial, já que ao ouvirem algum sinal, os agentes tendem a ir em direção ao emissor do sinal (comportamento de fonotaxis), e dessa forma pode perceber e identificar o tipo de fonte. Nesse caso, o tipo de arquitetura 1 predominou nos agentes durante toda a simulação. Na segunda simulação, a diferença de condições foi o aumento de custo para o aprendizado de coordenação motora para os agentes. Nesse cenário, não bastava ser o mais ativo de sua camada intermediária para ser disparado. Um neurônio só seria disparado se sua ativação estivesse um determinado valor acima do segundo neurônio mais ativo. Com isso, surgiram estratégias de

reaproveitamento de mapeamento, e no final da simulação prevaleceu a arquitetura do tipo dois entre os agentes. A arquitetura do tipo 2 foi utilizada para fazer um mapeamento da interpretação do sinal ouvido com o que já estava mapeado na camada visual, utilizando a entrada auditiva para tomar decisões “como se” tivesse ocorrido uma entrada visual (e nesse contexto não importa se houve erro no mapeamento, o fato mais importante é que a interpretação está sendo feita pelo agente), estabelecendo assim um sistema de comunicação simbólica na população.

Nesse experimento também encontramos algumas relações diretas com o trabalho proposto. É utilizada a teoria semiótica de Peirce, e são investigados os possíveis processos semióticos que surgem durante as simulações realizadas. Porém, podemos perceber que de maneira oposta à maioria dos trabalhos na área, há características comunicativas pré-estabelecidas (e.g. a existência de dois sinais, assim como parte do formato desses sinais) e, além disso, a análise é realizada em experimento desenvolvido pelos próprios autores, e o que pretendemos fazer aqui é uma análise em experimento proposto por outros autores, estabelecendo assim um diálogo com essa comunidade.

Podemos perceber que, de maneira geral, os trabalhos têm abordado várias perspectivas no estudo da emergência da comunicação. Essa diversidade é de fato esperada, justamente devido às características de simplificação e especificidade dos modelos em geral e, nesse caso específico, de modelos gerados para serem executados em uma simulação computacional (como discutimos acima). Entretanto, o fato de que a análise de processos de representação que podem existir nesses experimentos ser realizada e discutida em tão poucos trabalhos, indica que ainda apesar da diversificação, ainda há pontos e perspectivas a serem exploradas nesses experimentos.

REPLICAÇÃO DO EXPERIMENTO

4.1 ESCOLHA DO EXPERIMENTO REPLICADO

A proposta principal desse trabalho é realizar o estudo e a análise de processos de representação na emergência de comunicação em uma população de agentes artificiais replicando um experimento computacional previamente proposto. Pretendemos assim estabelecer diálogo com a comunidade envolvida nesse tipo de experimento, indicando a necessidade de mais discussões sobre esse tema, e produzir um *framework* que possa ajudar outros pesquisadores a realizarem análises similares. Esperamos com isso colaborar para um melhor entendimento do processo de representação no contexto desses experimentos, trazer resultados que possam apontar novas formas de tratar o processo de emergência de comunicação e, evidenciar como novas discussões e resultados podem ser adicionados aos trabalhos já realizados, e também como novos experimentos podem ser complementados com outros tipos de análises. Sendo assim, o primeiro passo de nosso trabalho foi escolher um experimento para ser replicado e analisado, e essa escolha teve que atender a alguns critérios para que pudéssemos ser capazes de gerar os resultados esperados.

No capítulo 3 vimos que redes neurais artificiais e algoritmos evolutivos são amplamente utilizados (e muitas vezes em conjunto) em experimentos que abordam a emergência de comunicação em uma população de agentes artificiais autônomos. Logo, seria razoável escolhermos um experimento que utilize essas duas técnicas. A utilização de redes neurais artificiais para controlar os agentes, além da representatividade de um grande espectro desses experimentos, traz consigo um desafio metodológico. Na literatura as redes neurais são muitas vezes descritas como “caixas pretas”, nas quais apenas as entradas e as saídas são verificáveis, e essa parece ser uma das maiores dificuldades para uma análise mais detalhada desses experimentos. De fato, recentemente vem havendo certa preocupação com a possibilidade de analisar mais detalhadamente os experimentos, e alguns pesquisadores vem adotando, dentre outras medidas, a diminuição da complexidade da arquitetura dos agentes existentes no experimento, justamente para possibilitar análises mais minuciosas (JÄGER et al., 2009). Dessa forma, é prudente escolhermos um experimento que não seja

trivial, mas também que seja simples o suficiente para que possamos replicá-lo e analisar detalhadamente os processos emergentes.

Por outro lado, entendemos, em acordo com a literatura, que quanto mais condições pré-estabelecidas, mais enviesado estará o experimento, principalmente em casos de pré-definições de características e comportamentos relacionados à comunicação. Esse enviesamento pode dificultar a procura pelo surgimento de algum tipo de comportamento semiótico não existente previamente, além de prejudicar também a procura pelas causas do aparecimento desse comportamento (caso ocorra). Por consequência, entendemos que o experimento escolhido para ser replicado, não deve ter tantas características pré-definidas de modo que não permita a verificação da emergência de um sistema de comunicação.

O experimento realizado por Mirolli e Parisi (2008) atende aos requisitos relacionados acima, sendo o experimento escolhido para ser replicado, modificado e analisado em detalhes, na busca por processos representativos nos agentes artificiais que fazem parte da população do experimento. Primeiro, o experimento utiliza redes neurais para controlar seus agentes, que são evoluídos através de um algoritmo genético, envolvendo *crossover* e mutação. Como já discutimos, essas são características bastante comuns, fornecendo assim grande representatividade ao experimento. A rede neural utilizada é do tipo *feedforward*, na qual a informação sempre flui em uma única direção, e possui três camadas. Por um lado, a existência de três camadas oferece possibilidades de análises que não existem em redes de duas camadas. De fato, a análise das ativações dos neurônios da camada intermediária foi um dos principais métodos utilizados nesse trabalho, como veremos nas seções seguintes, e o tipo de análise que realizamos não seria possível com uma rede neural com menor número de camadas. Por outro lado, redes do tipo *feedforward* evitam algumas complexidades extras das redes neurais recorrentes (que, como vimos no capítulo anterior, também são utilizadas nesse tipo de experimento), como, por exemplo, a análise do efeito acumulado das iterações passadas sobre o comportamento presente da rede. Quanto a representatividade do experimento, trata-se de um trabalho com inspirações na biologia, característica que vem sendo bastante utilizada não apenas nesses experimentos, mas também por grande parte da comunidade da IA. Além da inspiração, a simulação é guiada por princípios teóricos da biologia e, a utilização conjunta da simulação computacional com outras ferramentas e *frameworks* teóricos. E por último, foi possível obter todos os dados necessários para replicar a simulação. Para isso, além do trabalho de 2008, utilizamos dados de um trabalho anterior desses autores (MIROLLI; PARISI, 2005), que fornece informações complementares ao trabalho de 2008 sobre a implementação do mesmo experimento.

4.2 DESCRIÇÃO DO EXPERIMENTO ORIGINAL

Nessa seção iremos descrever as motivações de Mirolli e Parisi para realizarem o experimento, detalhar o cenário utilizado, com todas as variáveis envolvidas, além de apresentar os resultados obtidos por esses pesquisadores. Importante explicitar que durante essa seção as informações, definições e hipóteses apresentadas sobre o experimento foram retiradas dos próprios trabalhos dos autores, (MIROLLI; PARISI, 2005) e (MIROLLI;

PARISI, 2008), a menos que o contrário seja dito de forma explícita.

4.2.1 Motivações

Mirolli e Parisi (2008) levantam questões relacionadas ao surgimento da comunicação sobre a perspectiva da filogenética, propõem hipóteses para responder as questões, e realizam um experimento para testá-las. A filogenética é uma das quatro perspectivas definidas por Tinberge, através das quais podemos entender o comportamento animal (TINBERGEN, 1963). Nessa perspectiva é investigado como determinado comportamento surgiu e evoluiu dentro da história das espécies. Seguindo essa linha, o desenvolvimento do artigo toma como base a questão: como o comportamento comunicativo se desenvolveu no decorrer da história das espécies? Mais especificamente: como falantes e ouvintes puderam co-evoluir? Essas questões não são direcionadas a nenhuma espécie em específico, mas é genérica, pois o principal interesse é o estudo dessas questões de forma geral.

De fato, para que haja comunicação devem existir simultaneamente bons falantes e bons ouvintes, pois de outro modo não influenciaria na sobrevivência e reprodução dos indivíduos. Isso significa que ambas as características devem evoluir simultaneamente nos indivíduos de uma população. Bons falantes são indivíduos que produzem sinais que se correlacionam com situações e eventos relevantes para sobrevivência e reprodução. Para situações e eventos iguais, o bom falante emite os mesmos sinais. Já os bons ouvintes são indivíduos que reagem de forma apropriada às situações com as quais os sinais recebidos possuem covariância. Para sinais iguais, indivíduos que são bons ouvintes irão reagir da mesma forma (apropriada à situação). Bons falantes sem bons ouvintes são inúteis, pois não há indivíduos para reagirem apropriadamente aos sinais emitidos. Por outro lado, a existência de bons ouvintes sem bons falantes não faz sentido, pois não há a produção de sinais correlacionados com situações e eventos para que o bom ouvinte possa reagir de forma apropriada.

Esse entrelaçamento de dependências torna a emergência de comunicação uma questão bastante interessante sob a perspectiva da filogenética. Como pode emergir um sistema de comunicação que não beneficia igualmente a ouvintes e falantes? Por que emitir um sinal que apenas beneficiará o ouvinte (como sinais que avisam a existência e a localização de alimentos, ou que avisam sobre a presença de predadores nas proximidades)? Ou por que responder adequadamente a um sinal se apenas beneficiará ao emissor (como sinais agressivos emitidos durante a disputa por algum recurso, como briga por comida ou pelo acasalamento, ou sinais de filhotes que tentam chamar a atenção dos pais para ganharem mais alimentos)? Esses são alguns dos principais questionamentos feitos por Mirolli e Parisi (2008).

Segundo os autores, a literatura da biologia recente sugere que o comportamento comunicativo consiste essencialmente da manipulação do ouvinte pelos falantes. Essa hipótese encontra problemas quando confrontada com cenários nos quais a comunicação parece favorecer mais ao ouvinte, como é o caso de aviso de predadores ou de presença de comida. Uma possível explicação para o problema filogenético é encontrada na literatura da biologia através da ideia de exploração de uma pré-disposição do ouvinte.

Essa pré-disposição pode ser, por exemplo, um comportamento disparado por alguma característica do ambiente, e que tem um valor adaptativo para este ouvinte. O falante, por sua vez, produzirá sinais com objetivo de explorar essa pré-disposição já existente. Essa explicação parece adequada apenas para os sistemas de comunicação que favorecem prioritariamente os falantes. Porém, o mecanismo oposto pode ser utilizado para explicar os sistemas de comunicação que favorecem principalmente os ouvintes: estes tiram proveito da pré-disposição do falante. Se determinado organismo tem a pré-disposição em produzir determinados sinais e/ou comportamentos que são sistematicamente correlacionados (espaço-temporalmente) com alguma característica do ambiente que tem algum valor adaptativo para o receptor, esse comportamento pode ser utilizado como sinalização e dica para seus próprios propósitos. É sugerido por alguns trabalhos recentes que essa pré-disposição dos falantes em produzir sinais coerentes está relacionada com a necessidade adaptativa de categorizar experiências internamente. O experimento simulado corrobora com a ideia da pré-disposição do falante em emitir sinais apropriados (como bons falantes), baseado na necessidade de categorização interna de suas experiências. Na simulação proposta por Mirolli e Parisi (2008), é testada justamente essa hipótese.

4.2.2 Cenário do experimento

O cenário do experimento consiste em um corredor unidimensional com 12 posições. Nele são posicionados um cogumelo e duas criaturas artificiais. Existem 420 cogumelos diferentes, cujas características são representadas por um vetor de 10 bits, com valores -1 ou +1. Metade dos cogumelos é comestível e outra metade é venenosa, e é essa informação que esses bits codificam: se o cogumelo for comestível os dois bits menos significativos terão valor 1, caso contrário terão valor -1, o restante dos bits são valores binários ordenados entre 0 e 210, para ambos os tipos de cogumelo. Essa propriedade do cogumelo apenas são perceptíveis à criatura quando este estiver em uma posição do corredor imediatamente próxima ao cogumelo.

Em uma população de 100 criaturas, duas são utilizadas a cada rodada do experimento. Uma das criaturas é posicionada em uma das extremidades do corredor, oposta à posição do cogumelo, impedindo assim qualquer tipo de recebimento de informação visual. A partir do recebimento de um sinal comunicativo, essa criatura deve decidir ir em direção ao cogumelo ou ficar parada. Essa é a criatura ouvinte da rodada, e a cada geração todas as criaturas passam por essa função uma vez, pois é nela que a avaliação do desempenho da criatura (coleta de dados para o cálculo do *fitness*) ocorre. A outra criatura é justamente a que tem a função de falante. Ela é escolhida aleatoriamente, e colocada na extremidade oposta, imediatamente próxima ao cogumelo presente no ambiente, de modo que possa perceber suas informações visuais. Essa criatura não pode mover-se, e apenas emite sinais comunicativos que serão recebidos pelo agente ouvinte da rodada, baseada na percepção visual do cogumelo em sua vizinhança. A figura 4.1 mostra o modelo de cada rodada do experimento.

Para cada indivíduo esse processo é repetido 420 vezes, uma para cada cogumelo. Sendo a população de 100 indivíduos e havendo 420 cogumelos, temos, a cada geração um total de 42000 rodadas de avaliação. Cada rodada possui, no máximo, 11 instantes

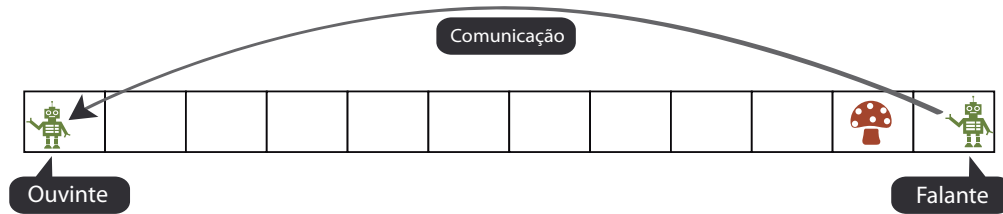


Figura 4.1 Modelo esquemático do experimento.

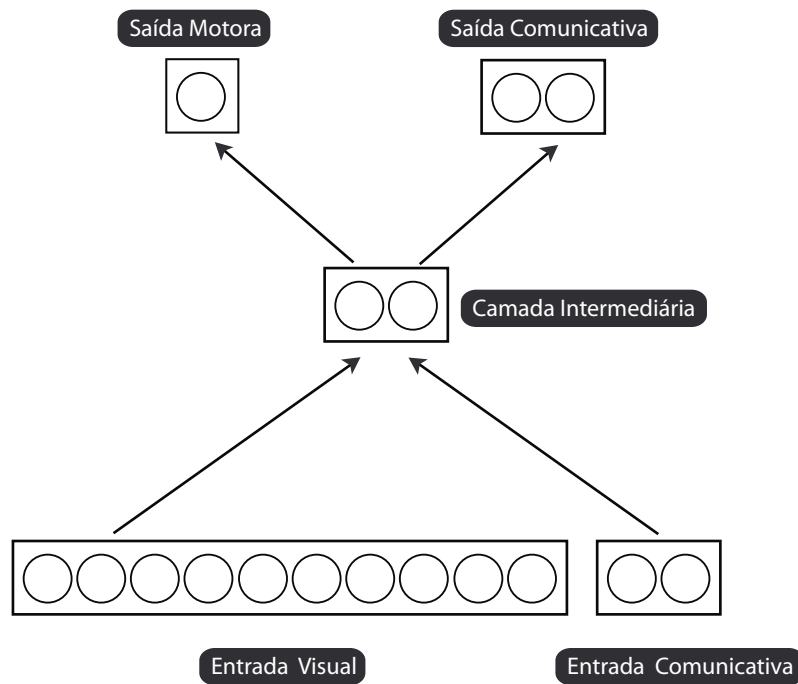


Figura 4.2 Arquitetura da rede neural que controla os agentes do experimento.

de duração (em que o falante emite um sinal comunicativo, que é recebido pelo ouvinte juntamente com dados perceptivos do cogumelo se estiver imediatamente a frente, devendo decidir se irá mover-se para frente ou ficar parado) e, ao final deste período, caso o agente avaliado tenha alcançado o cogumelo (avançado até a posição do cogumelo no corredor), é considerado que o agente o consumiu.

As criaturas são controladas por uma rede neural do tipo *feedforward* de três camadas, conforme mostra a figura 4.2. Essa rede neural recebe na camada de entrada informações visuais (perceptivas do cogumelo) e comunicativas (sinais emitidos por outra criatura), e tem saída motora (que indica se o agente irá se mover ou não) e comunicativa (que é copiada para a entrada comunicativa de outra criatura). Nessa configuração, a saída comunicativa do falante é copiada para a entrada comunicativa do ouvinte. Os pesos das conexões variam na faixa de $[-4, +4]$. A saída motora é formada por um valor binário -1 ou +1, após uma função de ativação degrau. Assim, se a ativação for acima

de determinado valor, o agente se desloca (à frente), caso contrário fica parado. Não encontramos nos trabalhos de Mirolli e Parisi (2008, 2005) o valor de limiar desse degrau e, após testar alguns valores (inclusive tratá-lo como mais um peso da rede, cujo valor pode ser ajustado), atribuímos o valor 0 para esse limiar, pois foi o valor testado cujos resultados mais se aproximaram dos resultados obtidos por Mirolli e Parisi (2008). A saída comunicativa, por sua vez, é um vetor bidimensional (a saída de dois neurônios) com valores reais, que variam entre $[-1, +1]$. Com exceção do neurônio da saída motora, nada é discutido sobre as funções de ativação dos demais neurônios nos trabalhos de Mirolli e Parisi (2008, 2005). Por serem amplamente utilizadas em redes neurais artificiais, atribuímos uma função de ativação de tipo sigmoide para gerar as ativações dos neurônios.

Com essa configuração básica, existe uma característica implícita neste experimento que achamos importante explicitar. O fato é que os agentes (tanto no papel de falante quanto no papel de ouvinte) não podem tomar certas atitudes, que poderiam influenciar bastante o experimento. Para o ouvinte, não é possível ignorar o sinal recebido. Já para o falante, sempre há um valor em sua saída comunicativa, que é copiado para entrada comunicativa do ouvinte. Não há a opção de “ficar calado” para o agente nesse papel.

De acordo com essa configuração, podem acontecer essencialmente três resultados para cada rodada:

- O agente ouvinte, após receber o sinal comunicativo, pode mover-se para frente até a proximidade do cogumelo e, nessa posição obter também seus dados perceptivos (aqui chamados de visuais), e, a partir do acréscimo desses dados, decidir ficar parado e assim não consumi-lo;
- O agente ouvinte, após receber o sinal comunicativo, pode se deslocar até a proximidade do cogumelo e, dessa posição obter também seus dados perceptivos, e, a partir do acréscimo desses dados, mover-se também para frente novamente e assim consumi-lo;
- O agente ouvinte, após receber o sinal comunicativo, pode decidir ficar parado desde o início do experimento, e nem chegar a obter dados perceptivos do cogumelo presente no ambiente.

Com base nas informações sobre o experimento que temos até esse ponto, podemos perceber que a existência de 12 posições no corredor do experimento pode ser questionável. Porque 12 e não quatro? Se após ver um cogumelo um agente emite determinada saída comunicativa, esta não mudará. Da mesma forma, se ao ouvir um sinal comunicativo, um agente toma determinada ação, esta não mudará nas mesmas circunstâncias. Mirolli e Parisi (2008) não discutem essa questão, mas acreditamos que ela possa ter relação com a função *fitness* que avalia o sucesso das criaturas, como veremos a seguir.

O sucesso adaptativo das criaturas (seu *fitness*) é avaliado positivamente pela quantidade de cogumelos comestíveis que a criatura consumiu, e negativamente pela quantidade de cogumelos venenosos consumidos e também pela quantidade de posições que a criatura percorreu, de acordo com a fórmula:

$$f(x) = \frac{an_c - bn_v - cn_m}{norm}$$

onde n_c é número de cogumelos comestíveis consumidos pela criatura x , n_v é o número de cogumelos venenosos consumidos pela criatura x , n_m é o número de movimentos dado pela criatura x . a , b e c são constantes com valores 30, 5 e 1, respectivamente. $norm$ é o fator de normalização, equivalente a 4200.

Não há no trabalho de Mirolli e Parisi (2008) uma discussão sobre os valores de a , b e c . Em nossa análise, verificamos que esses valores parecem favorecer a movimentação dos agentes, principalmente nas gerações iniciais. O benefício de consumir um cogumelo comestível é relativamente maior do que a perda sofrida com o consumo de um cogumelo venenoso. Se um agente ficar parado sempre, seu *fitness* será zero. Por outro lado, se um agente se movimenta sempre, e consome todos os cogumelos de forma indistinta, seu *fitness* final será de 0.25 $(30 \cdot 210 - 5 \cdot 210 - 1 \cdot 4200)/4200$. Ou seja, quando os agentes ainda não categorizaram os cogumelos, e não há um sistema de comunicação confiável estabelecido, os valores de a , b e c favorecem mais os agentes que tendem a se mover sempre, do que os agentes que tendem a ficar mais parados. Sem um sistema de comunicação, o comportamento ótimo para os agentes seria sempre se deslocar (mesmo gastando energia para essa ação) até a vizinhança do cogumelo, obter seus dados perceptivos, e consumi-lo ou não de acordo com sua natureza. Nesse caso, o *fitness* máximo do agente seria de 0.55 $((30 \cdot 210 - 5 \cdot 0 - 1 \cdot 3990)/4200)$. Utilizando comunicação confiável, o comportamento ótimo para um agente (que o levaria a ter o maior *fitness* possível) seria se deslocar e consumir o cogumelo, quando o cogumelo presente no ambiente for comestível, e não se deslocar quando o cogumelo presente no ambiente for venenoso, pois o deslocamento consome energia do agente. Nesse caso, o valor de *fitness* será 1.

As criaturas são evoluídas durante 2000 gerações e para fins de evolução, cada criatura é representada por um cromossomo de um vetor com 32 valores correspondentes aos pesos das conexões da rede neural que a controla. Após cada geração, a geração seguinte é criada selecionando-se os pais geradores dos novos indivíduos através do método da roleta com janelamento. No método da roleta, a chance do indivíduo ser selecionado é proporcional ao seu *fitness*. O janelamento é um complemento que permite ajustar os valores de *fitness* entre 0 e um determinado valor máximo, possibilitando a realização de uma normalização desses valores. São selecionados então dois indivíduos para realizar a reprodução utilizando *crossover*, com duplo ponto de corte e depois mutação com probabilidade de 0.1% para cada valor do cromossomo do novo indivíduo. A figura 4.3 ilustra a forma de reprodução. Se determinado valor for selecionado para sofrer mutação, é escolhido um valor aleatório, dentro do limite $[-4, +4]$ para substituir seu valor corrente. Esse procedimento se repete até que novos 100 indivíduos sejam gerados. O método de seleção não é determinístico, mas há maior probabilidade dos indivíduos com maior *fitness* serem selecionados para reprodução. Isto tende a tornar mais provável que determinadas características presentes em agentes com maior *fitness* se espalhem para a população.

4.2.3 Análise dos resultados encontrados por Mirolli e Parisi

Mirolli e Parisi (2008) coletaram vários dados para análise durante as 50 simulações realizadas. Os principais dizem respeito ao *fitness* e ao que foi chamado pelos autores de qualidade do sistema de comunicação, ou simplesmente qualidade da comunicação.

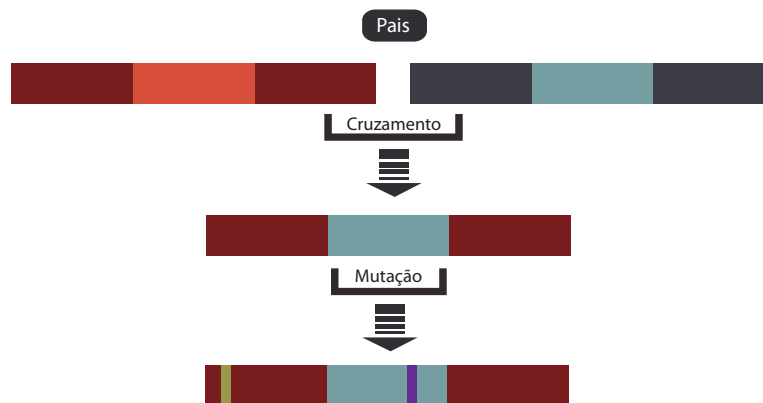


Figura 4.3 A geração dos novos indivíduos de cada geração é realizado em duas fases. Primeiro dois indivíduos são escolhidos para reprodução, e tem seus cromossomos cruzados, com duplo ponto de corte. Depois é aplicado o mecanismo de mutação, que pode alterar para um valor aleatório qualquer gene, como probabilidade de 0.1%.

Quanto a qualidade do *fitness*, já apresentamos como o cálculo é feito. Em relação ao sistema de comunicação, três critérios foram definidos para considerar-se sua existência:

1. Objetos da mesma categoria devem ser sinalizados da mesma forma.
2. Objetos de categorias diferentes devem ser sinalizados de maneiras diferentes.
3. Todos os indivíduos devem sinalizar da mesma forma (o sistema de comunicação é compartilhado pela população).

No caso do experimento considerado, os objetos são os cogumelos, e os sinais corresponderiam ao vetor bi-dimensional (com valores entre $[-1, +1]$) formado pelas saídas comunicativas dos agentes, que podem ser vistos como pontos em um sistema de coordenadas, no qual as coordenadas são os sinais do vetor bi-dimensional). Os autores consideraram então que o sistema de comunicação pode ser descrito por duas nuvens de pontos, uma formada pelos sinais emitidos por todos os indivíduos falantes quando cogumelos comestíveis estavam no ambiente, e outra formada pelos sinais emitidos por todos os indivíduos falantes quando cogumelos venenosos estavam no ambiente. Qualitativamente, a qualidade da comunicação melhora na medida em que essas nuvens se distanciam (sinais diferentes para objetos diferentes), e ficam mais compactas (com seus pontos mais próximos entre si, o que mostra que a população está sinalizando mesmos objetos de maneira semelhante). Quantitativamente, deve-se considerar dois cálculos: (a) a distância média entre cada ponto de uma nuvem para seu respectivo centro geométrico e (b) a distância entre os centros das duas nuvens. A qualidade do sistema de comunicação é calculada a partir da média aritmética de dois valores, uma vez que quanto menor o valor obtido por (a), e quanto maior o valor obtido por (b) melhor a qualidade da comunicação. Considerando que todos os valores estejam normalizados (entre 0 e 1),

o valor utilizado no cálculo final deve ser o complemento do valor obtido por (a), ou seja, se o resultado de (a) foi x , o valor considerado deve ser $(1-x)$.

No cenário descrito, a boa comunicação beneficia apenas as criaturas no papel de ouvinte, que poderá gastar energia com deslocamento apenas se houver um cogumelo comestível no ambiente. O bom falante, pelo contrário, seria prejudicado, uma vez que ajudaria concorrentes diretos pela seleção a melhorarem seus respectivos *fitness*, o que indiretamente reduziriam suas próprias chances de reprodução. Para o bom falante, esse seria uma característica puramente altruísta.

Mirolli e Parisi (2008) apresentam os resultados, fazem uma análise bastante detalhada, e buscam dados que possam explicar as características apresentadas pelas simulações. A seguir mostraremos os resultados principais encontrados por Mirolli e Parisi (2008), e algumas das explicações apresentadas por eles. Com isso pretendemos, além de apresentar a descrição completa do experimento, dar suporte para que possamos comparar os resultados da simulação original com a nossa replicação (mostrando que possuem o mesmo comportamento geral), poder esclarecer algumas diferenças em relação à nomenclatura e conceitos utilizados no experimento original em relação aos utilizados no escopo desse trabalho, e assim poder tornar mais clara nossa proposta metodológica.

A figura 4.4 mostra dez resultados que os autores consideraram representativos de suas simulações. Os dados da simulação mostram instabilidade tanto no *fitness* (linha mais clara) quanto na qualidade da comunicação (linha mais escura) média da população, ao longo das gerações. Os valores dessas grandezas variam de forma cíclica entre valores altos e baixos. Isso significa que periodicamente, um sistema de comunicação emerge, e depois é destruído. Esses ciclos, apesar de assumir diferentes formas, invariavelmente acontecem em todas as simulações.

A explicação apresentada para apontar o mecanismo causador da “destruição” do sistema de comunicação é a seguinte: após o sistema de comunicação emergir e se estabelecer, surgem indivíduos considerados enganadores (geralmente pela ocorrência de alguma mutação). Os enganadores tiram vantagem da situação, pois podem se beneficiar da informação fornecida pelos bons falantes, e por sua vez, enganam seus ouvintes, emitindo sinais opostos ao que emitiriam se estivessem seguindo o sistema de comunicação então estabelecido, induzindo os bons ouvintes a cometerem enganos. Com esse comportamento eles conseguem diminuir o *fitness* dos demais indivíduos do restante da população, e assim essa característica é espalhada nas próximas gerações. Note que o que esses indivíduos (enganadores) provocam não é o aumento absoluto de seus *fitness*, mas o aumento relativo pela redução do *fitness* médio da população, pois quando atuam como falantes é bastante provável que enganem os ouvintes. Os ouvintes enganados tomarão a ação oposta àquela que elevaria seu *fitness*: se houver um cogumelo comestível no ambiente, ele será induzido a permanecer parado; se houver um cogumelo venenoso, ele será induzido a se deslocar até ele, podendo consumi-lo ou não, mas já tendo gasto energia no deslocamento. Dessa forma, os falantes enganadores elevam a probabilidade de serem selecionados para reprodução e espalharem seus genes, que codificam essa característica, para a população da próxima geração. Com essa dinâmica, em poucas gerações, tanto o sistema de comunicação quanto o *fitness* médio da população podem cair bastante. E é o que acontece várias vezes durante as simulações, como mostra a figura 4.4.

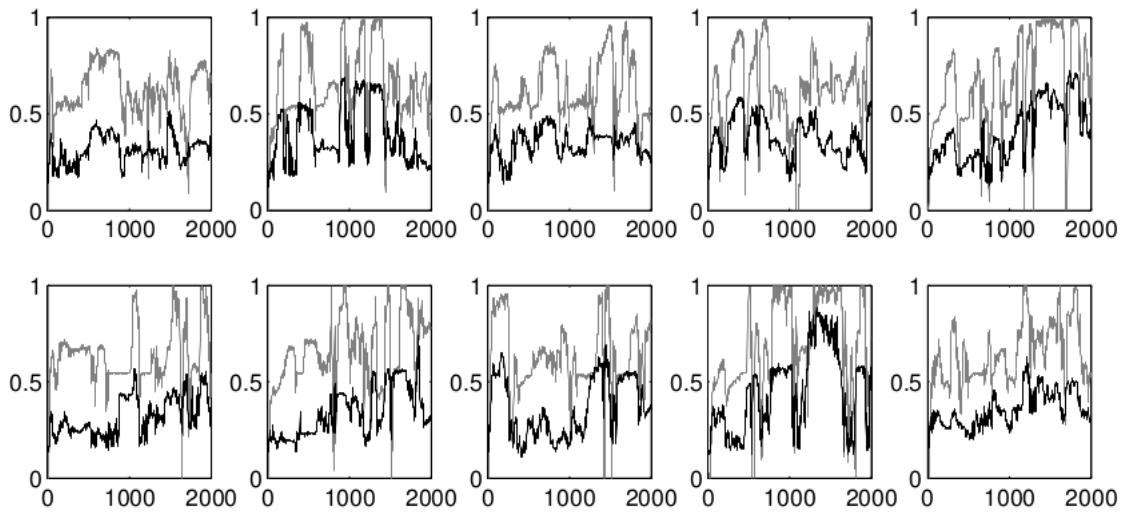


Figura 4.4 Resultados alcançados por Mirolli e Parisi em (MIROLI; PARISI, 2008). As linhas cinza são os resultados do *fitness* médio da população, e as linhas pretas são os resultados de qualidade de comunicação média da população.

Já a hipótese para o aparecimento e constante reaparecimento de um sistema de comunicação confiável considera dois fatores: a pressão cognitiva por categorização interna dos cogumelos, e a convergência genética da população. A pressão cognitiva por categorização foi um fator já levantado por Cangelosi e Parisi (1998), em um trabalho que teve influência direta sobre esse trabalho de Mirolli e Parisi (2008). Segundo essa perspectiva, a fim de ter maior assertividade em relação ao consumo dos cogumelos, é necessário que os agentes categorizem os cogumelos avistados, de acordo com suas qualidades. A partir dessa categorização (se for correta), o agente consumirá apenas cogumelos comestíveis, e evitará os venenosos. Para produzir esse comportamento, é necessário que essa informação seja apropriadamente categorizada nas unidades de sua camada interna. Ou seja, os pesos de suas conexões devem ser ajustados de tal forma que as ativações das unidades forneçam padrões similares para a percepção de cogumelos de mesma natureza, e diferentes para cogumelos de naturezas diferentes. E isso tende a acontecer durante as primeiras gerações. Porém, devido à arquitetura da rede neural, a pressão pela categorização e consumo assertivo dos cogumelos acaba tendo como efeito secundário também uma pressão para a produção de boa sinalização comunicativa. Isso acontece por que, assim como a saída da ação motora, a saída comunicativa também depende diretamente dos padrões de ativação gerados pela camada intermediária da rede neural, e se esse padrão varia conforme a natureza do cogumelo presente no ambiente, não só sua ação motora, mas também sua sinalização, vão seguir esse padrão.

Porém, esse fator sozinho não é capaz de explicar como o sistema de comunicação se espalha e se torna comum a toda população. Por que os indivíduos da população tendem a emitir os mesmos sinais para as mesmas situações (fator necessário para a boa comu-

nicação, como ela foi definida)? A explicação seria a convergência genética da população. Pelo processo de seleção, apenas alguns indivíduos são selecionados para reprodução, diminuindo a variedade genética da população. Dessa maneira, os genes (que codificam os pesos entre as conexões), em especial aqueles que conectam a camada intermediária com as saídas comunicativas, tendem a ter valores similares entre os indivíduos da população, com o passar das gerações.

Após mostrar os resultados e levantarem estas hipóteses, os autores fazem uma análise bastante detalhada de alguns pontos do experimento. Eles recortam uma simulação específica e acompanham algumas gerações, em um período em que há um ciclo de surgimento e depreciação de um sistema de comunicação, buscando avaliar se a dinâmica da comunicação se deve realmente à hipótese levantada. São então calculados medidas relacionadas à alguns fatores que podem ajudar a explicar a não estabilização do sistema de comunicação e do *fitness* da população. As medidas foram de diversas naturezas: convergência (e divergência) genética da população, sinais emitidos pelos agentes, acompanhamento das mudanças genéticas no melhor indivíduo e na média populacional, e medidas do que Mirolli e Parisi (2008) chamaram de qualidade do sistema de comunicação e qualidade do sistema de representação.

As medidas relacionadas à convergência genética utilizam os valores dos pesos de todas as conexões em todos os indivíduos da população para chegar aos resultados. Quanto aos sinais emitidos pelos agentes, é calculada a média de cada um dos valores (que variam entre $[-1, 1]$) do vetor bidimensional que compõem os sinais emitidos pela população. Calculando essa média para cada um dos dois valores do vetor da saída comunicativa, e para a sinalização utilizada para cada tipo de cogumelo (comestíveis e venenosos), teremos quatro valores diferentes. Para calcular o que foi chamado de qualidade de comunicação e qualidade de representação, no entanto, é preciso discutir um pouco mais, e avaliar a adequação dos termos utilizados, justamente para não cairmos nos problemas que discutimos no capítulo 2 (como falta de padronização e indefinição do significado dos termos).

Mirolli e Parisi (2008) propõem calcular a qualidade do sistema de comunicação e do sistema de representação, tanto a nível populacional como a nível individual. A qualidade do sistema de comunicação para a população é calculado como já descrevemos acima, com base na distância entre os centroides e na proximidade entre os pontos de cada nuvem, uma formada pelos pontos dos sinais emitidos na presença dos cogumelos comestíveis e outra formada pelos pontos emitidos na presença dos cogumelos venenosos, com todas as criaturas da população emitindo seus sinais uma vez para cada tipo de cogumelo. O que foi chamado de qualidade do sistema de comunicação individual é a média do valor, para cada indivíduo, da qualidade do sistema de comunicação. Cada cálculo considera os sinais emitidos por cada indivíduo falante para todos os 420 cogumelos para calcular os centroides e os demais valores necessários para se obter a qualidade do sistema de comunicação (como descrito acima). Após realizar esse cálculo para cada indivíduo, é obtida a média aritmética da qualidade do sistema de comunicação de cada indivíduo, que dará o valor dessa medida. O próprio trabalho de Mirolli e Parisi destaca que para haver um bom sistema de comunicação, é preciso haver bons sinalizadores, e bons ouvintes, como já colocamos aqui, porém, podemos perceber que o que é chamado de cálculo da

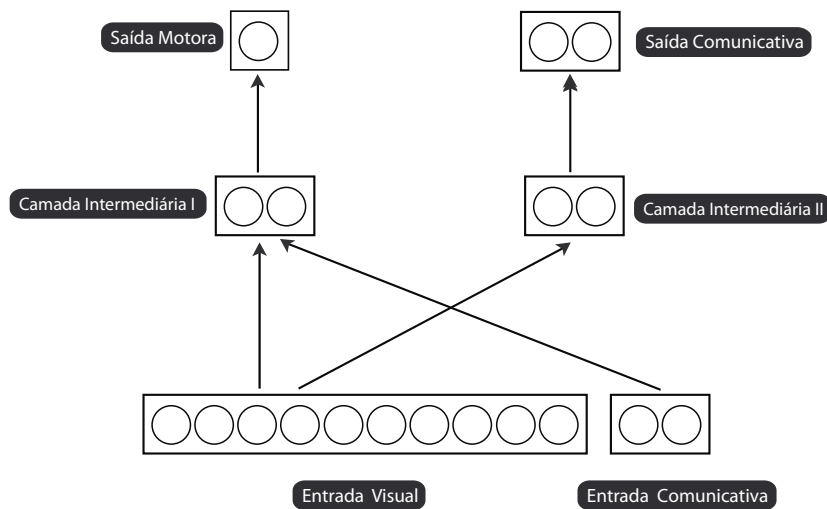


Figura 4.5 Arquitetura que evita o efeito da categorização visual na comunicação em (MIROLLI; PARISI, 2008).

qualidade do sistema de comunicação leva em consideração apenas a parte da emissão dos sinais comunicativos, sem considerar o comportamento dos ouvintes. Talvez, um nome mais apropriado para essa informação seria algo como “qualidade dos sinais emitidos”, uma vez que só considera esse aspecto da comunicação.

Cálculo semelhante foi realizado para a qualidade da representação populacional, e qualidade da representação individual. A única diferença é que no lugar de utilizar a saída comunicativa dos indivíduos falantes, são utilizados os valores das ativações da camada intermediária dos falantes (que também é um vetor bi-dimensional). Não é deixado explícito o que (nem se) é considerado como algum tipo de representação, nem o que estaria representado para quem, ou qual efeito essa representação causaria em seu intérprete. E mais uma vez, leva em consideração apenas dados relativos ao falante, pois os falantes estão em situação que podem perceber as propriedades do cogumelo presente no ambiente, coisa que o ouvinte inicialmente não pode. Porém, nos parece razoável supor que, sendo as criaturas no papel de ouvinte que devem tomar alguma decisão baseadas no sinal recebido, são elas que precisam interpretar os sinais recebidos como representação de algo. Essas medidas realizadas por Mirolli e Parisi (2008), portanto, não parecem revelar nada sobre representação, pelo menos não na maneira como definimos no capítulo 2, mas parecem revelar algo mais próximo à qualidade da categorização dos cogumelos.

Essas avaliações sobre uma simulação específica, segundo os autores, mostraram que as hipóteses levantadas podem explicar os resultados. Mas, para confirmação das hipóteses foram realizadas outras duas simulações controladas: uma sem presença de comunicação, e outra sem pressão cognitiva. No experimento sem comunicação, é utilizada a mesma

rede da simulação base, mas a saída comunicativa não pode ser copiada para o ouvinte. Ela mantém o estado interno, e a qualquer momento pode-se verificar quais sinais estão sendo gerados. Nesse caso, ao habilitar a comunicação para fins de teste, percebe-se que a qualidade do sistema de comunicação é boa (confiável), mostrando que a categorização interna continua ocorrendo, e que é a responsável pela geração de bons emissores, além de mostrar que há uma certa convergência genética na população. No experimento sem pressão cognitiva, a diferença em relação a simulação feita no experimento base é que na nova arquitetura da rede neural existem dois grupos independentes de neurônios intermediários ao invés de apenas um, como na simulação base. A figura 4.5 mostra a arquitetura da rede neural com essa modificação. Um dos grupos é ligado apenas à saída motora, enquanto o outro grupo era ligado apenas à saída de comunicação. Dessa forma separa-se o efeito da categorização visual na comunicação. Nesse caso, não há emergência de boa comunicação, mostrando que ao separar a categorização da geração dos sinais de comunicação, não é possível o surgimento de bons falantes. Esses resultados complementam os obtidos no experimento original, e parecem confirmar as hipóteses levantadas inicialmente pelos autores para explicar os resultados obtidos.

4.3 EXPERIMENTO ANALISADO SOB A PERSPECTIVA DA SEMIÓTICA DE PEIRCE

Apesar de o artigo fazer análises bastante detalhadas sobre a evolução da população, descrevendo o processo evolutivo e analisando dados precisos sobre o *fitness* e o sistema comunicativo durante o experimento, em nenhum momento foi discutido se, o quê e como os sinais comunicados representariam algo, nem para quem representariam. Mesmo quando é descrita a “qualidade da representação”, nomenclatura que pode fazer o leitor imaginar que se trate do uso de representações na perspectiva da semiótica, esse tema não é discutido.

Como já discutimos, aqui usaremos a teoria semiótica definida por Peirce para tentar descobrir se existem processos representativos durante a comunicação, e se existirem, de que tipo são. Para realizar essa busca, primeiro devemos fazer algumas definições de cenário. Queremos verificar se existe (ou se surge durante o experimento) algum processo de representação durante a comunicação. Nesse caso devemos definir quais são os elementos da tríade semiótica formada. Nossos intérpretes são as criaturas que formam a população da simulação, quando estão no papel de ouvintes, e o efeito interpretante pode ser dois: se mover ou ficar parado. Durante a comunicação, um sinal é enviado do falante para o ouvinte; esse sinal poderá ser o signo (apenas no momento em que for interpretado como tal). Esse signo, por sua vez, é algo emitido pelo falante após ver um cogumelo, e a ação do ouvinte deve ser relacionada ao cogumelo presente no ambiente, então seu objeto (do sinal emitido) deve ser o cogumelo. Dessa forma, iremos procurar um processo representativo envolvendo esses três elementos que devem estar em ação simultaneamente. Note que em relação a Mirolli e Parisi há uma mudança de perspectiva de análise quanto ao estabelecimento de um sistema de comunicação, pois não apenas incluímos, mas atribuímos um papel fundamental à criatura ouvinte, que de fato deverá ser o intérprete do signo, estabelecendo o processo representativo.

As classificações mais básicas de signo são ícone, índice e símbolo (para mais detalhes veja a seção 2.6 dessa dissertação). Dessa forma, devemos procurar algum processo que tenha esses três elementos, cujo signo, por mediação de um interpretante, estabeleça alguma dessas relações com seu objeto. No caso particular desse experimento, o signo comunicado não tem nenhuma similaridade com o objeto (as propriedades do objeto são formadas por dez valores binários, enquanto o signo é formado por dois valores reais), então não pode ser um ícone. Há a possibilidade desse sinal ser índice, principalmente nas gerações iniciais, quando podemos imaginar que o signo recebido "atraí a atenção" da criatura para o objeto, fazendo com que a criatura se mova até o cogumelo. Porém, até que o agente chegue na vizinhança do cogumelo, a criatura não poderá relacionar o signo ao objeto, porque o objeto não pode ser percebido pela criatura ouvinte, impedindo assim de que se estabeleça alguma relação temporal ou física, limitando as possibilidades. Por outro lado, um signo deve ser capaz de representar seu objeto em algum aspecto; no caso desse experimento, é importante que o agente consiga distinguir entre cogumelos comestíveis e venenosos, para que possa consumir o primeiro, e não se mover caso o segundo esteja no ambiente. Sendo um índice, não é possível para criatura distinguir cogumelos venenosos e comestíveis até visualizar o cogumelo, uma vez que para interpretar o índice o intérprete precisa perceber a existência do objeto. Porém, se o signo for um símbolo, que se relaciona com seu objeto através do uso de uma associação entre os dois, que foi de alguma forma adquirida pelo agente, este será capaz de distinguir o tipo de cogumelo sem necessidade de visualizar o objeto. Para seu sucesso adaptativo, a criatura deve ser capaz de distinguir o tipo de cogumelo (comestível ou venenoso) e não cada cogumelo individualmente, logo o objeto do signo é um geral. Por isso, nesse trabalho, para a busca da ocorrência de possíveis signos adotaremos a posição de que, se ocorrem, serão símbolos, de acordo com a concepção peirciana desse termo.

Definimos os elementos, mas ainda não o processo envolvido. Para que aconteça o processo representativo, o signo deve representar alguma coisa (seu objeto) em algum aspecto, para alguém (um intérprete), gerando neste um efeito (interpretante). Onde procurá-lo? Antes de chegar às saídas motora e comunicativa, a percepção do possível objeto (o cogumelo), e a recepção do signo (o sinal comunicativo) passam pela camada intermediária da rede neural, e a partir de suas ativações (juntamente com os pesos associados) é que as respostas motoras e comunicativas são determinadas. Assim, podemos supor que, se ocorre um processo representativo simbólico, podemos descobrir algo analisando essa camada da rede neural em busca de uma associação entre signo e objeto. Essa estratégia, de analisar a camada intermediária da rede neural em busca de representações já foi utilizada por Mirolli (2012), como comentamos no capítulo 3, mas com as diferenças essenciais de que no experimento analisado por Mirolli (2012) não havia nenhum tipo de sistema comunicativo (existia apenas um agente a cada simulação), e a definição de representação utilizada é diferente da que utilizamos aqui.

O signo deve ser capaz de representar seu objeto em algum aspecto e, no caso desse experimento, como comentamos anteriormente, para o sucesso dos agentes o ideal é que o signo seja capaz de representar o tipo do cogumelo presente no ambiente (comestível ou venenoso). Como podemos verificar isso? Se ao ver um cogumelo (o objeto) de determinada natureza observamos determinado padrão de ativação na camada intermediária

do agente e, se após esse agente ouvir um sinal comunicativo (o signo) emitido por outro agente, relacionado a um cogumelo de mesma natureza, observarmos o mesmo padrão de ativação, podemos afirmar que houve uma interpretação por parte do primeiro agente, fazendo com que a entrada auditiva gere o mesmo efeito pragmático que a entrada visual, para cogumelos do mesmo tipo. Ou seja o signo (sinal comunicativo emitido pelo falante e recebido pelo ouvinte) foi capaz de, através de um processo de mediação, representar o cogumelo no aspecto de sua natureza, de forma a gerar um efeito interpretante na criatura ouvinte, como se ela estivesse vendo o objeto. Se encontrarmos algo semelhante a esse cenário, podemos afirmar que encontramos um processo representativo.

4.3.1 Comparação de ativações da camada intermediária da rede neural

Para conseguir avaliar e quantificar a relação signo-objeto que ocorre na interpretação realizada pelo ouvinte, podemos observar a dinâmica de ativação da rede neural que controla esta criatura. Para isto, é preciso identificar os padrões de ativações da camada intermediária da criatura quando um cogumelo é visto, e também os padrões de ativação quando uma criatura recebe um sinal comunicativo. Assim, coletamos esses dados tanto quando o agente visualiza um cogumelo como quando ele ouve sobre o cogumelo. Uma forma de quantificar a relação entre o signo ouvido e um tipo de objeto visto, é avaliar a similaridade entre esses padrões.

Quando um sinal comunicativo é ouvido, a camada intermediária da rede neural produz um padrão correspondente a um par de valores de ativação, que pode ser visualizado como um ponto no espaço bidimensional. Da mesma forma, quando um cogumelo é visto, também é produzido um padrão correspondente a um par de valores de ativação. As figuras 4.6 e 4.7 ilustram a rede neural de um agente recebendo entrada comunicativa e visual, respectivamente, e apresentam uma representação gráfica das ativações geradas na camada intermediária a partir dessas entradas. As ativações originadas da entrada comunicativa foram coletadas após o agente receber as sinalizações comunicativas em todas as 420 rodadas de avaliação, e as ativações originadas da entrada visual foram coletadas após o agente visualizar todos os 420 cogumelos.

Ilustramos as ativações originadas por entrada comunicativa com o formato de cruz, enquanto as ativações originadas por entrada visual tem forma circular. Além disso, nos gráficos das figuras 4.6 e 4.7 estão representados os seguintes elementos:

- ativações em azul são ativações que levam a criatura a se mover quando há um cogumelo comestível no ambiente;
- ativações em verde são ativações que levam a criatura a se mover quando há um cogumelo venenoso no ambiente;
- ativações em vermelho são ativações que levam a criatura a ficar parada quando há um cogumelo venenoso no ambiente;
- ativações em laranja são ativações que levam a criatura a ficar parada quando há um cogumelo comestível no ambiente;

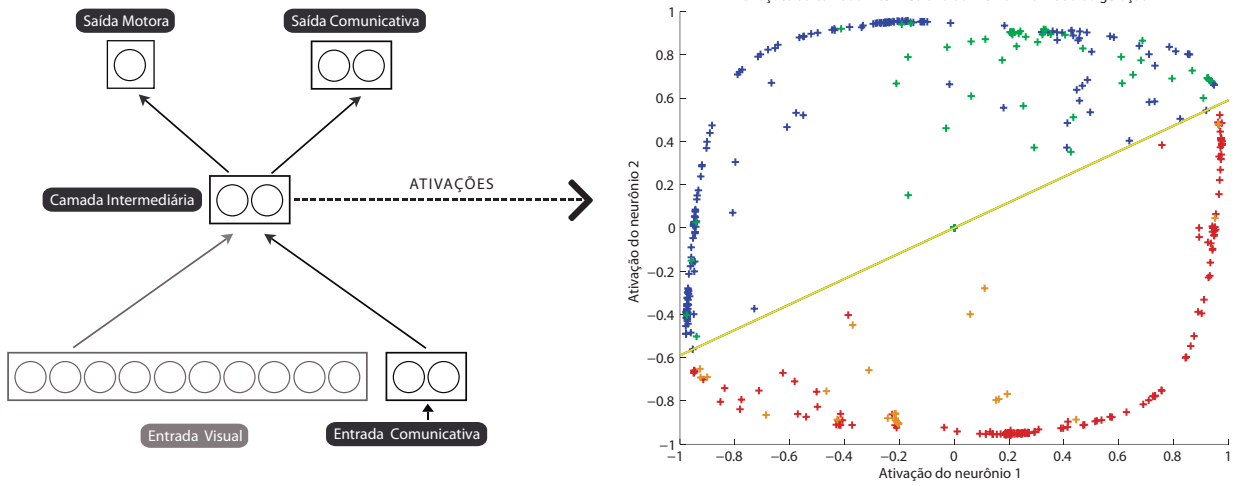


Figura 4.6 Ativações geradas a partir do recebimento de sinais comunicativos no grupo de neurônios da entrada comunicativa.

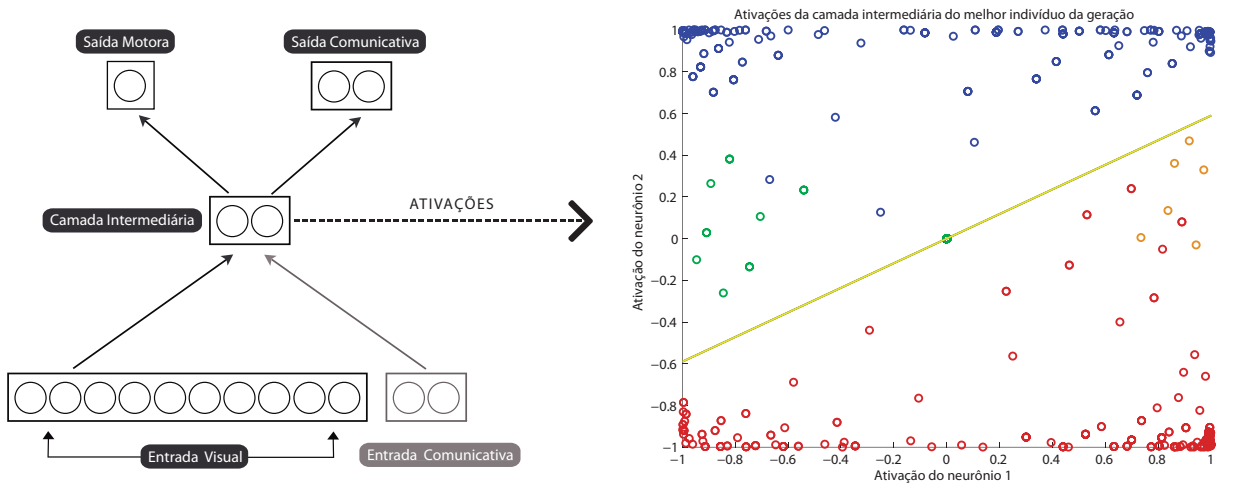


Figura 4.7 Ativações geradas no grupo de neurônios da entrada visual a partir da visualização de cogumelos.

- a linha amarela é a linha de decisão da ação motora do agente.

A linha de decisão é calculada da seguinte forma: a saída motora é uma função degrau (se sua ativação for igual ou maior que zero o agente se move, e fica parado caso contrário), e sua ativação a_{sm} é dada pelo somatório dos produtos da ativação de cada neurônio da camada intermediária (a_1 e a_2) por seu respectivo peso (w_1 e w_2), resultando na fórmula:

$$a_{sm} = w_1 \cdot a_1 + w_2 \cdot a_2$$

Se igualarmos essa equação a zero, que corresponde ao limiar que gera mudança da saída do neurônio em questão, e substituímos os valores dos pesos (que são conhecidos), podemos obter uma reta em função do valor das ativações variando de $[-1, +1]$, que são seus valores limites. Ativações sobre essa reta resultam em saída motora com valor igual a zero. Assim, para descobrir sob quais condições as ativações da camada intermediária resultam em ativação da saída motora maior que zero (que resultam em movimento da criatura), basta calcular a inequação $0 < w_1 \cdot a_1 + w_2 \cdot a_2$, realizando as substituições dos pesos.

Discutimos que, nesse experimento, se houver signo, ele é um símbolo (genuíno), e se refere a um tipo de cogumelo e não a um cogumelo específico. Por isso, a avaliação da relação dos padrões é feita comparando a ativação produzida por um signo específico com todos os possíveis cogumelos vistos de cada tipo. Ou seja, avaliamos a relação signo-objeto através de dois cálculos: um que mede a relação entre as ativações produzidas a partir do recebimento de um signo (para cada cogumelo) e as ativações originadas a partir da visualização do mesmo tipo de cogumelo presente no ambiente no momento em que o signo foi recebido, e outro que mede a relação entre as ativações produzidas a partir do recebimento de um signo (para cada cogumelo) e as ativações originadas a partir da visualização dos cogumelos do tipo diferente do presente no ambiente no momento em que o signo foi recebido.

Para avaliar a similaridade entre uma ativação originada por entrada comunicativa e ativações originadas por entrada visual, para o mesmo tipo de cogumelo, calculamos a distância entre essas ativações, como ilustra a figura 4.8. Nessa figura é colocada a ideia dos cálculos de distância entre ativação originada da entrada comunicativa (cruz), e as ativações originadas de entrada visual (círculos), para um mesmo tipo de cogumelo. No caso dessa figura, de acordo com as definições apresentadas nessa seção, o cogumelo presente no ambiente no momento em que o sinal comunicativo foi recebido é do tipo comestível, já que a cruz apresentada no gráfico é azul. Lembremos que azul e laranja são as cores que indicam que o cogumelo presente no momento do recebimento da entrada (seja comunicativa ou visual) é do tipo comestível.

Para avaliar a similaridade entre uma ativação produzida por entrada comunicativa e ativações originadas a partir entrada visual, para tipos de cogumelo diferentes, também calculamos a distância entre essas ativações, como ilustra a figura 4.9. Nela é tratada a ideia dos cálculos de distância entre as ativações originadas da entrada comunicativa, e as ativações originadas de entrada visual, para tipos de cogumelo distintos. Assim como na figura 4.8, o cogumelo presente no ambiente no momento em que o sinal comunicativo foi recebido é do tipo comestível. Logo, deve-se calcular a distância dessa ativação para

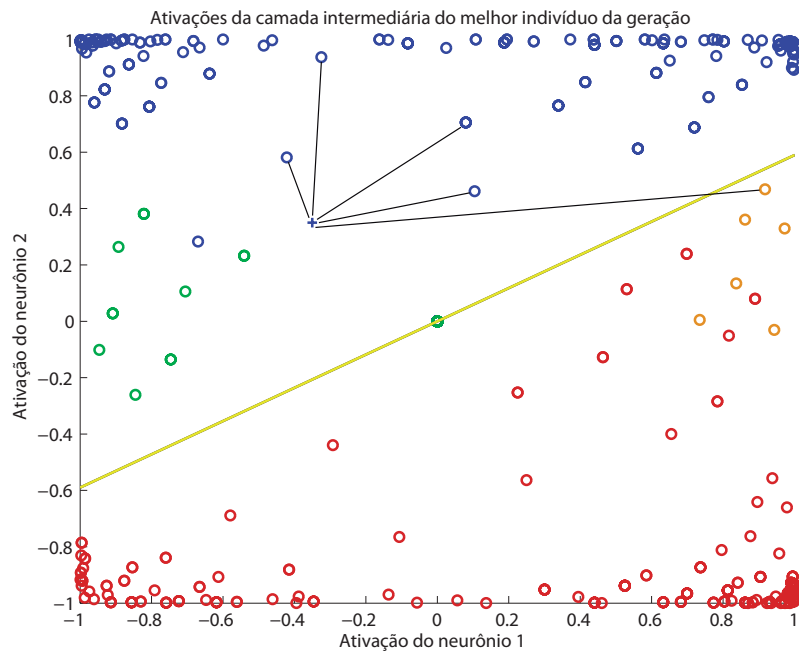


Figura 4.8 Ilustração do cálculo de similaridade entre uma ativação originada por entrada comunicativa e ativações originadas por entrada visual, para o mesmo tipo de cogumelo.

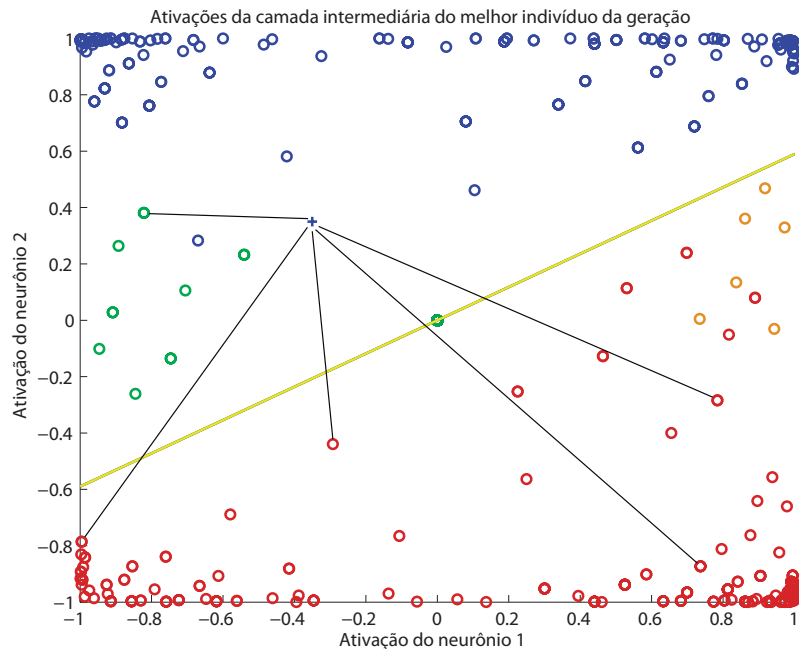


Figura 4.9 Ilustração do cálculo de similaridade entre uma ativação produzida por entrada comunicativa e ativações originadas a partir entrada visual, para tipos de cogumelo diferentes.

as ativações geradas a partir de entrada visual de cogumelos do tipo oposto, nesse caso, venenoso. Lembremos que vermelho e verde são as cores que indicam que o cogumelo presente no momento do recebimento da entrada (seja comunicativa ou visual) é do tipo venenoso.

Note que esses cálculos são realizados de acordo com o tipo de cogumelo no ambiente, e não de acordo com a ação tomada pelo agente. Por isso, as ligações entre cruzes e círculos podem ocorrer entre lados diferentes da fronteira de decisão. Além disso, se as ativações originadas de entradas diferentes estiverem próximas, há então uma alta similaridade entre essas ativações, e se estiverem distantes, há então uma baixa similaridade entre essas ativações. Ou seja, podemos considerar a similaridade como o complemento das distâncias entre as ativações, uma vez que quanto menor a distância, maior a similaridade, e quanto menor a distância, maior a similaridade.

Logo, para calcular da média da similaridade (MS) entre ativações originadas a partir de entrada comunicativa e o padrão de ativações originadas pela visualização de um tipo de cogumelo, como ilustram as figuras 4.8 e 4.9, devemos utilizar a média do complemento da distância entre as ativações, conforme a fórmula:

$$MS_{ij} = \frac{\sum_{k=1}^{210} 1 - \sqrt{(a_{ij}^{1c} - a_{ijk}^{1v})^2 + (a_{ij}^{2c} - a_{ijk}^{2v})^2}}{210 \cdot 2\sqrt{2}},$$

na qual a_{ij}^{1c} e a_{ij}^{2c} são, respectivamente, a ativação do neurônio 1 e do neurônio 2, geradas a partir de uma entrada comunicativa, na criatura i , na rodada j ; e a_{ijk}^{1v} e a_{ijk}^{2v} , são, respectivamente, as ativações do neurônio 1 e do neurônio 2, também da criatura i e rodada j , para todos os k cogumelos do mesmo tipo, ou do tipo oposto ao presente no ambiente no momento da recepção do signo. A divisão por $2\sqrt{2}$ é para normalizar o valor de MS entre 0 e 1.

A média de similaridade avalia a semelhança entre dois padrões de ativação, dessa forma os cálculos serão feitos para mensurar a similaridade entre padrões de ativação originados a partir de entradas diferentes para um mesmo tipo e para tipos diferentes de cogumelos. Para conseguir realizar os cálculos necessários e quantificar essa similaridade, registramos a ativação da camada intermediária de cada criatura para cada signo ouvido a cada rodada de avaliação, assim como o tipo (comestível/venenoso) do cogumelo presente nesta rodada. Também apresentamos cada um dos 420 cogumelos para registrar a ativação da camada intermediária quando cada cogumelo é visto.

Para mensurar a relação signo-objeto, precisamos considerar a média de similaridade MS a cada geração, para todas as criaturas, e para ambos os tipos de cogumelos (tanto presente quanto ausente). Dessa forma, definimos essa relação da seguinte maneira:

1. **Relação signo para o objeto presente (R_p):** média aritmética das similaridades entre cada ativação originada a partir do recebimento de um signo e todas as ativações originadas a partir da visualização do tipo de cogumelo presente no ambiente no momento em que o signo foi recebido, para os 100 indivíduos da população.
2. **Relação signo para o objeto ausente (R_a):** média aritmética das similaridades entre cada ativação originada a partir do recebimento de um signo e todas as

ativações originadas a partir da visualização do tipo de cogumelo que não está presente no ambiente no momento em que o signo foi recebido, para os 100 indivíduos da população.

De modo que o cálculo das medidas R_p e R_a são feitos utilizando a fórmula:

$$R_x = \frac{\sum_{i=1}^{100} \cdot \sum_{j=1}^{420} MS_{ij}}{100 \cdot 420},$$

modificando-se apenas o tipo de cogumelo para os quais será calculada a média das distâncias (MS).

A medida de relação signo-objeto presente (R_p), quantifica o quão similares são as ativações originadas a partir de entradas diferentes para um mesmo tipo de cogumelo. Como discutimos durante essa seção, essa medida é derivada de medições de distância entre esses padrões. Utilizamos a média do complemento das distâncias, pois quanto mais próximas estiverem as ativações, menor o valor das distâncias, e mais semelhantes são os padrões de ativação, melhorando assim a associação feita pela camada intermediária entre signo e objeto. Logo, quanto maior R_p , melhor a relação signo-objeto.

A medida de relação signo-objeto ausente (R_a) também se baseia nas distâncias entre ativações, mas entre tipos diferentes de objetos. Quanto menor a relação entre essas ativações, menor o risco de ocorrerem associações equivocadas por parte dos agentes, evidenciando uma melhor a relação signo-objeto. Por isso, também utilizamos o complemento das distâncias entre as ativações, uma vez que, menor relação é caracterizado por uma maior distância entre esses padrões de ativação. Logo, quanto menor R_a , melhor a relação signo-objeto.

Pretendemos avaliar se as medidas de relação signo-objeto têm alguma influência em valores como o *fitness*, uma vez que a ocorrência de uma boa relação do signo com o tipo de cogumelo presente indica que a tendência é de haver poucos equívocos das criaturas em suas ações (ou seja, conseguem determinar os tipos de cogumelos a partir de informações comunicativas da mesma forma como os classificariam se os estivessem vendo), e uma boa relação do signo com o tipo de cogumelo ausente, indica a tendência de haver mais equívocos (ou seja, não conseguem determinar os tipos de cogumelos a partir de informações comunicativas da mesma forma como os classificariam se os estivessem vendo, tomando assim ações diferentes).

4.3.2 Outras medidas

Outra maneira de avaliar a qualidade da relação signo-objeto é avaliando o efeito interpretante, ou seja, identificando as consequências pragmáticas da relação. Para isso utilizaremos os dados armazenados sobre a ação tomada pelo agente, em ambas as ocasiões, quando vê e quando ouve sobre um cogumelo. Dessa maneira podemos mensurar a quantidade média de ações iguais que as criaturas da população tomaram ao ver e ouvir sobre um mesmo objeto, ou ainda a quantidade de ações acertadas (ativações em azul e em vermelho) e equivocadas (ativações em verde e em laranja), originadas a partir de cada entrada (considerando que o certo é a decisão que vá resultar em melhor *fitness*). A

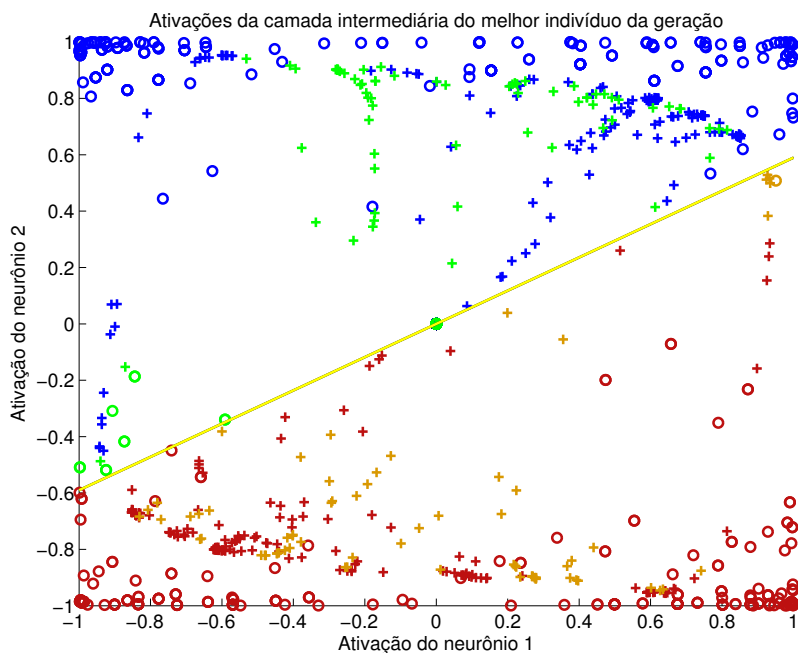


Figura 4.10 Gráfico com os pontos gerados pelas ativações da camada intermediária da rede neural da criatura com maior *fitness* em uma geração.

figura 4.10 apresenta um gráfico semelhante aos apresentados nas figuras 4.6 e 4.7, com as ativações originadas de entradas diferentes apresentadas no mesmo gráfico.

É importante notar que os sinais ouvidos por cada agente são originados de um outro agente escolhido aleatoriamente na população, e diferentes agentes no papel de ouvintes podem receber sinais diferentes para um mesmo cogumelo presente no ambiente. O que se espera de um bom sistema de comunicação, no entanto, não é que os sinais sejam exatamente iguais, mas que sejam semelhantes o bastante para que o agente ouvinte possa tomar a decisão correta (que os sinais possam representar o objeto no aspecto de sua natureza). Por isso devemos ter muito cuidado ao analisar os resultados, e não descartar nenhuma evidência de interpretação.

Alguns outros dados e detalhes dos resultados também são relevantes, os quais procuramos explorar e analisar com mais detalhes no nosso trabalho. Eles também fazem parte dos nossos objetivos, seja por que de alguma maneira (direta ou indireta) contribuem para a construção do objetivo principal (da análise de processos representativos), seja por que contribuem para também para o estabelecimento de diálogo com a comunidade, e pode fornecer mais colaborações no que diz respeito a análises relacionadas com esse tipo de experimento.

Outra nova medida que propomos para esse experimento está relacionada aos sinais emitidos pelos falantes. O trabalho original apresenta somente a média destes sinais ao longo das gerações, mas é provável que pelo menos durante alguns períodos da simulação a dispersão dos sinais emitidos pela população seja tão grande que esse dado não seja muito significativo. Como informação complementar, podemos avaliar a variância dos

sinais emitidos. De fato, pelos mesmos motivos a medida da variância vai ser útil também na análise das ativações e distâncias das medidas de relação signo-objeto.

4.4 REPLICAÇÃO E MODIFICAÇÕES PROPOSTAS NO EXPERIMENTO ORIGINAL

O primeiro passo para começarmos a procurar respostas às perguntas lançadas, é a replicação do experimento original que traz o trabalho de Mirolli e Parisi (2008). Com o experimento replicado poderemos então realizar algumas modificações que nos ajudarão a ter os dados suficientes para responder as questões levantadas. Buscamos principalmente elementos que possam evidenciar processos representativos durante a comunicação e, como já comentamos, é nas ativações da camada intermediária da rede que temos a maior probabilidade de encontrar as respostas.

Uma maneira de chegarmos a algumas respostas é através da descoberta de relação entre determinados fatores. Por exemplo, temos a hipótese de que existe alguma relação entre o *fitness* e as medidas de relação signo-objeto. Devemos então realizar a coleta de dados e os cálculos apropriados para verificar se essa hipótese se confirma, ou se deixamos de levar em consideração outros fatores presentes na suposta relação, e que fazem os resultados diferentes do esperado.

Fizemos três modificações no experimento original, que resultaram em três experimentos diferentes do original. Uma dessas variações com modificação da arquitetura da rede neural, e as outras duas com modificações na dinâmica de interações entre os agentes e na seleção para reprodução.

4.4.1 Arquitetura modificada

Aqui vamos utilizar a terminologia de arquitetura original para designar o experimento original, e arquitetura modificada para se referir ao experimento com arquitetura modificada. O que pretendemos com essa arquitetura é separar influências visuais e auditivas, de forma que tanto a saída comunicativa quanto a saída motora sejam influenciadas apenas pelos fatores que tem ligação direta com cada uma e, em algum grau separar a produção de signos do controle motor do agente. Quando o agente está no papel de falante, apenas as entradas visuais são relevantes para ele, pois tudo o que importa é que ele veja o cogumelo e tenha sua saída comunicativa copiada para a entrada comunicativa do agente ouvinte da iteração, e a entrada comunicativa desse agente não é utilizada (ficando com valor constante 0). A figura 4.11 apresenta o modelo dessa arquitetura modificada. Com ela, o processo de representação, da forma como discutimos, é impossibilitado, uma vez que as duas camadas intermediárias não têm nenhuma relação.

Para realizar essas avaliações, a arquitetura modificada, como pode ser visualizado na figura 4.11, possui dois grupos (grupo 1 e grupo 2) de dois neurônios, ou seja, a camada intermediária original foi duplicada, e temos algo como duas camadas intermediárias no mesmo nível. Ao grupo 1 se ligam apenas os neurônios que recebem as informações perceptivas do cogumelo, e ao grupo 2 se ligam apenas os neurônios que recebem a entrada comunicativa dos agentes. O grupo 1 se conecta com todos os neurônios da saída

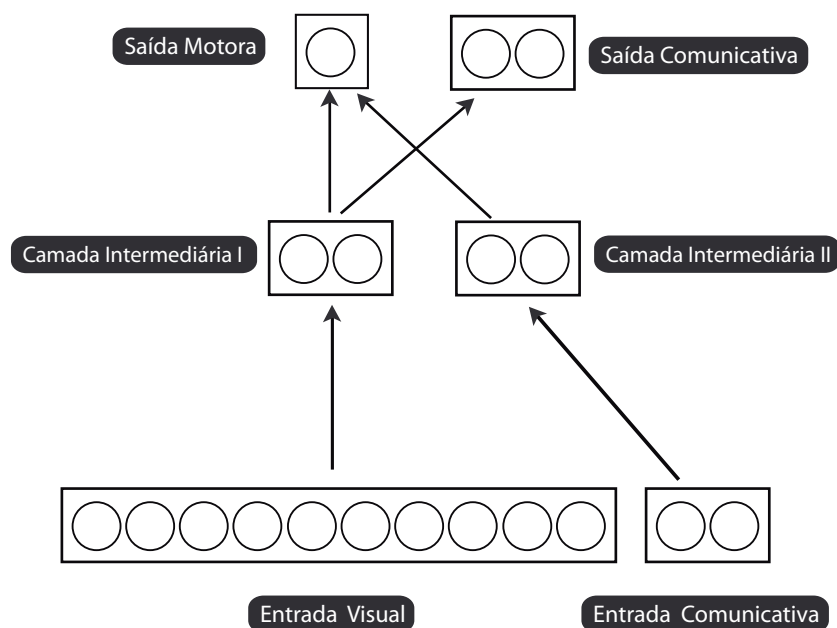


Figura 4.11 Arquitetura modificada proposta, que pretende dificultar a utilização de signos.

(tanto o responsável pela saída motora, quanto o responsável pela saída comunicativa). Já o grupo 2 se conecta apenas com o neurônio responsável pela saída motora, não influenciando a saída comunicativa. Todos os cálculos, interações e procedimentos da simulação são os mesmos, adicionando-se apenas os dois pesos a mais que a rede neural ganhou. Essa arquitetura tem alguma semelhança com a arquitetura mostrada na figura 4.5, cujo objetivo era separar a categorização da comunicação. De forma semelhante, pretendemos separar a influência da entrada visual e da entrada comunicativa, e verificar a hipótese de que essa arquitetura dificulta a utilização de signos. Apesar da semelhança entre as arquiteturas, por possuírem dois grupos de neurônios na camada intermediária, diferenças nas ligações entre as camadas, mostram uma maneira diferente de obter essa separação de influências.

Nessa arquitetura modificada (figura 4.11), o grupo 2 da camada intermediária não é conectado a saída comunicativa, mas apenas o grupo 1. Por outro lado, quando o agente é ouvinte, apenas a saída motora interessa. Para o cálculo dessa saída, a entrada comunicativa é utilizada, então o grupo 2 se conecta à saída motora. Mas quando o agente chega na vizinhança do cogumelo, ele vê suas propriedades, e então utiliza as duas informações (auditiva e visual) para tomar a decisão de consumir ou não o cogumelo. Logo, o grupo 1 da camada intermediária também irá se conectar com a saída motora. Em resumo, a entrada visual influencia tanto a saída motora quanto a saída comunicativa, e a entrada auditiva influencia apenas a saída motora, e isso utilizando camadas intermediárias separadas. Perceba que, diferente do que acontece na modificação realizada por Mirolli e Parisi (2008) (figura 4.5), não ligamos entradas diferentes a uma mesma camada intermediária, mas cada entrada possui a sua própria camada intermediária. O efeito esperado

é que os cálculos sejam realizados sobre os valores de entrada separadamente, e só então entregues para a camada de saída, prevenindo a influência de um fator sobre o outro.

4.4.2 Criaturas geneticamente semelhantes

Essa modificação teve o objetivo de verificar algumas condições sob as quais um sistema de comunicação confiável pode surgir e permanecer de forma constante durante as gerações, e como (ou se) poderemos avaliar essa situação a partir das medidas de relação signo-objeto. Caso um sistema de comunicação confiável e estável se estabeleça, e possamos analisar alguma relação das medidas de relação signo-objeto com esse comportamento, podemos verificar quais as condições mais favoráveis para o estabelecimento de processos representativos.

Para isso, a modificação realizada foi inspirada no trabalho de Floreano et al. (2007). Nessa configuração não se escolhe de forma aleatória o agente falante da rodada, mas colocamos em seu lugar uma cópia geneticamente idêntica do agente ouvinte. Floreano et al. (2007) utilizaram uma configuração parecida com essa em seu experimento, com a diferença principal de que os indivíduos geneticamente iguais estão distribuídos em grupos (FLOREANO et al., 2007). Esse tipo de configuração se baseia em alguns estudos que indicam que, para que um indivíduo membro de uma população transmita informações “honestas” para outros membros, mesmo se isso acarretar custo para ele, é preciso considerar fatores como parentesco, similaridade genética, e conflito (ou confluência) de interesses por parte dos indivíduos envolvidos nesse processo (FLOREANO et al., 2007; SMITH, 1991). Com os agentes interagindo sempre com um agente geneticamente idêntico, o grau de parentesco e similaridade genética será o maior possível, e não haverá nenhum conflito de interesse, pois o sucesso (no sentido de conseguir se reproduzir e passar seus genes adiante) do ouvinte é também o sucesso do falante. Dessa forma, pretendemos que com essa configuração a comunicação seja facilitada, o *fitness* médio das simulações seja maior, e que não ocorram as variações cíclicas observadas no experimento original.

4.4.3 Interação e seleção por grupo

Uma outra forma de favorecer a cooperação na população, mas de modo menos determinístico é a divisão da população em grupos que interagem entre si, nos quais a probabilidade de reprodução dos indivíduos dependem não apenas dos desempenhos de cada um individualmente, mas do desempenho de todos os indivíduos do grupo. Floreano et al. (2007) também utilizaram uma técnica de divisão e seleção por grupo em seu trabalho. Os agentes artificiais, com o experimento nessa configuração, apresentaram mais propensão à cooperação, segundo seus resultados. A modificação que realizamos na dinâmica do experimento foi exatamente a divisão da população em grupos, dentro dos quais os agentes interagem, e sobre os quais o mecanismo de seleção atua.

De forma semelhante à modificação descrita na seção 4.4.2, a seleção por grupo tende a favorecer o surgimento de um sistema de comunicação confiável e permanente (sem ciclos de ruptura). Mas a justificativa para esse comportamento é diferente. A interação e a seleção por grupo favorece a cooperação entre os membros de um mesmo grupo devido a

pressão seletiva. Vejamos por que existe essa pressão seletiva. A modificação realizada no experimento das simulações analisadas nessa seção divide a população de 100 agentes em 20 grupos de cinco indivíduos, de maneira aleatória. A interação entre os agentes ocorre apenas entre indivíduos de um mesmo grupo, ou seja, quando um agente é ouvinte, apenas agentes pertencentes ao mesmo grupo podem ser selecionados para se tornarem falantes. A seleção para reprodução acontece por grupos, e estabelece que quanto maior a soma dos *fitness* dos indivíduos de um grupo, maior a chance daquele grupo ser selecionado para reprodução. Para a reprodução, são sempre escolhidos dois grupos, e um indivíduo de cada grupo é escolhido para fazer o cruzamento e (após o resultado do cruzamento passar pelos operadores de mutação) gerar um novo indivíduo. Depois do processo de reprodução os indivíduos são novamente divididos de forma aleatória em grupos, e o ciclo recomeça. Dessa forma, quanto maior o *fitness* de cada indivíduo do grupo, maiores as chances de seleção do grupo e, conseqüentemente, os grupos nos quais houverem mais cooperação terão mais chances de ser selecionados para reprodução.

Após realizar essas simulações com modificações, podemos comparar os dados dessas simulações com os das outras simulações, e analisar quais as modificações nos comportamentos das medidas de relação signo-objeto. Como, por exemplo, se há, de fato melhora nessas medidas, de que forma elas variam, e como influenciam em outras medidas, como comunicação e *fitness* da população.

4.5 TESTES ESTATÍSTICOS

Devido à natureza estocástica dos experimentos, o resultado de duas simulações de um mesmo experimento nunca é exatamente o mesmo, podendo mesmo ter diferenças bastante significativas; e isso vale tanto para o experimento original quanto para os experimentos com as modificações propostas. Dessa maneira, para ter maior confiabilidade nos resultados e poder verificar de modo mais significativo as observações realizadas, realizamos várias simulações de cada experimento. Nesse contexto, em muitas situações, tanto no experimento original quanto nas modificações propostas acima, iremos realizar algum tipo de verificação entre séries numéricas, seja para medir covariância, seja para verificar se uma série é maior que a outra. Para essas verificações utilizaremos alguns testes estatísticos, que permitem por exemplo, tanto comparar várias medidas de uma mesma variável para um mesmo experimento, quanto comparar os valores de variáveis entre duas simulações, e verificar de modo confiável relações de covariância e superioridade de valores entre duas variáveis.

Dados originados por algoritmos de computação evolutiva, em geral (e é o caso na simulação utilizada nesse trabalho) possuem características como a não-normalidade e distribuição assimétrica. Nessas condições, é recomendado o uso de testes não-paramétricos (LAHOZ-BELTRA; PERALES-GRAVAN, 2010). Apesar de testes paramétricos podem utilizar dados com mais informações, para utilizá-los é necessário que os dados atendam a três restrições: sejam sobre eventos independentes, tenham distribuição normal (gaussiana), e que as amostras tenham as mesmas variâncias (homocedasticidade) (PICEK; GOLUB; JAKOBOVIC, 2012). Desse modo utilizamos testes como o teste de Mann-Whitney e o coeficiente de correlação de postos de Spearman para auxiliar na rea-

lização de algumas de nossas análises, além de calcular sempre o chamado *valor - p* dos testes.

O teste de Mann-Whitney é um teste não-paramétrico (não pressupõe normalidade ou homoscedasticidade das amostras) que verifica se há evidências para se acreditar que valores de um grupo A são superiores aos valores de um grupo B. Com ele, podemos testar, por exemplo, se a média do fitness no experimento com determinada alteração é maior ou menor que no experimento base. O cálculo de coeficiente de correlação de postos de Spearman verifica a relação entre duas variáveis, determinando quão bem a relação entre elas pode ser descrita por uma função monotônica. Quando a descrição é perfeita, o coeficiente retornado é 1 (para correlação positiva entre as variáveis), ou -1 (para uma correlação negativa). Quanto mais próximo de 0 o valor, menor a relação entre as variáveis avaliadas. Correlação positiva indica que ambas as variáveis tendem a ter variações de mesma natureza (quando uma aumenta, a outra também aumenta, e quando uma diminui a outra também diminui); a correlação negativa indica que as variáveis tem comportamentos inversos (quando uma aumenta a outra diminui e vice-versa). Com o cálculo do coeficiente de Spearman podemos verificar, por exemplo, o grau de co-variância entre variáveis, como por exemplo, variância dos sinais emitidos pela população e o fitness médio em cada geração de uma simulação.

Em ambos os casos (tanto no teste de Mann-Whitney quanto na correlação de Spearman) há o conceito de hipótese nula, que é a hipótese de que o resultado esperado pelo pesquisador é falso. Por exemplo, quando o pesquisador testa se os valores do grupo A são maiores que o do grupo B, a hipótese nula é de que não são; ou quando testa se os valores de uma variável C pode ser significativamente correlacionada com uma variável D, a hipótese nula é de que não estão. O *valor - p* indica a probabilidade de se obter uma estatística de teste igual ou mais extrema que aquela observada sob a hipótese nula. Em outras palavras, é a probabilidade do efeito observado ser devido ao acaso, e não aos fatores que estão sendo estudados (para mais detalhes sobre hipótese nula e *valor - p*, ver (PICEK; GOLUB; JAKOBOVIC, 2012)). Praticamente todos os testes estatísticos utilizam o *valor - p* para aceitar a validade de um resultado que rejeite a hipótese nula, e na literatura em geral o valor de 5% ou menos é geralmente adotado para o valor-p na grande maioria dos testes. Nesse trabalho adotamos o valor limite de 5%.

Esses cálculos estatísticos são muito importantes para que se possa mensurar o grau de relação entre dois fatores ou variáveis, para que se tenha melhor noção da quantidade de simulações que tornam os dados minimamente confiáveis, e principalmente para que não se cometa o erro de aceitar ou refutar determinada hipótese sem a devida segurança, antes descartando (ou pelo menos reduzindo muito) a possibilidade do acontecimento ter ocorrido por acaso.

RESULTADOS E ANÁLISES DOS PROCESSOS DE REPRESENTAÇÃO

No capítulo 4 descrevemos o experimento que replicamos, as informações adicionais que queríamos coletar, e as modificações que pretendíamos realizar sobre o experimento original (modificações tanto na arquitetura quanto na dinâmica da simulação). Apresentamos neste capítulo os resultados obtidos pelas simulações, com suas várias configurações diferentes. Discutiremos e analisaremos os resultados encontrados, realizando comparações, tanto entre configurações diferentes quanto dentro de uma mesma configuração. Com isso pretendemos compreender melhor a dinâmica dos experimentos, a fim de poder responder às questões levantadas nesse trabalho, e contribuir com os resultados e discussões geradas.

Iniciamos mostrando os resultados encontrados com a replicação do experimento original, e comparando os dados principais, como *fitness* e qualidade da comunicação, com os resultados obtidos por Mirolli e Parisi (2008). Após mostrar que a replicação do experimento tem resultados similares aos resultados do experimento original, iremos analisar novos dados que coletamos, abordar relacionamentos entre eles, e o que eles podem nos dizer a respeito de processos de representação. Depois, apresentaremos os resultados das simulações modificadas, também abordando relações entre os dados coletados, e realizando comparações, tanto com o experimento original quanto com os outros experimentos modificados, discutindo as relações entre vários fatores, e apontando possíveis causas de determinados comportamentos, tendo como base principal análises estatísticas sobre vários dados coletados.

5.1 RÉPLICA DO EXPERIMENTO

Replicamos o experimento original de Mirolli e Parisi (2008), com base nos dados dos trabalhos de Mirolli e Parisi (2008, 2005), conforme detalhado no capítulo 4. Os resultados obtidos por nossas simulações são compatíveis com os resultados do trabalho original. O gráfico da figura 5.1 mostra um exemplo representativo dos resultados que obtivemos.

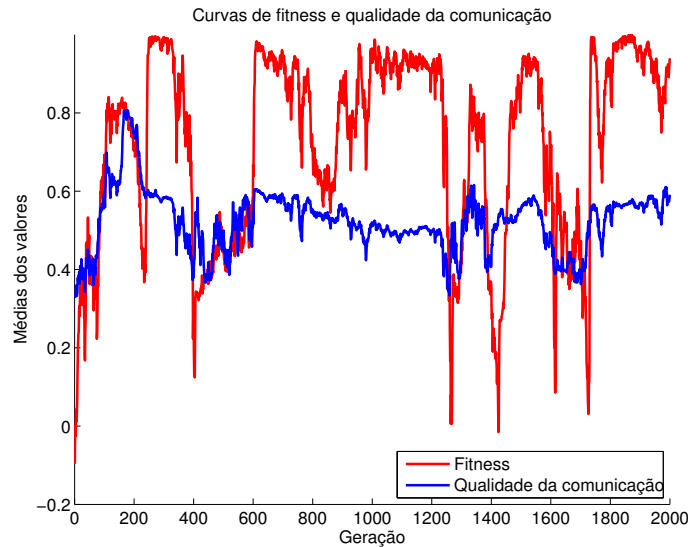


Figura 5.1 Amostra representativa do *fitness* e qualidade da comunicação das 40 simulações realizadas com a replicação do experimento.

Na seção 4.2.3, no capítulo 4, apresentamos a amostra representativa dos resultados das simulações realizadas por Mirolli e Parisi (2008), que podem ser comparadas com a figura 5.1, na qual apresentamos uma amostra representativa dos resultados encontrados nas 40 simulações da réplica do experimento original realizadas por nós. Essa comparação é feita tomando como base os resultados de *fitness* e de qualidade da comunicação (de acordo com a definição dada por Mirolli e Parisi (2008)), os quais, segundo os autores, são os dados principais do experimento. Pela própria natureza desse experimento, e pelas várias configurações que em algum momento são ajustadas de maneira aleatória (como os pesos da rede neural na população inicial, o sorteio do agente falante a cada rodada, e o valor aleatório atribuído a um peso após ocorrência de mutação), não é possível obter resultados idênticos. Podemos perceber, no entanto, que a característica principal do experimento, sua dinâmica oscilatória de surgimento e desaparecimento de um sistema de comunicação e, conseqüentemente, da não estabilização do *fitness* da população, está presente de maneira bastante similar em ambos os experimentos. Outra característica importante que também podemos perceber no experimento que replicamos, é a relação entre a qualidade da comunicação e o *fitness*, na qual o *fitness* tende a seguir a queda ou subida da qualidade da comunicação.

Outros dados explorados no trabalho de Mirolli e Parisi também foram coletados e calculados, como é o caso da variância genética da população e do deslocamento dos centroides dos pesos entre duas gerações da simulação. Como mostram as figuras 5.2 e 5.3, essas duas medidas estão divididas em grupos de pesos relacionados a funções comunicativas e não comunicativas desempenhadas pelos agentes, sendo estes os grupos: pesos da Entrada Comunicativa (EC); pesos Não-Comunicativos (NC); pesos da Saída Comunicativa (SC). Os pesos da EC correspondem aos pesos das conexões entre os neurônios

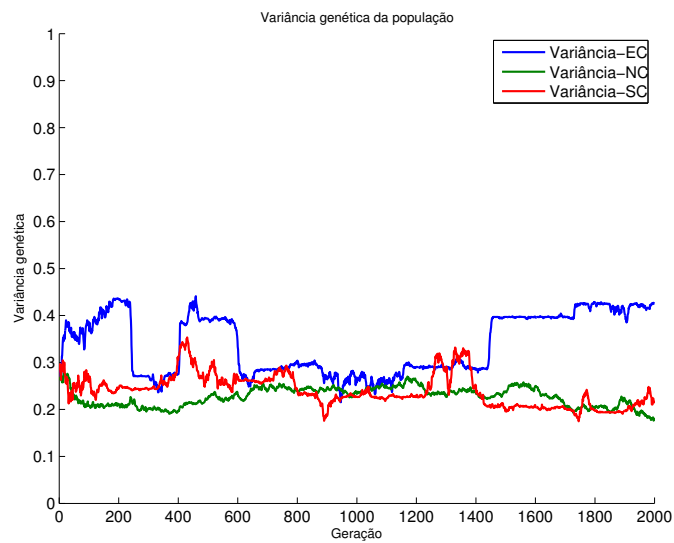


Figura 5.2 Amostra representativa da curva de variância genética da população, divididos por grupos de pesos. EC, NC, e SC se referem, respectivamente, aos grupos de pesos da Entrada Comunicativa, pesos Não-Comunicativos, e pesos da Saída Comunicativa.

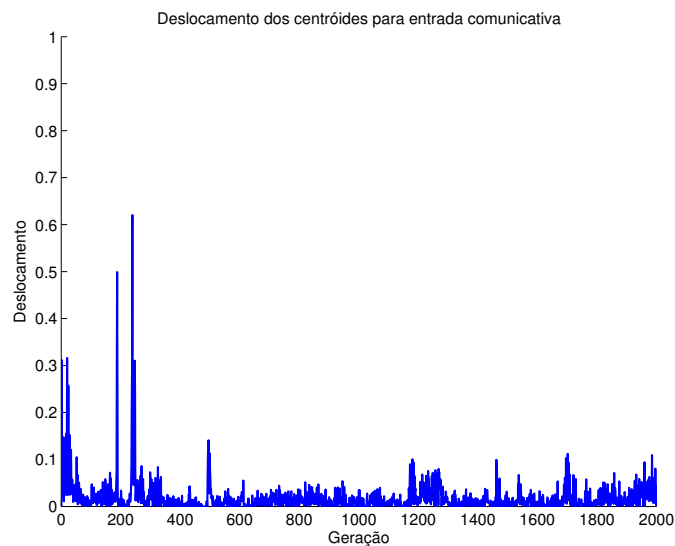


Figura 5.3 Amostra representativa da curva do deslocamento dos centróides dos pesos entre duas gerações na população.

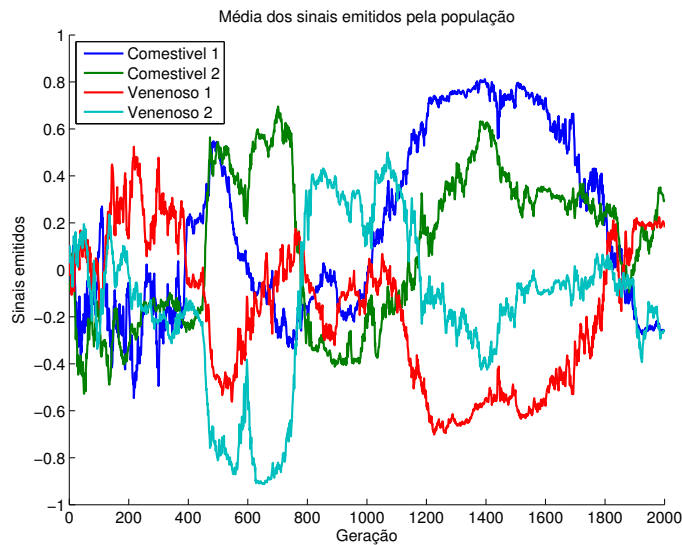


Figura 5.4 Amostra representativa dos sinais emitidos pela população.

da entrada comunicativa e os neurônios da camada intermediária (totalizando quatro pesos). Os pesos da SC correspondem aos pesos das conexões entre os neurônios da camada intermediária e os neurônios da saída comunicativa (também totalizando quatro pesos). Todos os demais pesos foram classificados como Não-Comunicativos.

Por conta da forma da curva de deslocamento dos centroides entre duas gerações, que impediria de ver de forma compreensível as três curvas na mesma figura, mas que possuem formas bastante parecidas, o gráfico da figura 5.3 apresenta apenas os dados relativos aos pesos da entrada comunicativa de uma simulação. Também calculamos a média dos sinais emitidos pelos agentes durante as gerações, como podemos ver no exemplo da figura 5.4. Todos esses gráficos apresentam resultados qualitativamente compatíveis com os resultados apresentados no trabalho de Mirolli e Parisi (2008). A partir dessas comparações entre o experimento original e a nossa replicação podemos ter mais segurança de que replicamos de maneira adequada o experimento analisado.

5.1.1 Relação signo-objeto

Discutimos no capítulo 4 a utilização da teoria semiótica de Peirce na definição e procura por representação no experimento analisado nesse trabalho. Inferimos que, se houver representação, ela é simbólica, uma vez que o signo só poderá se relacionar com seu objeto através do uso de uma associação entre os dois, que foi adquirida pelo intérprete. Para formar a tríade semiótica (signo-objeto-interpretante) do processo representativo que procuramos, chegamos às seguintes definições: o signo é o sinal emitido pelo agente na posição de falante (e recebido pelo agente na posição de ouvinte); o objeto é o cogumelo; e o efeito interpretante no agente ouvinte (que é o intérprete) poderá ser dois, se mover ou permanecer parado. Nesse caso, o efeito interpretante está diretamente relacionado à saída motora da rede neural do agente. Acontece que, antes de chegar à saída

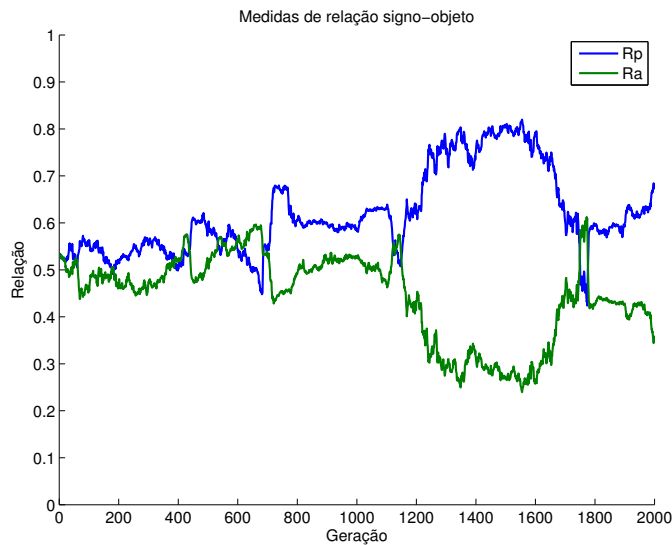


Figura 5.5 Valores médios de R_p (medida de relação signo para o objeto presente) e R_a (medida de relação signo para o objeto ausente) durante uma simulação.

motora, tanto a percepção visual do possível objeto (cogumelo), quanto o signo recebido (sinal comunicativo), passam pela camada intermediária. E, de fato, são as ativações da camada intermediária (e as ativações associadas), que determinam a saída motora e, consequentemente, o efeito interpretante.

Tendo então a camada intermediária essa característica, é possível que possamos descobrir se ocorre o processo representativo analisando dados relativos a essa camada, como já fizeram em outros trabalhos (e.g. (MIROLI, 2012)), mas com perspectivas e conceitos diferentes. Uma forma de procurar por associação entre signo e objeto é comparando os padrões de ativação da camada intermediária quando um agente vê e quando ouve sobre um mesmo tipo de cogumelo. Se um agente exibe um padrão de ativação similar na camada intermediária de sua rede neural, para os dois tipos de entrada e para o mesmo tipo de cogumelo, podemos afirmar que houve interpretação, fazendo com que a entrada auditiva gerasse o mesmo efeito pragmático que a entrada visual (para cogumelos do mesmo tipo).

No capítulo 4 discutimos ainda sobre o cálculo das medidas de relação signo-objeto, discutindo as definições e as fórmulas para o cálculo de R_p e R_a (para mais detalhes sobre essas medidas, veja seção 4.3.1). A figura 5.5 apresenta os resultados do cálculo desses valores para uma simulação, escolhida livremente dentre as 40 simulações realizadas na configuração do experimento original.

Apesar de cada simulação gerar dados diferentes, a figura 5.5 apresenta um gráfico que podemos considerar ilustrativo das características que os cálculos sobre a relação signo-objeto possui. Para que ocorra um processo de comunicação, é preciso que o signo seja capaz de indicar seu objeto para o intérprete, e que o intérprete seja capaz de relacionar o signo com seu objeto. Para o signo poder ser relacionado com um objeto, os valores de

R_p e R_a devem ter certo grau de distinção, pois de outra forma o intérprete não poderá estabelecer a relação com algum tipo de cogumelo, visto que a relação signo-objeto para ambos os tipos de cogumelo serão bastante semelhantes. R_p se refere à relação entre as ativações geradas a partir de entrada comunicativa e as ativações geradas a partir de entrada visual com o mesmo tipo de cogumelo presente no ambiente no momento em que o agente recebeu a entrada comunicativa. Logo, quanto melhor essa relação, mais próxima estará a interpretação do signo da interpretação dada ao agente à própria visualização do objeto. Por outro lado, R_a se refere à relação entre as ativações geradas a partir de entrada comunicativa e as ativações geradas a partir de entrada visual com cogumelo do tipo oposto ao presente no ambiente no momento em que o agente recebeu a entrada comunicativa. Nesse caso, quanto menor a relação, menor o risco de ocorrer uma interpretação equivocada, e o signo acabar sendo associado ao objeto diferente pelo agente. Resumidamente, quanto maior o valor de R_p e quanto menor o valor de R_a , melhores são as condições para a ocorrência de um processo de representação.

Outras duas situações podem ocorrer em relação às medidas R_p e R_a . Elas podem ter valores iguais (ou bastante similares), o que tende a criar condições adversas ao estabelecimento de qualquer processo de representação ou comunicativo, visto que os agentes estabelecerão relação do signo com um objeto de maneira praticamente aleatória. Se, por outro lado, R_p tem valor menor que R_a , há condições favoráveis para a comunicação “enganadora”. Isso por que, uma vez que as ativações da camada intermediária originadas a partir do recebimento do signo, no momento em que um tipo de cogumelos está presente no ambiente, estão mais próximas das ativações originadas a partir da visualização do outro tipo de cogumelo (o que estava ausente), os agentes são levados a associarem o signo ao objeto ausente. Nesse cenário, a relação estabelecida pelos agentes do signo com o objeto ausente é maior do que com o objeto presente. Note que ainda nesse caso pode haver processos representativos e comunicação ocorrendo normalmente, porém, ela tende a ser desfavorável ao ouvinte. Existem todos os elementos da tríade semiótica: um signo, um objeto, e um efeito interpretante no agente, pois mesmo uma interpretação equivocada, ainda é uma interpretação.

O gráfico apresentado na figura 5.6 mostra a curva da diferença entre R_p e R_a ($R_p - R_a$) para a mesma simulação apresentada na figura 5.5. Ela auxilia na rápida identificação das três situações relatadas: valores positivos relativamente altos indicam R_p maior que R_a ; valores próximos a zero indicam R_p e R_a com valores similares; e valores negativos significativamente altos indicam R_p menor que R_a . Porém, a partir dessa informação não podemos discernir se a medida variou em função da variação de R_p , de R_a , ou de ambos, o que pode ser uma informação importante em algumas situações. Sendo assim, ao longo das análises iremos utilizar a medida mais adequada para o contexto.

Analisando a figura 5.5, podemos observar que assim como a curva do *fitness* médio da população, as curvas de R_p e R_a não se estabilizam. Iniciam com valores praticamente iguais e, até aproximadamente a geração 600, há um afastamento e aproximação frequentes e de pouca amplitude entre os valores dessas medidas. Até próximo da geração 600, a diferença entre as medidas variou dentro de uma faixa próxima a 0, incluindo momentos em que R_p foi menor do que R_a (como podemos observar, por exemplo, no pico negativo próximo à geração 600, no gráfico da figura 5.6). Esse trecho tem, em sua maior parte

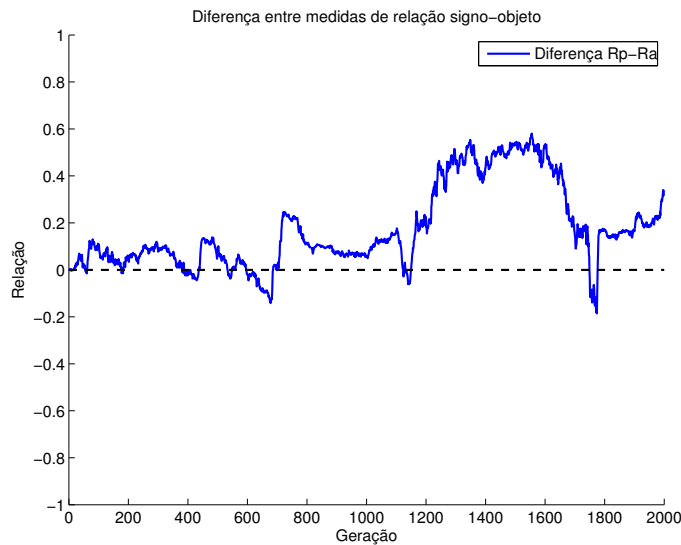


Figura 5.6 Valores médios da diferença entre R_p e R_a durante uma simulação.

características que acreditamos serem adversas ao estabelecimento de processos de representação, e nos picos negativos há características de um processo de representação com interpretação “equivocada” do agente, como descrevemos acima.

Logo depois da geração 600, até próximo da geração 1200, há um período de relativa estabilidade e afastamento dos valores de R_p e R_a , tendo D_a valores significativamente menores que R_p . Esse cenário foi descrito acima como favorável ao estabelecimento de processos representativos. Próximo à geração 1200 essa estabilidade e afastamento entre as curvas foi desfeita e, por um curto espaço de gerações, R_p teve valor mais baixo que R_a , voltando ao cenário onde ou não há processo representativo ou ele é “enganoso”.

Aproximadamente entre as gerações 1200 e 1700 há um grande distanciamento entre R_p e R_a . Esse afastamento entre as medidas ocorre tanto em função da diminuição de R_a quanto do aumento de R_p , o que configura uma condição muito boa para a ocorrência de processos representativos, e para a comunicação confiável. Mas, mais uma vez esse período se encerra praticamente com a inversão dos valores durante um curto espaço de gerações, no qual D_p tem seu valor significativamente menor do que D_a , indicando cenário ainda favorável ao aparecimento de processos de representação, porém à comunicação “enganadora”. Após essa momentânea inversão (apresentado como pico negativo na figura 5.6 próximo a geração 1800), R_p volta a ser significativamente maior que R_a , e a simulação chega ao fim.

5.1.2 Fitness e relação signo-objeto

Uma vez que as curvas que medem a qualidade da relação signo-objeto possuem oscilações que também se encaixam nos padrões de oscilação das curvas de *fitness*, levantamos a hipótese de que a medida de relação signo-objeto pode influenciar o *fitness* da população. Se as medidas de relação favorecem o surgimento de processos representativos (ou seja,

R_p e R_a tem valores distintos), a consequência esperada é que a comunicação entre os indivíduos da população seja melhorada, o que influenciaria diretamente no *fitness*. Por comunicação melhorada queremos dizer que tanto a sinalização do falante quanto a interpretação do ouvinte tendem a não ser aleatórios, mas a seguir determinado comportamento específico, a depender do tipo de cogumelo presente no ambiente. A influência sobre o *fitness*, por sua vez, pode ser positiva ou negativa, como já comentamos. Se R_p for maior que R_a , a influência tende a ser positiva, pois o intérprete tende a associar o signo com o objeto presente no ambiente, vindo a tomar a decisão de acordo com sua categorização prévia dos tipos de cogumelo. Por outro lado, se R_p for menor que R_a , a influência tende a ser negativa, visto que o intérprete tende a associar o signo com o objeto ausente do ambiente, vindo a tomar a decisão de ação contrária à que tomaria se tivesse visto o cogumelo.

Podemos perceber que essa análise é feita sob a perspectiva ouvinte-intérprete, algo que não foi abordado por Mirolli e Parisi (2008). Aqui avaliamos a existência de alguma interpretação por parte do ouvinte: se e como ele associa signo e objeto; se e como o signo recebido pode representar algo para a criatura no momento em que é interpretado. Fazemos isso essencialmente avaliando as distâncias médias entre as ativações da camada intermediária quando as criaturas ouvem sobre um cogumelo e quando veem, tanto o mesmo tipo de cogumelo sobre o qual ouviram quanto o tipo diferente. Isso nos permite observar qual a associação que a criatura está estabelecendo com o signo recebido. Se é para o tipo de cogumelo presente no ambiente ($R_p > R_a$), se é para o ausente ($R_p < R_a$), ou se é indefinido ($R_p \approx R_a$).

As medidas apresentadas nas figuras 5.5 e 5.6 podem ser utilizados para verificar a validade da hipótese de que o *fitness* é influenciado pelas medidas de relação signo-objeto. Para isso, avaliamos a correlação do *fitness* com as três medidas de relação signo-objeto: relação signo-objeto para o objeto presente (R_p), relação signo-objeto para o objeto ausente (R_a), e diferença entre as proximidades signo-objeto para o objeto presente e para o objeto ausente ($R_p - R_a$). As figuras 5.7, 5.8 e 5.9 mostram, respectivamente, os gráficos de dispersão do *fitness* e destas medidas, apresentando o valor de correlação entre essas variáveis, para uma das execuções da simulação. Os resultados desses testes estatísticos variaram em alguns valores entre as execuções, no entanto, todas as simulações mantiveram as mesmas tendências de correlações positivas e negativas, e mais de 70% obtiveram índice de correlação que podemos considerar significativos (com medida estatística ρ acima de 0.5). Os resultados que apresentamos nas figuras 5.7, 5.8 e 5.9 são exemplos que nos permitem expor os vários aspectos das simulações que observamos através desses testes.

A figura 5.7 apresenta um gráfico de dispersão dos dados de relação do signo com o objeto presente e dados de *fitness*. O valor indicado por ρ é a correlação de Spearman, uma medida estatística não paramétrica que mede a dependência entre duas variáveis, de forma a mensurar quão bem a relação entre elas pode ser descrita por uma função monotônica. O valor de ρ igual a 1 ou -1, significa que cada variável é uma função monótona perfeita da outra. Valores de ρ tendendo a zero sugerem fraca relação entre as variáveis. Podemos perceber pela figura 5.7 que há uma forte correlação entre o *fitness* e a relação signo-objeto com o objeto presente (R_p). Os valores mais altos do *fitness* ocorrem

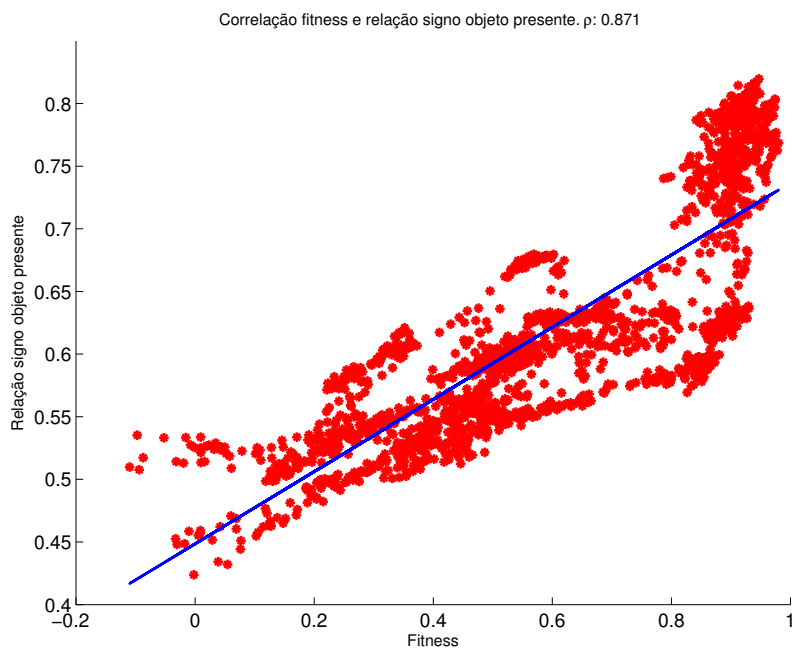


Figura 5.7 Gráfico de dispersão do *fitness* e relação do signo com o objeto presente no ambiente (R_p), em uma execução do experimento.

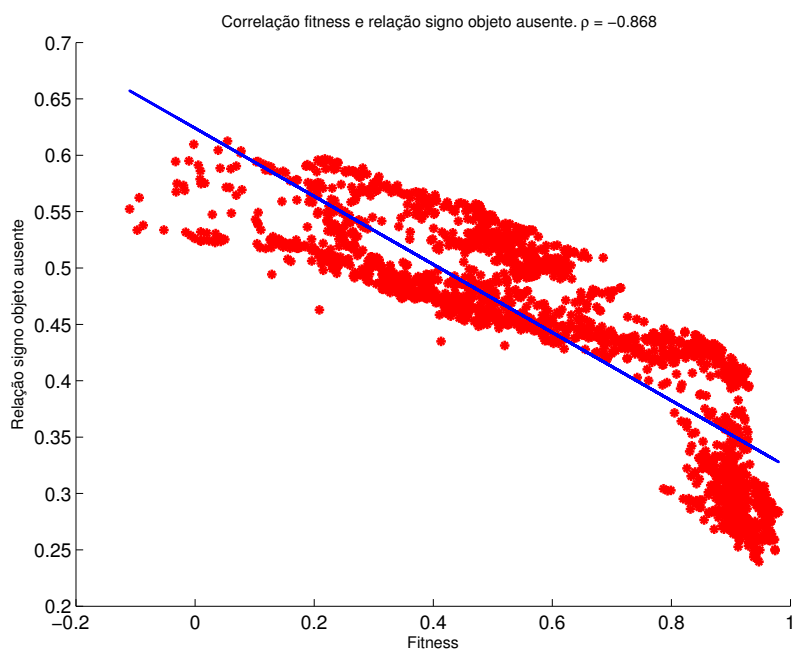


Figura 5.8 Gráfico de dispersão do *fitness* e relação do signo com o objeto ausente do ambiente (R_a), em uma execução do experimento.

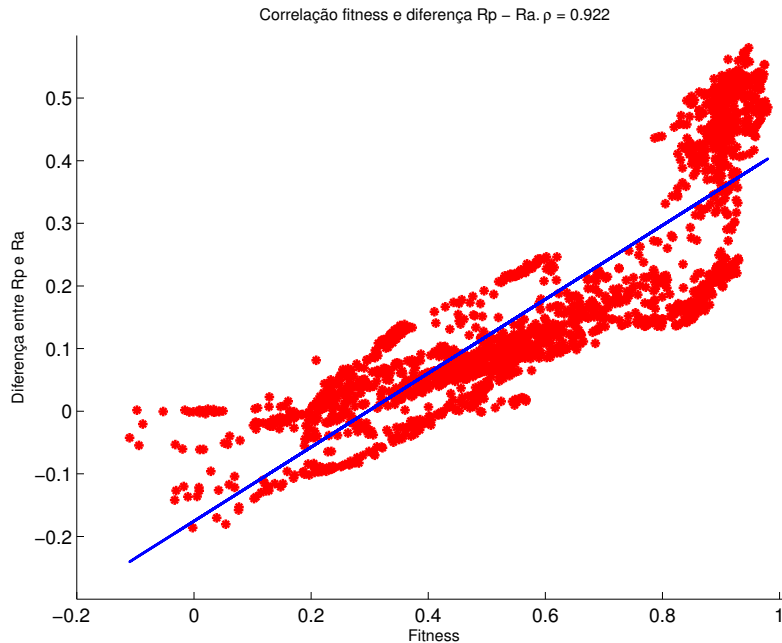


Figura 5.9 Gráfico de dispersão do *fitness* e diferença entre a medida de relação do signo com o objeto presente no ambiente e a relação do signo com o objeto ausente do ambiente ($R_p - R_a$), em uma execução do experimento.

em momentos nos quais também ocorrem os valores mais altos de R_p , e vice-versa. Ou seja, o *fitness* mostra a tendência de aumentar na medida em que a relação signo-objeto com o objeto presente aumenta, e de diminuir na medida em que essa relação se degrada. Corroborando com a percepção visual do gráfico, o valor de ρ , aproximadamente 0.87, indica uma forte correlação entre as variáveis.

A figura 5.8 apresenta um gráfico de dispersão dos dados de relação do signo com o objeto ausente e o *fitness*, de maneira similar a figura 5.7. Podemos perceber que há uma clara correlação negativa entre a medida de relação signo-objeto para o objeto ausente e o *fitness*, o que é corroborado pelo valor de ρ de aproximadamente -0.87. A correlação negativa, como discutimos na seção 4.5, indica que quando R_a aumenta, o *fitness* diminui, e quando R_a diminui, o *fitness* aumenta. Como podemos perceber no gráfico da figura 5.8, o *fitness* cresce na medida em que a relação signo-objeto com o objeto ausente diminui, e diminui quando este aumenta.

Juntando os dados apresentados nos gráficos da figura 5.7 e 5.8, chegamos aos dados apresentados na figura 5.9. Essa figura apresenta a relação entre *fitness* e a diferença entre as medidas de relação signo-objeto com o objeto presente e com o objeto ausente (valor de $R_p - R_a$). Esse valor pode ser entendido como uma ponderação entre a relação das ativações da camada intermediária, geradas a partir da auditiva e as ativações geradas a partir de entrada visual para o mesmo tipo de objeto, e a relação para objetos de outro tipo. Ou seja, leva em conta tanto a relação signo-objeto para o objeto presente, quanto para o objeto ausente. Nessa simulação, o valor de ρ para essa medida é de

aproximadamente 0.92, indicando também uma forte correlação entre ela e o *fitness*.

Nos gráficos de correlação apresentados nessa seção (figuras 5.7, 5.8 e 5.9) podemos observar que o *fitness* tende a aumentar nas seguintes condições: (1) quando a relação R_p aumenta, (2) quando a relação R_a diminui, e (3) quando a diferença entre R_p e R_a ($R_p - R_a$) aumenta; e tende a diminuir nas condições contrárias. Argumentamos no início da seção 5.1.1 e no início dessa seção, que as medições de R_p e R_a são importantes para a busca por processos representativos e, uma vez existindo processo representativo, este influenciaria na comunicação, que por sua vez influenciaria no *fitness* da população. Os resultados obtidos indicam que há de fato uma forte correlação entre as medidas de distância signo-objeto e o *fitness* médio da população, e está de acordo com a hipótese de que essas medidas são importantes para mensurarmos os processos de representação que podem ocorrer durante as simulações, e que a interpretação dos signos e a determinação da ação do intérprete, podem influenciar assim no *fitness* médio da população.

5.1.3 Medidas de correlação e variância

Todos os dados apresentados anteriormente apoiam a hipótese de que as medidas de relação signo-objeto tem forte relação com o comportamento e desempenho dos agentes. Quanto melhores essas medidas, melhores seriam representados os tipos de objeto e, conseqüentemente, mais adequada a ação resultante, impactando diretamente no *fitness*. Mas esses números ainda podem ser considerados muito gerais, podendo não serem capazes de fornecer informações mais detalhadas sobre essa relação. Por exemplo, podemos concluir que certamente há uma relação entre esses valores, porém, não podemos concluir sobre a causalidade existente, que indicaria que as medidas de relação signo-objeto influenciam diretamente o *fitness*, e são responsáveis por sua variação. Apesar de, intuitivamente, tendermos a atribuir à relação signo-objeto causalidade sobre o *fitness*, não podemos descartar a possibilidade de que um terceiro fator esteja influenciando tanto as medidas de proximidade signo-objeto quanto o *fitness*, causando assim a semelhança de variação entre essas curvas. Além disso, não podemos descartar a informação de que nem todas as simulações apresentam valores de correlação tão significativos quanto os valores apresentados nas figuras 5.7, 5.8 e 5.9.

Para investigarmos melhor essas influências e obtermos mais detalhes sobre essa relação, fizemos uma comparação mais detalhada entre *fitness* e as medidas de distância signo-objeto. Para essa comparação, escolhemos outra execução da simulação (a mesma que forneceu os dados para as figuras 5.5 e 5.6), cujos dados permitem ilustrar melhor as conclusões de acordo com as análises realizadas sobre os dados. Nessa execução, identificamos alguns intervalos de gerações nos quais os valores médios das medidas de correlação signo-objeto com o *fitness* são diferentes entre si, e então analisamos esses intervalos com mais detalhes. Além disso, como informação complementar, utilizaremos um outro dado que nos dará uma possível explicação sobre a diferença de correlação entre relação signo-objeto e *fitness* em alguns intervalos ao longo da execução: a variância dos valores da relação signo-objeto.

A variância de uma variável aleatória mede sua dispersão estatística, uma medida sobre o quão longe os valores estão da média de determinada variável. A variância de

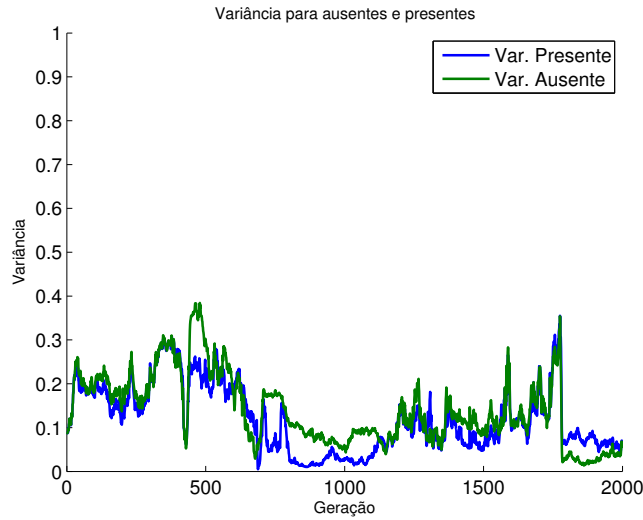


Figura 5.10 Variância média da relação signo-objeto R_p (Var. Presente) e R_a (Var. Ausente).

um conjunto de dados é medida através do somatório do quadrado da diferença entre os valores da variável e a média, conforme a fórmula

$$var(x) = \frac{\sum_{i=0}^N (x_i - \mu)^2}{N},$$

onde μ é a média da variável, e N é a quantidade de amostras.

Nosso conjunto de dados envolve as medidas de relação signo-objeto. As medidas de relação signo-objeto são calculadas através da média da similaridade entre padrões de ativação, dessa forma é através desses valores que podemos calcular as médias das variâncias das medidas de relação signo-objeto. Assim, o cálculo da variância das medidas de relação signo-objeto em uma geração, para o objeto presente ou para o objeto ausente, são realizados de acordo com a fórmula:

$$V_r = \frac{\sum_{i=1}^{42000} (R_x - MS_i)^2}{42000},$$

onde R_x é a relação signo-objeto, pois ela é justamente a média dos valores de média da similaridade (MS) entre todos os agentes em todas as rodadas ($42000 = 100 \cdot 420$). A figura 5.10 mostra a variância calculada para os dois valores na simulação em questão (a mesma das figuras 5.5, 5.6, 5.7, 5.8 e 5.9). Note que, indiretamente, a medida da variância da relação signo-objeto nos dá também uma noção da variância das ativações da camada intermediária para entradas comunicativas, pois seu valor está diretamente relacionado com o espalhamento dos pontos de ativação da camada intermediária gerados a partir da entrada comunicativa.

A figura 5.11 apresenta o gráfico com as três curvas nas quais estamos interessados para essa análise. Em azul temos a média de *fitness* da população, em verde o valor de diferença entre a relação signo-objeto com o objeto presente e com o objeto ausente

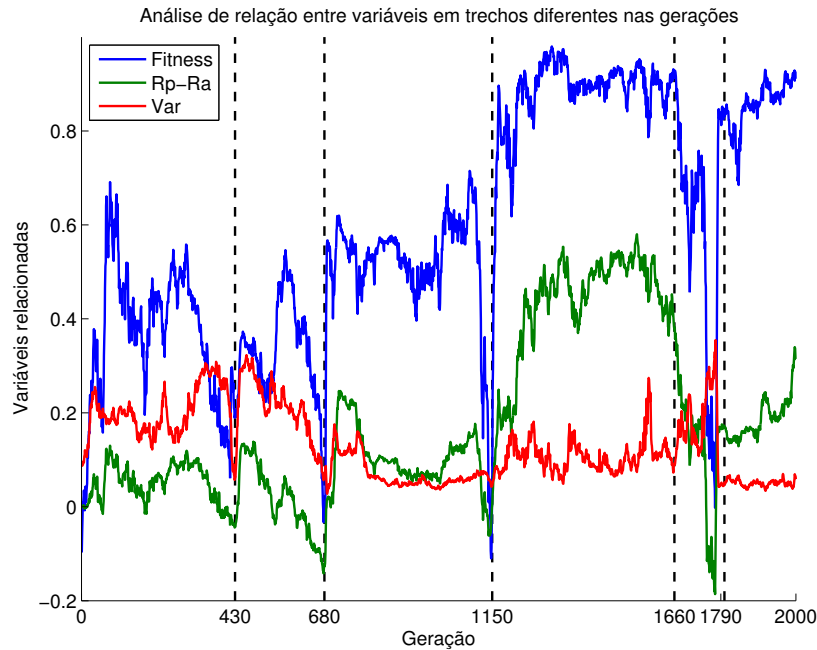


Figura 5.11 Gráfico que apresenta *fitness*, diferença entre as relações signo-objeto com objeto presente e com ausente ($R_p - R_a$), e a médias das variâncias dessas medidas (Var). Está dividido em seis intervalos, os quais possuem características de relações diferentes entre essas três variáveis.

(como na figura 5.6), e em vermelho a média aritmética das variâncias das relações signo-objeto apresentadas na figura 5.10. As linhas tracejadas na vertical, dividem o gráfico nos trechos que desejamos destacar. O gráfico foi dividido em seis trechos, entre os quais há alterações significativas ou do *fitness* médio da população ou da diferença entre a distância signo-objeto para o objeto ausente e para o objeto presente. As médias dos principais dados de cada uma dessas partes estão resumidas na tabela 5.1.

São oito os dados apresentados nessa tabela:

- **Trecho:** enumeração dos trechos em análise, para fins de organização e futuras referências no texto;
- **Intervalo:** o intervalo de gerações no qual os dados foram calculados;
- **Média $R_p - R_a$:** média da diferença entre a relação signo-objeto com o objeto presente e com o objeto ausente;
- ***Fitness* médio:** *fitness* médio da população;
- **Média das Variâncias:** média das variâncias das medidas de relação signo-objeto;
- **Correlação R_p :** medida de correlação entre o *fitness* e a relação entre signo e objeto presente no ambiente, no intervalo de gerações considerado;

Trecho	Intervalo	Correlação R_p	Correlação R_a	Correlação $R_p - R_a$	Média $R_p - R_a$	Média variâncias	<i>Fitness</i> médio
1	0-430	-0.67	0.91	0.90	0.0494	0.1998	0.3621
2	431-680	-0.35	0.28	0.32	0.0156	0.2106	0.3285
3	681-1150	-0.73	0.57	0.67	0.1010	0.0696	0.5130
4	1151-1660	-0.62	0.71	0.67	0.4293	0.1063	0.8835
5	1661-1790	-0.71	0.93	0.83	0.1287	0.1715	0.5916
6	1791-2000	-0.94	0.91	0.94	0.1867	0.0506	0.8573

Tabela 5.1 Tabela com dados calculados em intervalos distintos de gerações de uma execução.

- **Correlação R_a :** medida de correlação entre o *fitness* e a relação entre signo e objeto ausente no ambiente, no intervalo de gerações considerado;
- **Correlação $R_p - R_a$:** medida de correlação entre o *fitness* e a diferença da relação entre signo e objeto presente, e signo e objeto ausente no ambiente, no intervalo de gerações considerado.

A hipótese levantada, e corroborada pelos dados apresentados na seção 5.1.2, é de que as medidas de relação signo-objeto influenciam no *fitness*, na medida em que pesam na determinação da ação pelas criaturas intérpretes, por meio da qualidade de representação do objeto pelo signo.

O primeiro trecho apresentado na tabela engloba as primeiras 430 gerações da simulação. Durante as gerações iniciais (aproximadamente as 100 primeiras), há grande frequência de oscilações no *fitness*, em torno de valores baixos, até mesmo negativos. Nas primeiras gerações não há um sistema comunicativo efetivo, e os agentes ainda não categorizaram corretamente os cogumelos, sendo o motivo principal desse “caos” inicial. De maneira geral, durante esse trecho a média de *fitness* oscila bastante, e a média (do *fitness* desse trecho) é a segunda mais baixa dentre os trechos das gerações que dividem o gráfico. Por outro lado, a diferença $R_p - R_a$ também oscila em torno de zero, e vai a valores negativos nesse trecho, ficando também com o segundo menor valor médio absoluto entre os todos os trechos. Esse valor tão próximo a zero, pode indicar que as ativações geradas a partir do signo indicativo de um tipo de cogumelo estão muito próximas das ativações geradas pelo signo indicativo dos dois tipos de cogumelo. Além disso, a variância desse trecho é também a segunda maior dentre todos os intervalos. A baixa diferença $R_p - R_a$, somada à alta variância, produz uma dispersão de valores de ativação na qual é praticamente impossível o agente estabelecer o referente correto para todas as sinalizações comunicativas recebidas. A figura 5.12 pode ajudar a entender melhor essa questão. Ela apresenta as ativações da camada intermediária do melhor agente de uma geração do trecho analisado (da execução analisada), de acordo com o explicado no capítulo 4. Podemos perceber que há uma grande dispersão das ativações, originadas em ambas as entradas, e também uma grande quantidade de ativações originadas a partir de entrada comunicativa que disparam ações que não correspondem ao cogumelo presente no ambiente, em ambos os lados da linha de decisão.

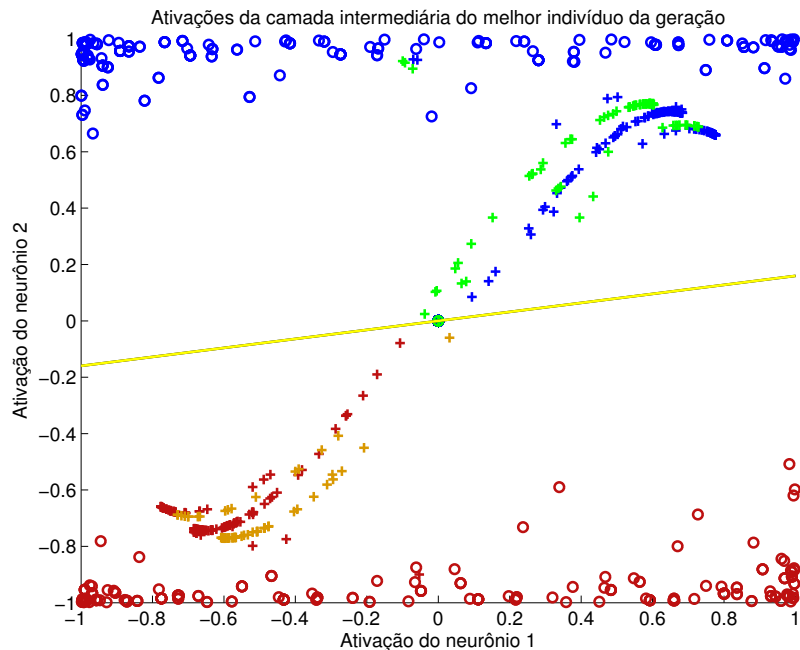


Figura 5.12 Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração do trecho 1.

Podemos ainda observar que, no trecho 1, as correlações entre as medidas de relação signo-objeto e o *fitness* são bastante significativas, mesmo com valores mais baixos de *fitness*. Nesse caso, é possível que a alta variância das ativações seja uma das responsáveis por essa boa correlação. O que ocorre é que estando as ativações próximas da linha de decisão e espalhadas, como nesse caso, pouca variação das ativações, em qualquer direção, resulta em grande número de ativações ocorrendo em lado diferente da fronteira de decisão, influenciando assim a interpretação do agente, modificando seu comportamento, e finalmente agindo sobre o *fitness* médio da população.

No trecho 2, percebemos uma queda gradual e consistente dos valores de diferença $R_p - R_a$ e de sua variância. Esse movimento é seguido pela curva do *fitness* e, na parte final, o *fitness* cai de modo mais acentuado que as outras duas curvas. Esse fato leva à relativa pouca correlação entre as medidas de distância signo-objeto e o *fitness*. De fato, os valores de correlação encontrados nesse intervalo não são significativos para nenhuma medida de distância signo-objeto, e são os menores dentre todos os intervalos. Os valores médios apresentados na tabela 5.1 podem nos ajudar a entender o que aconteceu. De todos os dados analisados, esse intervalo possui os piores valores. Em relação ao trecho 1, a diferença $R_p - R_a$, que já era pequena, neste segundo trecho é ainda bastante reduzida. Na prática, um valor de diferença entre distâncias signo-objeto tão pequeno e com variância tão alta, indica que não há uma relação representacional entre signo e objeto que permita determinar o tipo do cogumelo, com $R_p \approx R_a$, como argumentamos na seção 5.1.2.

No final do intervalo, a curva de diferença $R_p - R_a$ alcança valores negativos, o que parece também influenciar negativamente o *fitness*, que alcança valores próximos a zero

no final desse trecho. A variância, por sua vez, apesar de seguir uma tendência de queda, começa o trecho com valores elevados, e decai lentamente, o que resulta em uma média de variância no trecho 2 levemente maior do que no trecho 1 (aproximadamente 5%). Desses fatores podemos extrair possíveis explicações para a baixa correlação entre as medidas de distância signo-objeto e o *fitness*. Primeiro, uma possível explicação é que nesse trecho a diferença $R_p - R_a$ é muito menor, e em boa parte das gerações com valores negativos, indicando tendência ou à não comunicação, ou à comunicação “enganosa” dentro da população. Em conjunto com esse fato, devemos levar em conta também a variância das distâncias, que aumentou em relação ao trecho 1. Ou seja, o que acontece nessa situação é que além das distâncias entre signo e objetos diferentes serem bastante parecidas, elas ainda estão mais dispersas que no trecho anterior. Esse cenário aumenta as chances dos agentes cometerem enganos em relação ao sinal recebido, resultando em um *fitness* médio aproximadamente 10% menor que no trecho 1.

O terceiro trecho apresenta um aumento significativo da média de *fitness*. Nesse trecho, a diferença $R_p - R_a$ aumentou em aproximadamente 2 vezes em relação ao trecho 1, e quase sete vezes em relação ao trecho 2, não mais variando entre valores positivos e negativos, mas apenas positivos. Devemos lembrar que essa diferença de relação signo-objeto, quando positiva, indica na média o quanto a relação signo-objeto para o objeto presente é melhor que a relação signo-objeto para o objeto ausente. Logo, é de grande importância para avaliarmos o quão bom está o processo de representação que ocorre durante a simulação. Além do aumento da diferença $R_p - R_a$, a variância das distâncias desse trecho é aproximadamente 3 vezes menor que nos 2 trechos anteriores. Dessa forma, temos o seguinte cenário: as ativações da camada intermediária originadas a partir de entrada comunicativa estão mais próximas das ativações geradas a partir de entrada visual para o tipo de cogumelo presente no ambiente no momento em que o signo foi recebido e, além disso, estão mais concentradas. Esse cenário aumenta as chances de se estabelecer um sistema de comunicação confiável, uma vez que a interpretação correta do signo por parte do agente ouvinte é facilitada, com as ativações originadas a partir do recebimento do signo concentradas próximas às ativações originadas a partir da entrada visual, diminuindo bastante a possibilidade do agente cometer um “equivoco” na ação tomada.

As partes inicial e final deste trecho apresentam ainda um exemplo de como a variância das distâncias pode influenciar na relação entre diferença $R_p - R_a$ e *fitness*. No início desse trecho ocorre o maior pico de distância $R_p - R_a$, seguido por uma queda no valor e um outro pico, menor que o inicial, no final desse intervalo. Porém, apesar do pico inicial ser maior, é no período do pico final que o *fitness* alcança maiores valores. Entretanto, é possível verificar visualmente na figura 5.11, que durante o pico inicial a variância da relação signo-objeto era significativamente maior que durante o pico final, o que pode ter compensado o menor índice de relação signo-objeto, e facilitado a boa interpretação dos sinais comunicativos. De maneira geral, a média do *fitness* nesse trecho aumentou aproximadamente 35% em relação ao trecho 2.

A figura 5.13 ajuda a entender esse cenário. Há poucas ativações próximas à fronteira de decisão, e a região de maior densidade de concentração está aproximadamente a meio caminho entre a fronteira de decisão e à concentração de ativações geradas a partir de

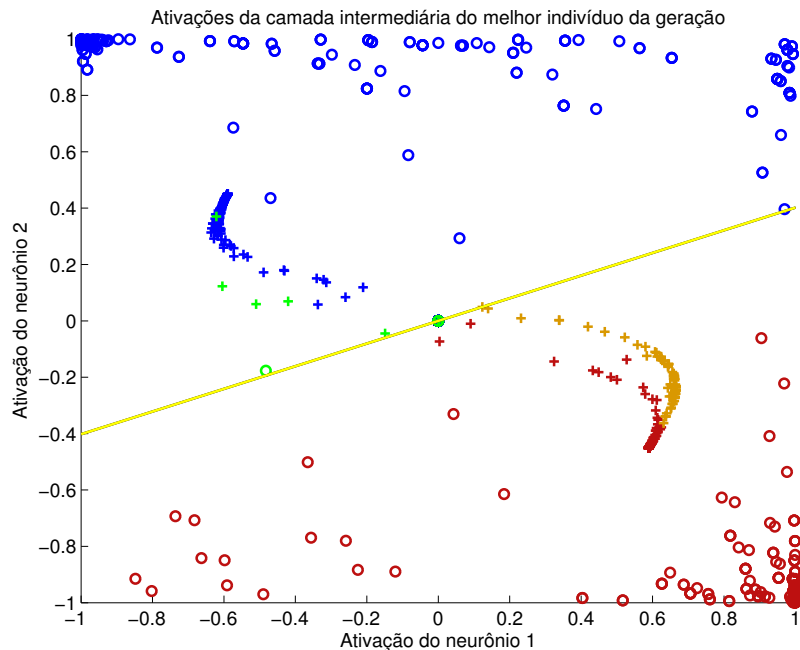


Figura 5.13 Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração do trecho 3.

entrada visual. Por isso, também a correlação entre as medidas de distância signo-objeto foram significativas. As ativações originadas a partir de entradas comunicativas estão a certa distância da linha de decisão, significativamente maior que zero, enquanto praticamente todas as ativações originadas a partir de entradas visuais ficam sempre próximas de valores extremos de um dos eixos, afastando-se ao máximo da linha de decisão. Dessa forma, se a distância entre as ativações originadas de entradas diferentes, para um mesmo tipo de cogumelo, aumenta, indiretamente podemos inferir que a distância das ativações originadas a partir de entradas comunicativas ficam mais próximas da fronteira de decisão, aumentando assim o risco de “equivoco” do agente. E, do contrário, essas ativações se afastam da fronteira de decisão, diminuindo a possibilidade de “equivoco” do agente.

No quarto trecho a diferença $R_p - R_a$ aumentou aproximadamente 4 vezes, em relação ao trecho três, alcançando seu maior valor médio. Porém, a média do *fitness* aumentou em aproximadamente 40% em relação ao trecho 3. Notamos porém que, apesar desse grande aumento na diferença $R_p - R_a$, que favorece os processos de representação, há, em contrapartida um aumento de praticamente 35% da variância média em relação ao trecho 3. Assim, mesmo com valor médio de diferença $R_p - R_a$ muito maior do que todos os outros trechos, o *fitness* médio não parece refletir proporcionalmente esse fato, uma vez que outro fator, que tende a contribuir negativamente para ele, aumentou seu valor. Por outro lado, mesmo tendo uma variância média significativamente maior que o trecho 3, a média de *fitness* no trecho 4 é significativamente maior que no trecho dois. O aumento de um fator parece compensar o aumento do outro. Vejamos como isso pode ocorrer.

Considere a figura 5.14, que apresenta dados relativos a ativações da camada inter-

mediária do melhor agente em uma geração entre 1151 e 1660 da simulação em questão. Nela vemos que a maioria das ativações, geradas tanto a partir de entrada comunicativa quanto de entrada visual se concentram nos extremos do gráfico. A maior concentração das ativações originadas a partir de ambas as entradas estão bastante próximas, em seus respectivos lados da linha de decisão. E a distância entre as médias das ativações originadas de entradas comunicativas que geram determinada ação, estará praticamente na distância máxima das ativações originadas a partir de entradas visuais que geram a ação contrária. Em outras palavras, a média das ativações da camada intermediária que está sendo gerada a partir de entradas comunicativas quando determinado tipo de cogumelo está presente no ambiente é bastante próximo da média das ativações da camada intermediária quando os agentes visualizam esse mesmo tipo de cogumelo, caracterizando uma boa relação do signo com o objeto presente (R_p), e distante da média das ativações de quando os agentes visualizam o outro tipo de cogumelo, caracterizando uma baixa relação do signo com o objeto ausente (R_a). Essa proximidade de ativações da camada intermediária para entradas originadas a partir de entradas diferentes, mostra que o agente consegue fazer uma boa interpretação signo-objeto, consequentemente conseguindo associar o signo recebido à ação correta. Se a média da distância é muito boa, por outro lado, há uma variância considerável nessas distâncias, maior que a apresentada no trecho anterior. Porém, como na média as ativações originadas com tipos de cogumelos diferentes presentes no ambiente estão bastante longe umas das outras, pouquíssimas das ativações espalhadas chegam próximas da linha de decisão do agente. Dessa forma, o risco do agente realizar a interpretação de forma equivocada, diminui bastante. Por esse motivo, mesmo com a variância maior que no trecho anterior, o aumento da distância foi tal que compensou o maior espalhamento das ativações, indicando a possibilidade de existir um sistema de comunicação confiável.

O quinto trecho apresenta uma dinâmica com alta frequência de grandes variações nos valores, principalmente do *fitness*. Nesse caso, faz pouco sentido se falar em médias de valores, já que praticamente não há estabilidade nas curvas. Especificamente, esse intervalo mostra uma queda abrupta do *fitness*, seguida de uma alta, também repentina. Esse movimento é acompanhado de uma queda também abrupta da diferença $R_p - R_a$, que alcança seus menores valores de toda a simulação (próximos a -0.2, como podemos verificar na figura 5.11), e também de sua seguida alta. Nesse intervalo, o *fitness*, apesar de cair muito, não alcança seus menores valores da simulação. Por exemplo, valores menores podem ser verificados próximos a geração 1150. Nesse momento, a variância das distâncias signo-objeto subiram bastante, alcançando seus maiores valores. Utilizaremos a figura 5.15 para nos ajudar a entender o por quê do motivo de o *fitness* não ter caído tanto, ter sido justamente a alta variância das distâncias signo-objeto.

Os valores negativos na diferença $R_p - R_a$ indicam a existência de um número significativo de “enganadores” na população e, de fato, podemos notar que houveram várias ações equivocadas por parte do melhor agente da geração em questão, quando este recebia sinais comunicativos. Por outro lado, as ativações originadas a partir de sinais comunicativos estão muito espalhadas por quase todo o gráfico. Assim, mesmo que a média das ativações originadas a partir de entradas comunicativas esteja se afastando das ativações originadas a partir de entradas comunicativas para o mesmo tipo de cogumelo,

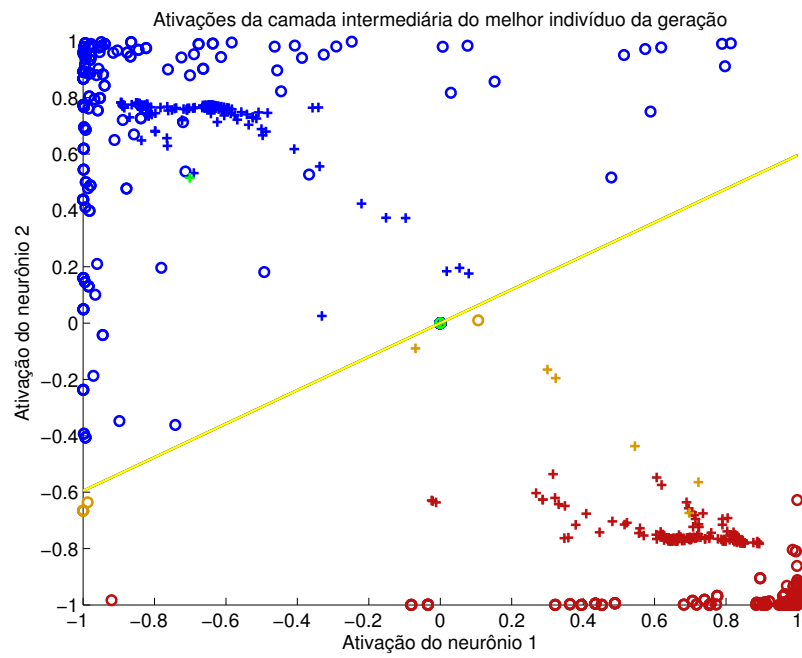


Figura 5.14 Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração do trecho 4.

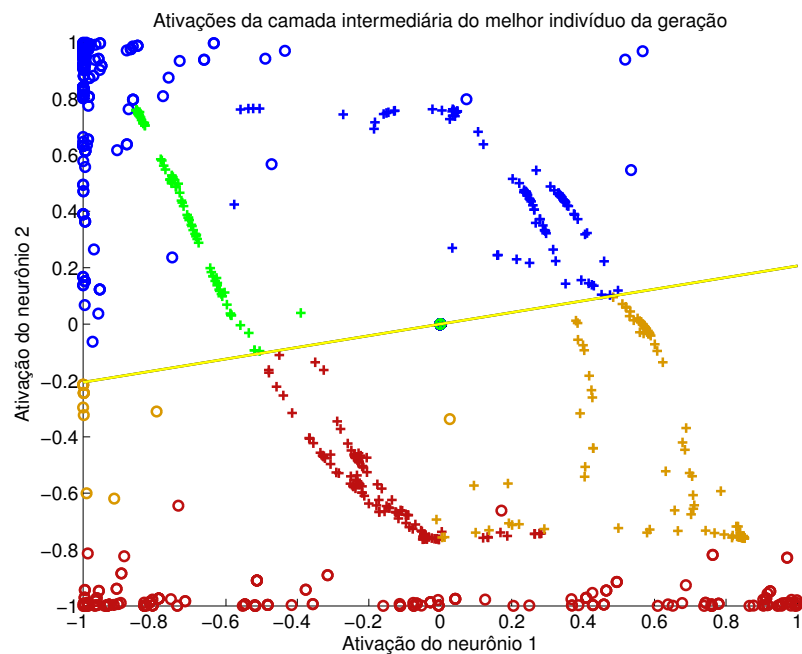


Figura 5.15 Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração do trecho 5.

se deslocando no sentido de cruzar a linha de decisão, sempre restarão algumas ativações significativamente distantes da fronteira de decisão, de forma que o agente poderá realizar a interpretação de forma correta. E esse número de ativações distantes da fronteira de decisão pode ser maior na medida em que a variância é maior. Por isso o *fitness* não cai tanto como em outras gerações, mesmo com a diferença $R_p - R_a$ em seu menor valor.

Esse caso parece ser o inverso de outros casos ocorridos em outros intervalos de gerações (talvez com exceção do primeiro trecho). Enquanto se tem um valor de diferença $R_p - R_a$ positivo, a pouca variância diminui o risco de interpretações “equivocadas” por parte dos agentes, uma vez que reduz a possibilidade de uma ativação originada a partir de um sinal comunicativo ocorrer longe do outro lado da fronteira de decisão. Porém, quando a diferença $R_p - R_a$ é negativa, a probabilidade maior é que as ativações originadas a partir de sinal comunicativo sejam interpretados de forma “incorreta” pelo agente, ou seja, estejam do lado “errado” da fronteira de decisão e, nesse caso, a maior variância aumenta a probabilidade de que algumas ativações ocorram do lado “certo” da fronteira de decisão, aumentando a probabilidade de interpretações corretas por parte do agente.

No sexto e último trecho há uma pequena diminuição do *fitness* médio em relação ao trecho 4 (cerca de 3%). Podemos supor que o principal motivo para essa pequena queda seja a grande diminuição da média da diferença $R_p - R_a$, que é menor do que a metade da registrada no trecho quatro. Porém, tamanha diminuição na diferença $R_p - R_a$, foi em parte compensada pela também diminuição da variância das relações, que é a menor registrada dentre todos os trechos. A figura 5.16, que apresenta dados relativos a ativações da camada intermediária do melhor agente em uma geração entre 1791 e 2000 da simulação em questão, pode nos ajudar a entender melhor esse cenário. Note que as ativações originadas a partir de entradas comunicativas estão praticamente todas alinhadas, bastante coesas e organizadas. Por outro lado, estão relativamente distantes das ativações geradas a partir de entradas visuais para o mesmo tipo de cogumelo, ao mesmo tempo que algumas poucas ativações ficam muito próximas da linha de decisão do agente. Também por esses motivos, as correlações entre o *fitness* e as medidas de distância signo-objeto alcançaram seus maiores valores dentre todos os trechos analisados, uma vez que nesse caso, qualquer variação na distância $R_p - R_a$ significa indiretamente que a média das ativações da camada intermediária geradas a partir de entradas auditivas se aproximaram ou se afastaram da linha de decisão e, como a variância é pequena, essas variações irão influenciar fortemente na possibilidade de equívoco de associação signo-objeto por parte do agente, em sua interpretação, conseqüentemente influenciando bastante suas tomadas de decisão.

Com essas avaliações, podemos perceber que o valor de correlação entre *fitness* médio da população e as medidas de médias da relação signo-objeto parecem dependentes também da variância das medidas de relação signo-objeto (e indiretamente da variância das ativações da camada intermediária da rede). Quanto mais dispersas as ativações, maior a chance dos agentes da população de “se enganarem”, e tomarem a ação “equivocada” (em relação ao tipo de cogumelo presente no ambiente). Por isso, parece razoável que, mesmo com medidas de relação signo-objeto tendo um bom valor médio, se a dispersão dos seus valores (que é medida através da variância) for grande, há realmente maior chance de haver menor correlação entre o *fitness* e essas medidas. Também o contrário

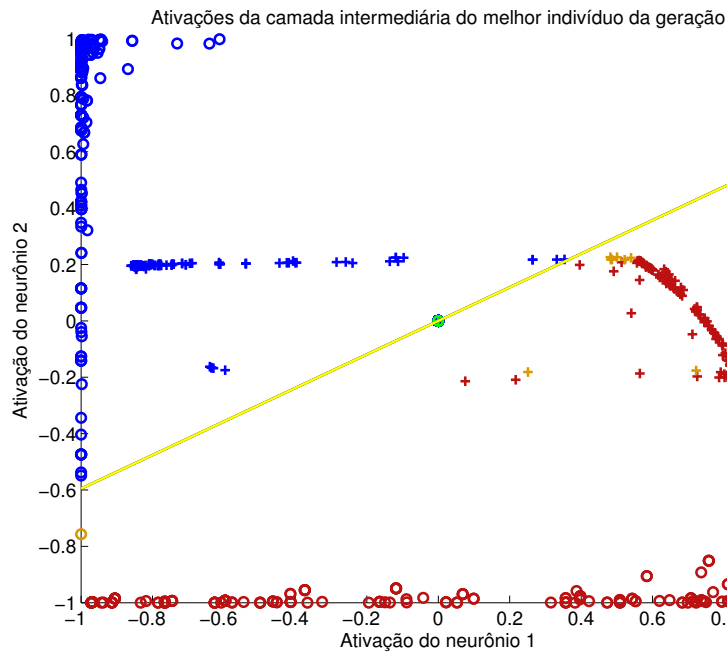


Figura 5.16 Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração do trecho 6.

é uma possibilidade razoável, mesmo com medidas de relação signo-objeto não sendo tão boas, se a variância das ativações da camada intermediária for bastante pequena, as chances de ações “equivocadas” dos agentes diminuí, e a influência sobre o *fitness* pode ser maior, uma vez que relacionar signo e objeto fica mais fácil para o agente. Inclusive, podemos notar que nos 4 pontos nos quais o *fitness* tem forte queda nessa execução, estas quedas acompanham quedas também bruscas da diferença $R_p - R_a$ e/ou aumentos repentinos da variância.

Os momentos nos quais o *fitness* médio da população cresce coincidem com momentos nos quais os ouvintes conseguem relacionar de forma apropriada o signo recebido com o objeto presente no ambiente. Nesses momentos, o signo é capaz de representar seu objeto em algum aspecto e o intérprete é capaz de relacionar signo e objeto de forma correta (do ponto de vista do objeto presente no ambiente), levando a um efeito interpretante benéfico para o ouvinte. Pudemos avaliar, pelo menos em parte, a maneira pela qual isso ocorre através da análise das ativações da camada intermediária da rede. Quando as ativações originadas a partir de entradas diferentes (visual e comunicativa) e para um mesmo tipo de cogumelo estão próximas (e distantes se forem cogumelos de tipos diferentes), e as ativações da camada comunicativa estão concentradas, tem-se as melhores condições para que esse processo ocorra. Note que a diferenciação entre os cogumelos não precisa ser entre cada um deles (que são de fato diferentes, dando um total de 420), mas entre os tipos existentes (apenas dois, comestível e venenoso). Disso podemos derivar duas conclusões: uma é que o aspecto que o signo deve representar é justamente o tipo de cogumelo presente no ambiente, ou seja, deve indicar de forma genérica a classe a qual

aquele objeto pertence, e não indicar um objeto em específico; e que a interpretação feita dessa maneira genérica é uma característica das relações simbólicas, que ocorrem nesse experimento.

5.2 FALANTES E OUVINTES COM SIMILARIDADE GENÉTICA

Com a modificação apresentada nessa seção, pretendemos que a comunicação seja facilitada, o *fitness* médio das simulações seja maior, e que não ocorram as variações cíclicas observadas no experimento original. Dessa forma, realizamos na dinâmica do experimento uma modificação que foi não escolher de forma aleatória o agente falante da rodada, mas colocar em seu lugar uma cópia geneticamente idêntica do agente ouvinte. Floreano et al. (2007) utilizam uma configuração parecida com essa em seu experimento, com a diferença principal de que os indivíduos geneticamente iguais estão distribuídos em grupos. Esse tipo de configuração se baseia em alguns estudos que indicam que, para que um indivíduo membro de uma população transmita informações “honestas” para outros membros mesmo se isso acarretar em custo para ele, é preciso considerar fatores como parentesco, similaridade genética, e conflito (ou confluência) de interesses por parte dos indivíduos envolvidos nesse processo (FLOREANO et al., 2007; SMITH, 1991). Com os agentes interagindo sempre com um agente geneticamente idêntico, o grau de parentesco e similaridade genética será o maior possível, e não haverá nenhum conflito de interesses, pois o sucesso do ouvinte (no sentido de conseguir se reproduzir e passar seus genes adiante) é também o sucesso do falante.

Foram realizadas 40 execuções do experimento com essa modificação e todas apresentaram valores de *fitness* muito parecidos entre si, sempre com valores de média de *fitness* para as 2000 gerações bastante altos (próximo a 1), e sem variar ciclicamente, como ilustra a figura 5.17, que mostra o gráfico do *fitness* médio da população para uma das execuções nessa configuração. Podemos perceber que os valores iniciam baixos (próximos a zero), como na configuração original, por conta da geração aleatória dos pesos iniciais dos agentes, mas em poucas gerações o *fitness* atinge valores próximos ao máximo de 1, e, a partir desse ponto, passam sofrer pequenas variações, sempre com valores muito próximos a 1. Tal característica fez com que sua média de *fitness* durante as simulações fosse de aproximadamente 0.96, ante a média de 0.66 obtida no experimento original. Comparando os valores médios do *fitness* em todas as 2000 gerações das execuções do experimento original e das execuções com falantes e ouvintes geneticamente idênticos, constata-se que de fato as execuções com essa última configuração possuem valores significativamente maiores (teste de Mann-Whitney, $p \cong 6.79 \cdot 10^{-8}$) do que as simulações com experimento original. Podemos analisar então, se as medidas de relação signo-objeto seguem comportamento proporcional ao experimento original, e se existe o mesmo tipo de correlação que existe no experimento original.

A figura 5.18 mostra os dados da diferença entre a relação signo-objeto com o objeto presente, e a relação signo-objeto com o objeto ausente ($R_p - R_a$) de duas execuções do experimento nessa configuração. Essas execuções foram escolhidas de modo a apresentar algumas características de interesse para nossa discussão. A figura 5.18(a) traz um gráfico cujos valores variam em torno de 0.2 (com média de 0.24), enquanto a figura

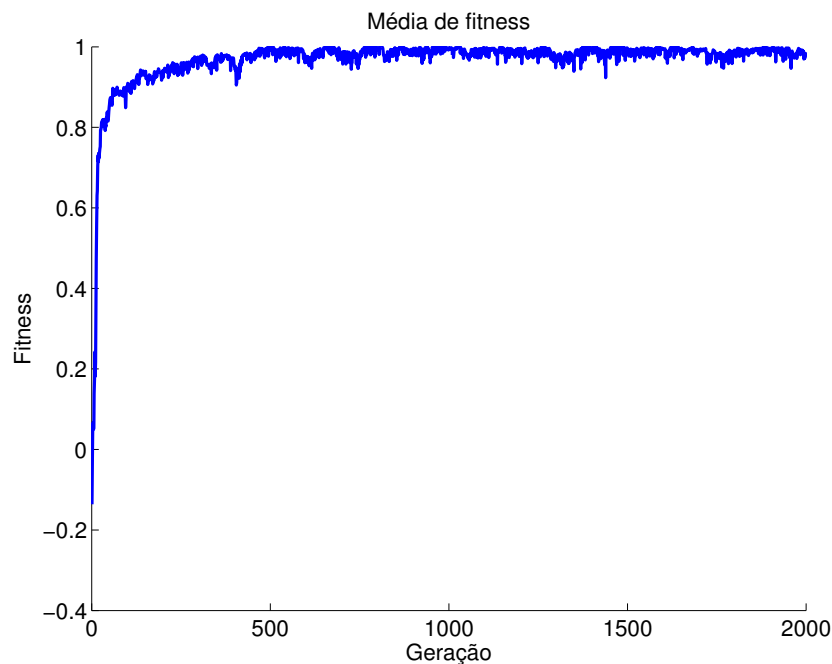
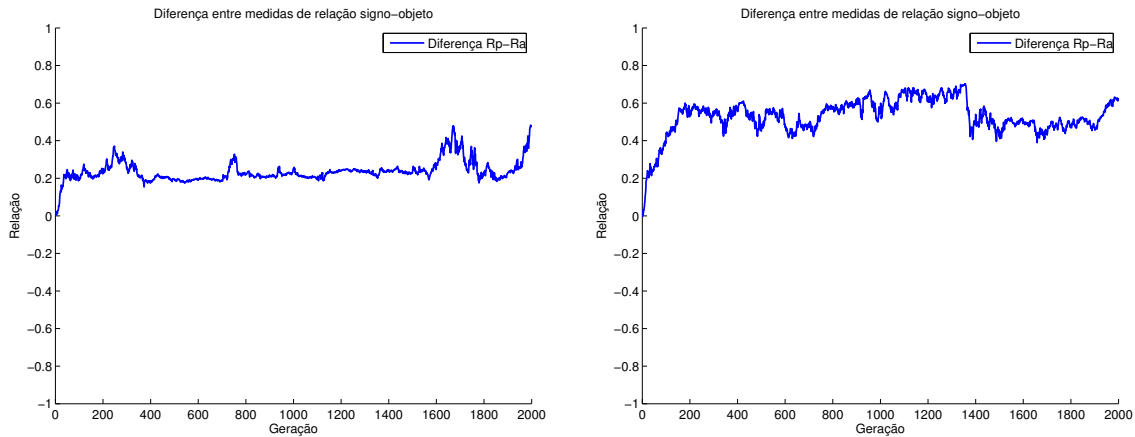


Figura 5.17 Exemplo de curva de *fitness* obtidas nas execuções da simulação com agentes ouvinte e falante geneticamente similares.

5.18(b) traz um gráfico cujos valores variam em torno de 0.6 (com média de 0.53). Os valores alcançados por essas medidas são valores que normalmente são alcançados em simulações do experimento original, porém, são muito mais constantes, e além disso, não houve casos de valores negativos, que indicariam a tentativa de atuação de agentes enganadores. Apesar de se manterem muito constantes ao longo das gerações, de forma similar ao *fitness*, há uma diferença bastante interessante entre essas duas curvas. Ao contrário do *fitness*, que é praticamente igual para todas as simulações, as medidas de relação signo-objeto variam bastante entre as execuções, nem sempre alcançando valores altos. Além disso, apesar de poucas e geralmente de pequena amplitude, as variações nas medidas de relação signo-objeto não parecem gerar nenhum efeito sobre o *fitness*. Ou seja, mesmo com essas medidas variando, a capacidade dos intérpretes de estabelecer uma forte relação entre signo o objeto, gerando um efeito interpretante correto (do ponto de vista do cogumelo presente no ambiente), permanece praticamente inalterado.

Uma consequência dessas características é que a correlação entre as medidas de relação signo-objeto e o *fitness* é bastante baixa, ao contrário do que ocorre nas simulações com a configuração original do experimento. Na figura 5.19, os gráficos apresentados em 5.19(a), 5.19(b) e 5.19(c) apresentam os gráficos de dispersão dos dados de *fitness* e das medidas de relação signo-objeto (de forma semelhante as figuras 5.7, 5.8 e 5.9). Os índices de correlação ρ são, aproximadamente, 0.38, -0.2 e 0.27, para relação signo-objeto com o objeto presente, com o objeto ausente, e para a diferença entre esses dois valores, respectivamente. Devemos notar que praticamente todos os pontos que indicam *fitness*



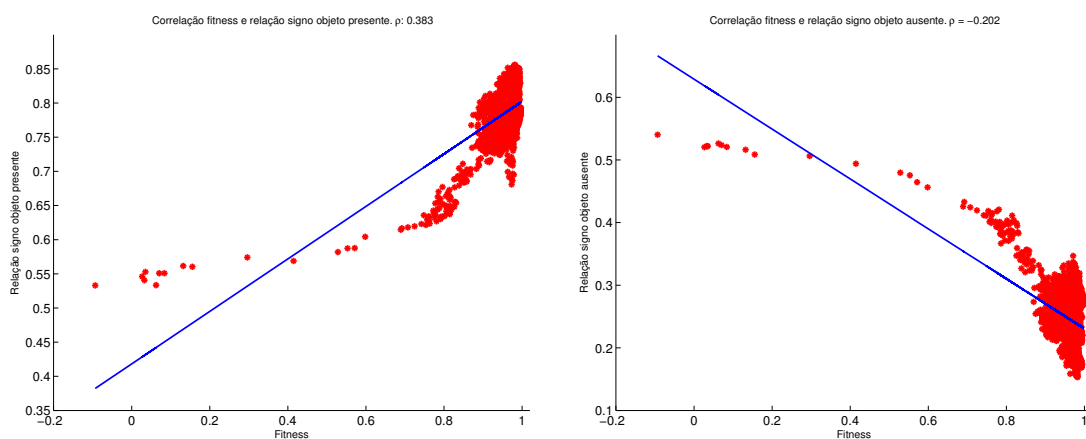
(a) Execução com valores da diferença $R_p - R_a$ com variações em torno de 0.2. (b) Execução com valores da diferença $R_p - R_a$ com variações em torno de 0.6.

Figura 5.18 Exemplos de duas execuções da simulação com interação entre agentes geneticamente similares, nos quais há grande diferença na média dos valores da diferença $R_p - R_a$.

estão aglomerados entre 0.8 e 1, pouco importando qual o valor do índice de proximidade, o que impede que uma função monotônica possa descrever bem essa relação.

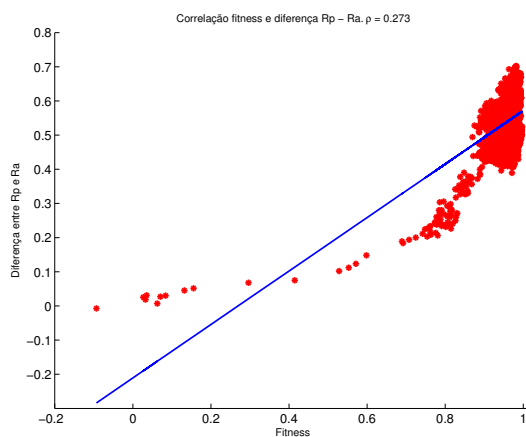
Mas, se pouco parece importar para os valores de *fitness* o valor das medidas de relação signo-objeto, já que os valores de *fitness* são sempre altos, mesmo que as medidas de relação signo-objeto sejam relativamente ruins, a que se deve esse *fitness* tão elevado? Primeiro é importante deixar claro que, apesar de nem sempre apresentar valores relativamente elevados em todas as simulações, as medidas de distância signo-objeto sempre apresentaram valores relevantes (maiores que zero). Esse já é um fator essencial para o surgimento de um bom sistema de comunicação, e conseqüentemente um bom *fitness*. Além disso, de acordo com o que discutimos na seção anterior, um outro fator, que além de auxiliar no surgimento de um bom sistema de comunicação, pode ainda explicar por que o *fitness* alcança valores elevados mesmo para valores relativamente baixos dos índices de relação signo-objeto, é justamente a variância dos valores das medidas de relação signo-objeto.

Na seção 5.1.3 vimos justamente que, a depender dessa variância, podem ocorrer *fitness* semelhantes para medidas de relação signo-objeto diferentes, por conta da menor possibilidade de “engano” dos agentes ouvintes. Nessa configuração do experimento, os valores de variância são significativamente menores do que os valores de variância no experimento na configuração original, de acordo com o teste de Mann-Whitney ($p = 0.00033$). A figura 5.20 apresenta valores de variância para a configuração modificada com interação entre agentes com alta similaridade genética. Como todas as execuções apresentam curvas com valores e padrões similares, escolhemos livremente a execução que originou essa figura. Note que os valores que esse gráfico apresenta são bem menores que os valores apresentados no gráfico da figura 5.10, que mostra os dados de variância de uma simulação com a configuração original.



(a) Gráfico de dispersão do *fitness* e relação signo-objeto com o objeto presente.

(b) Gráfico de dispersão do *fitness* e relação signo-objeto com o objeto ausente.



(c) Gráfico de dispersão do *fitness* e diferença de relação signo-objeto do objeto presente para o ausente.

Figura 5.19 Gráficos que mostram a correlação praticamente inexistente entre as medidas de relação signo-objeto e o *fitness*.

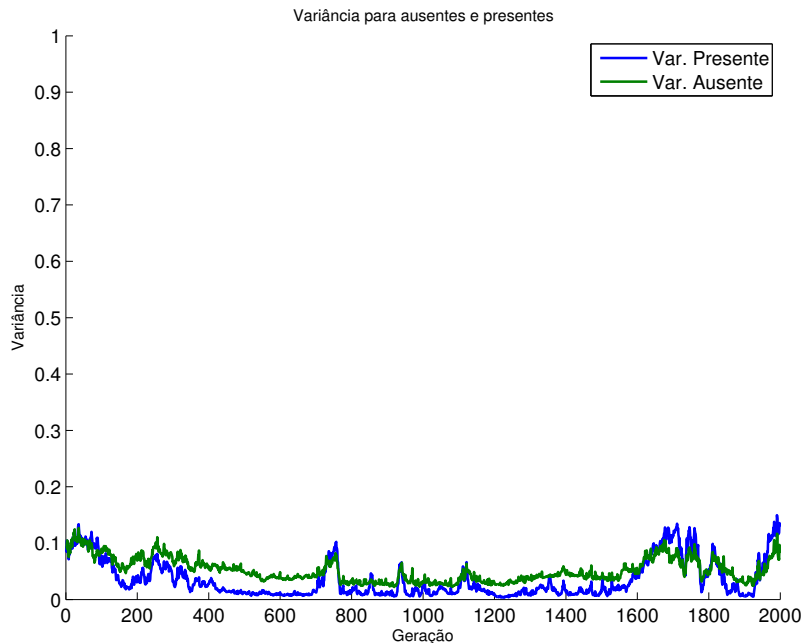


Figura 5.20 Exemplo de curva de variância das relações signo-objeto R_p e R_a em execução de simulação com interação entre agentes geneticamente similares.

A baixa variância das ativações da camada intermediária faz com que, mesmo valores de medidas de relação signo-objeto relativamente baixos sejam suficientes para a boa interpretação, e que a variação dos valores de relação signo-objeto não tenham correlação com as variações do *fitness*. Para ajudar a entender o porquê desse comportamento, devemos considerar a mudança do fator genético na população, e considerar novamente a linha de decisão das ações do agente (apresentada em 4.3.2), e de que forma eles influenciaram nessa configuração do experimento. Nessa configuração, o ouvinte e o falante são os mesmos indivíduos, ou seja, são geneticamente iguais e possuem o maior grau de parentesco possível (apesar de, teoricamente, não ser necessário tamanho grau de parentesco para que haja cooperação), assim, tanto falante quanto ouvinte tem o “objetivo” de replicação dos mesmos genes, e não haverá tentativa de engano por parte dos falantes.

A figura 5.21 apresenta o gráfico de ativações da camada intermediária do melhor indivíduo de uma das simulações realizadas com essa configuração, em uma geração após o período inicial de crescimento do *fitness*. Notamos claramente que as ativações originadas a partir das entradas comunicativas no gráfico da figura 5.21 estão muito mais concentradas que todos os gráficos dessa natureza mostrados até aqui neste trabalho, e também estão razoavelmente distantes da linha da fronteira de decisão. Uma explicação razoável para a maior coesão das ativações nesse gráfico é a menor variação genética da população, que tende a gerar sinais semelhantes, e a interpretá-los também de forma semelhante. A maior coesão das ativações geradas a partir das entradas comunicativas é justamente o fator que resultará em uma variância menor na relação signo-objeto. Logo,

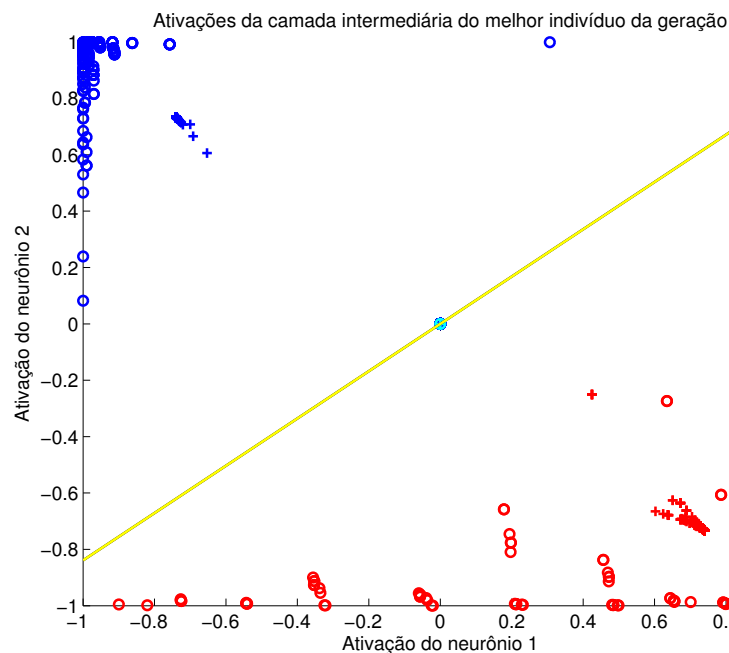


Figura 5.21 Gráfico de pontos de ativação da camada intermediária do melhor indivíduo de uma geração durante uma execução do experimento com interação entre agentes geneticamente similares.

a pouca correlação entre as medidas de relação signo-objeto e *fitness* pode ser explicada por essa pouca variância das ativações, uma vez que estando coesas dessa forma, a média da relação pode variar bastante antes que a interpretação dos signos por parte do agente seja comprometida, pois é preciso que as distâncias apresentadas na figura 5.18 sejam muito baixas para que as ativações de um lado da fronteira de decisão “invada” o outro lado, causando “equivocos” na interpretação dos agentes.

Já para o distanciamento dessas ativações da linha de decisão, há outra explicação. No experimento original são comuns as mudanças de interpretação dos sinais emitidos, visto que enganadores surgem e se espalham pela população. Para escapar dos enganadores, os agentes têm que se adaptar, muitas vezes invertendo as ações tomadas para determinados sinais recebidos (fenômenos que geram as “quebras” do sistema de comunicação, e conseqüentemente do *fitness*, conforme foi discutido na descrição do experimento, na seção 4.2 do capítulo 4). Por conta dessa característica, as ativações geradas a partir de sinais comunicativos tendem a ficar próximas da fronteira de decisão, uma vez que é menos custoso para o agente que assim estiver organizado realizar essa inversão de tomada de ação, visto que precisará alterar relativamente pouco os seus genes para se adaptar rapidamente as constantes mudanças no sistema de comunicação, mudando as ativações de um lado para o outro da fronteira de decisão mais facilmente. Por outro lado, quando agentes geneticamente iguais são postos para interagir, não há o fenômeno da inversão de valores, pois, como já discutimos, tanto ouvinte quanto falante estão cooperando para um objetivo comum. Por isso, as ativações geradas a partir de sinais comunicativos para

essa configuração, não precisam ficar próximos à fronteira de decisão. Como não há essa necessidade e ativações próximas à fronteira de decisão estão sujeitas a um risco maior de engano, a tendência é que nessa configuração elas se afastem da fronteira de decisão.

Com todas essas características, as criaturas conseguem sempre relacionar signo e objeto de maneira correta mais facilmente do que na configuração original do experimento. Os signos recebidos sempre geram ativações na camada intermediária que ficam bastante próximas das ativações geradas pela visualização dos respectivos tipos de cogumelo (objetos), o que faz com que as interpretações estejam quase sempre corretas, gerando um efeito interpretante adequado ao tipo de cogumelo presente no ambiente, resultando na sempre elevada média de *fitness* da população.

5.3 SELEÇÃO POR GRUPO

A outra modificação realizada na dinâmica do experimento foi a divisão da população em grupos, dentro dos quais os agentes interagem, e sobre os quais o mecanismo de seleção atua. De forma semelhante à modificação descrita na seção 5.2, a seleção por grupo tende a favorecer o surgimento de um sistema de comunicação confiável e duradouro (sem ciclos de ruptura), uma vez que favorece a cooperação entre os membros de um mesmo grupo, devido à pressão seletiva atuar sobre o grupo e não sobre o indivíduo. A modificação realizada no experimento das simulações analisadas nessa seção divide a população de 100 agentes em 20 grupos de cinco indivíduos, de maneira aleatória. A interação entre os agentes ocorre apenas entre indivíduos de um mesmo grupo, ou seja, quando um agente é ouvinte, apenas agentes pertencentes ao mesmo grupo que ele podem ser selecionados para se tornar falantes. A seleção para reprodução acontece por grupos, e estabelece que quanto maior a soma dos *fitness* dos indivíduos de um grupo, maior a chance daquele grupo ser selecionado para reprodução. Para a reprodução, são sempre escolhidos dois grupos, e um indivíduo de cada grupo é escolhido para fazer o cruzamento e gerar um novo indivíduo. Depois do processo de reprodução os indivíduos são novamente divididos de forma aleatória em grupos, e o ciclo recomeça.

Os valores obtidos pelas 40 execuções realizadas com essa configuração foram intermediários, ficando entre os valores obtidos com a simulação do experimento original e os obtidos com as simulações com o experimento com interação apenas entre agentes similares geneticamente. Começamos pela avaliação do *fitness*. Para termos uma noção melhor de como os valores de *fitness* médio estão colocados em relação ao experimento original e à simulação com agentes iguais interagindo, a figura 5.22 mostra a média de *fitness* de uma simulação com o experimento com grupos. Comparando com as figuras 5.1 e 5.17 podemos perceber seu valor intermediário. Constata-se que os valores médios do *fitness* do experimento com grupos é significativamente maior que os valores médios do experimento normal (teste Mann-Whitney, $p \cong 9.2 \cdot 10^{-5}$), e significativamente menor que o experimento de interação entre agentes geneticamente iguais (teste Mann-Whitney, $p \cong 6.8 \cdot 10^{-8}$).

Podemos notar ainda no gráfico da figura 5.22 que também a forma da curva do experimento com grupos parece intermediária entre as outras duas. Na simulação com o experimento original o *fitness* médio da população sobe e desce muito constantemente,

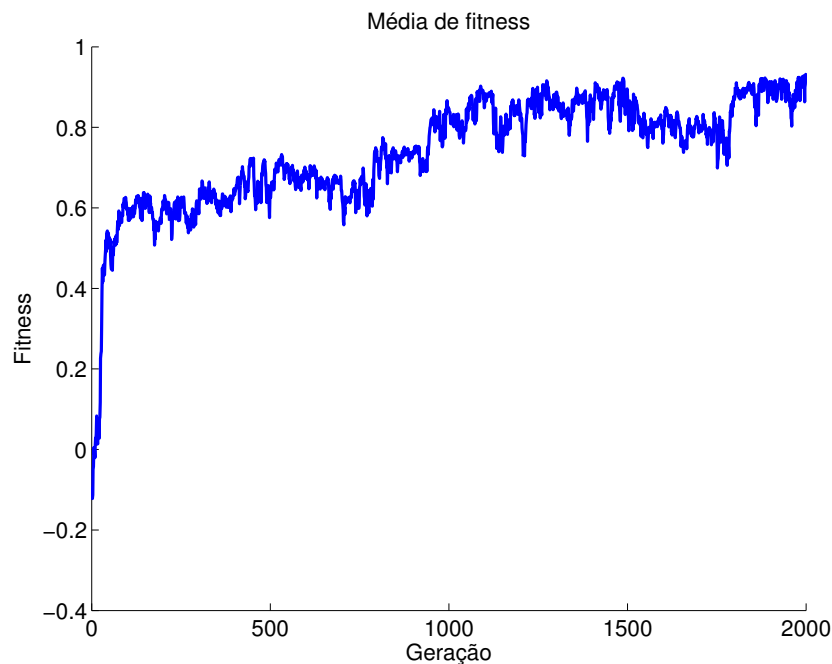


Figura 5.22 Exemplo de curva de *fitness* obtidas nas execuções da simulação com interação e seleção por grupo.

e com amplitudes bastante grandes (entre valores negativos e próximos a 1). No caso do experimento original, entre uma simulação e outra há muita variação de formas de curva, sendo que a partir de uma curva é possível apresentar a natureza da variação cíclica dos valores do gráfico, mas não é possível representar a forma de todas as curvas obtidas nas simulações. Já nas simulações com experimento de interações entre agentes similares geneticamente, o *fitness* médio da população sobe rapidamente e permanece praticamente constante próximo ao valor 1, sofrendo alguma flutuação por conta dos cruzamentos genéticos e de mutações. No caso dos agentes geneticamente similares, as curvas da média do *fitness* entre execuções diferentes são praticamente iguais, com um exemplar sendo capaz de representar com razoável grau de fidelidade todas as demais curvas obtidas em outras execuções, tanto em valor quanto em forma. As simulações com interação e seleção por grupo, por sua vez, apresentam um comportamento intermediário, seu valor sobe rapidamente até um patamar menor que a média do experimento com interação entre agentes iguais, e passa a oscilar um pouco, não tanto como no experimento original (em menor frequência, e em menor amplitude, não chegando nem perto de valores negativos), mas visivelmente mais que no experimento de interação entre agentes iguais. A partir de uma curva é possível apresentar bem suas características e natureza, mas entre as curvas de execuções diferentes há pequenas variações que podem ser facilmente percebidas visualmente.

Esses resultados para o *fitness* médio podem indicar que esse experimento possui os mecanismos necessários para que surja e se mantenha um sistema de comunicação

confiável nos grupos. Mas, ao mesmo tempo, também parece mostrar que esse experimento está exposto a alguns fatores que interferem de maneira significativa na estabilidade desse sistema. Para analisar melhor quais são esses fatores, vamos novamente verificar os dados apresentados pelos gráficos de relação signo-objeto, em conjunto com suas respectivas variâncias e correlações com o *fitness*, e adicionar ainda uma análise das medidas de variação genética da população.

A figura 5.23 mostra um gráfico com comportamento típico das curvas de relação signo-objeto, apresentando as curvas de relação do signo com o objeto presente e com o objeto ausente do ambiente. Esse gráfico utiliza os dados da mesma simulação que resultou no *fitness* médio apresentado na figura 5.22. Podemos perceber que a tendência das curvas apresentadas na figura 5.23 é de se afastarem cada vez mais com o passar das gerações, com a relação do signo com o objeto presente aumentando, e a relação do signo com o objeto ausente diminuindo (relativamente rápida nas primeiras gerações, e depois seguindo a tendência lentamente), mesmo com algumas oscilações. Por outro lado, a média de *fitness* apresentada na figura 5.22, apesar de sua significativa variação, mostra uma tendência de crescimento ao longo das gerações.

Mesmo com relativamente pouco aumento ou diminuição ao longo das gerações, parece existir uma boa correlação entre os índices de proximidade signo-objeto e o *fitness*. De fato, essa percepção visual se confirma, conforme mostra a figura 5.24, que, ainda utilizando a mesma execução das figuras 5.22 e 5.23, apresenta os dados de relação signo-objeto para o objeto presente e o *fitness*. O valor da correlação em aproximadamente 0.91 indica que há uma forte correlação entre essas medidas.

Essa alta correlação existe por que, mesmo com grande concentração do *fitness* médio em torno de valores relativamente altos, tal concentração é menor do que nas simulações com interação entre agentes geneticamente iguais. Um dos motivos para essa concentração ser menor é justamente a maior variação desses valores. Outro é que o crescimento inicial do *fitness* ocorre em uma taxa menor do que nas simulações com interação entre agentes geneticamente iguais. Dessa forma, seus valores ficam mais distribuídos ao longo do gráfico, possibilitando maior correlação. Devemos notar, entretanto, que apesar de praticamente todos os valores medidos nas execuções do experimento com seleção e interação por grupo possuírem valores intermediários entre as outras duas simulações vistas até aqui, o comportamento de correlação entre relação signo-objeto e *fitness* não é intermediário, mas apresenta valores bem próximos aos valores obtidos nas execuções da replicação do experimento original (por vezes os superando).

Segundo o teste estatístico de Mann-Whitney, não há diferença significativa entre o experimento original e o experimento com grupos em relação aos valores de relação signo-objeto, e de correlação entre essas medidas e o *fitness*. Mesmo assim, os valores de *fitness* do experimento com grupos é significativamente maior que os valores do experimento original (ainda de acordo com o teste de Mann-Whitney). Assim como nos experimentos com interação entre agentes geneticamente iguais, a variância das ativações da camada intermediária podem fornecer algumas respostas a esse respeito. No caso da simulação com interação e seleção por grupo, as médias das variâncias das distâncias signo-objeto é significativamente menor do que essa mesma média no experimento original, e praticamente igual a essa variância no experimento com interação entre agentes geneticamente

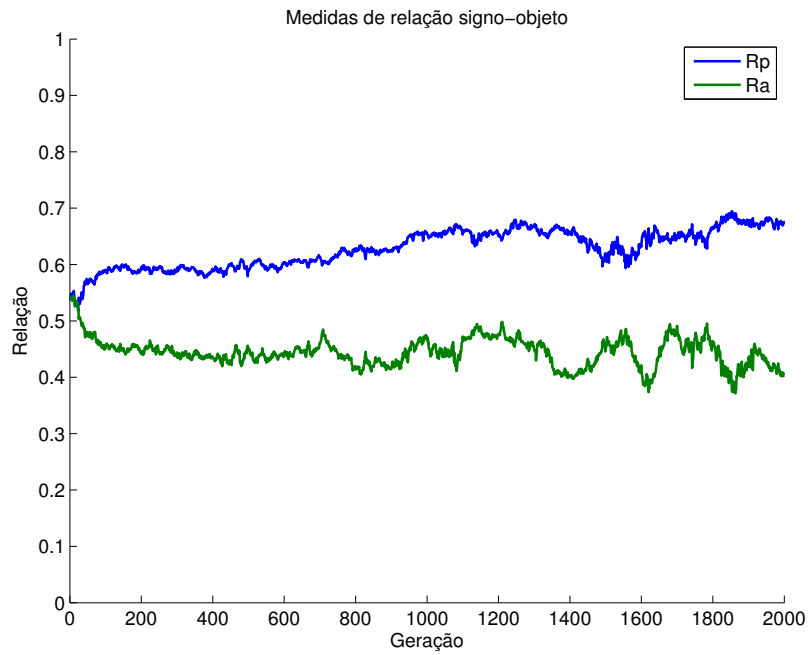


Figura 5.23 Exemplo de curva de medidas de relação signo-objeto com objeto presente (R_p) e com objeto ausente (R_a), em uma execução do experimento com grupos.

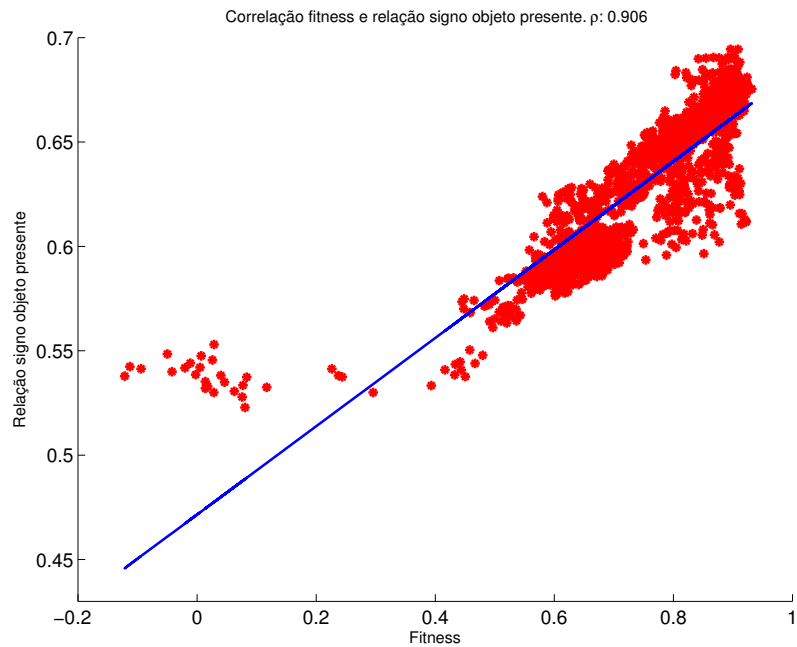


Figura 5.24 Gráfico de dispersão do *fitness* e relação signo-objeto com o objeto presente em uma execução do experimento com interação e seleção por grupo.

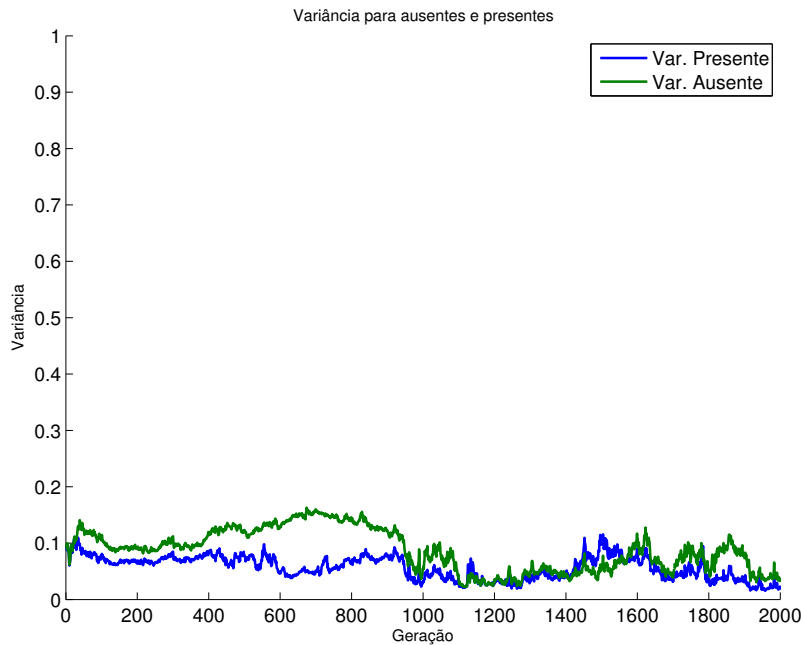


Figura 5.25 Variância das medidas de relação do signo com o objeto presente e ausente, durante uma execução da simulação com interação e seleção por grupo.

iguais, como podemos observar na figura 5.25, que apresenta a forma característica das curvas de variância das medidas de relação signo-objeto para as simulações do experimento com interação e seleção por grupo. É possível perceber que os valores tendem a sofrer pequenas oscilações ao longo das gerações, mas mantém certa tendência de queda. Um comportamento análogo aos gráficos com valores de *fitness*, com a diferença essencial de que o *fitness* mantém uma tendência ascendente.

Podemos identificar na simulação do experimento com interação e seleção por grupo, portanto, três comportamentos interessantes e generalizáveis em relação às curvas de *fitness* médio da população e às medidas de relação signo-objeto, que parecem se relacionar de algum modo:

- A curva do *fitness* médio cresce rapidamente até aproximadamente a geração 200, e depois segue uma tendência de crescimento, mas de forma bem mais suave e sujeita a pequenas oscilações (menores do que no experimento original, mas maiores do que no experimento com interação entre agentes geneticamente idênticos) (figura 5.22);
- As curvas de medidas de relação signo-objeto com o objeto presente e com o objeto ausente tendem a se afastar gradualmente ao longo das gerações. Ou seja, a relação signo-objeto com o objeto presente tende a aumentar, enquanto para o objeto ausente tende a diminuir (figura 5.23);
- As curvas de variância das ativações da camada intermediária, tendem a alcançar valores mais baixos com o passar das gerações, apesar de sofrerem muitas oscilações

(figura 5.25).

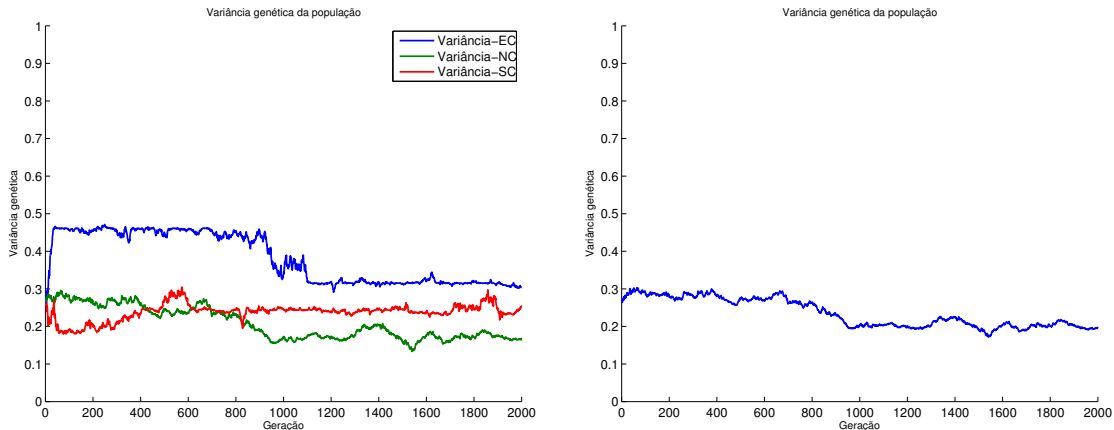
Para entendermos por que essa dinâmica acontece, devemos atentar para a forma de seleção, reprodução e interação da configuração por grupo, e para como os valores listados acima são calculados. Enquanto as interações são realizadas dentro de um grupo, e o cruzamento para a reprodução é feita entre indivíduos de grupos diferentes (conforme descrito no início desta seção), os cálculos de todos os valores realizados para essas simulações são iguais aos realizados tanto para as simulações do experimento original, quanto para o experimento com interações entre indivíduos geneticamente iguais; ou seja, os cálculos são realizados considerando-se a população como um todo, e não é feito nenhum cálculo em específico sobre os grupos.

Ao vincular a chance de sucesso de reprodução do indivíduo ao sucesso dos demais indivíduos do grupo, o mecanismo de seleção tende a selecionar para reprodução os indivíduos dos grupos nos quais houve maior cooperação entre os agentes. Essa cooperação seria basicamente a produção, pelo falante, de um signo que fizesse o ouvinte tomar a decisão de maior ganho de *fitness*, de acordo com o cogumelo presente no ambiente (ficar parado para cogumelos venenosos, e andar para cogumelos comestíveis), em conjunto com a interpretação adequada do ouvinte. Para que isso aconteça, é necessário que os agentes falantes sempre sinalizem de forma coerente. Dessa maneira o *fitness* de cada indivíduo do grupo aumenta, aumentando também as chances de reprodução de seus indivíduos. Entretanto, a reprodução não ocorre entre indivíduos do mesmo grupo, mas necessariamente entre indivíduos de grupos diferentes, e isso pode acarretar algumas situações que vão de encontro à cooperação entre os indivíduos. Por exemplo, é possível que signos diferentes sejam produzidos em cada grupo. Se dois grupos que utilizam signos bastante diferentes entre si são selecionados para reproduzir seus indivíduos, provavelmente os indivíduos gerados a partir do cruzamento entre esses indivíduos desses grupos não interpretarão signos tão bem quanto seus pais em seus respectivos grupos. Além disso, depois que os novos indivíduos são gerados, estes são distribuídos em grupos de maneira aleatória, de forma que não há como garantir que descendentes de grupos semelhantes estarão em mesmos grupos na geração seguinte.

Consideremos novamente a questão da reprodução, que pode misturar as características dos indivíduos de cada grupo de forma a deteriorar o repertório de signos utilizados. Para que uma nova geração não dependa do grupo de origem nem do grupo de destino dos seus indivíduos para que sempre se mantenha um sistema de comunicação confiável funcionando em todos os grupos, é necessário que toda a população compartilhe o mesmo repertório. Uma das premissas para que isso aconteça é que haja um processo de convergência genética da população que, por conta do processo de seleção e reprodução descritos acima, pode ser relativamente lento.

A figura 5.26 apresenta dados sobre a variância genética da população durante uma simulação do experimento com grupos, e mostra um exemplo do processo de convergência genética da população ao longo das gerações. A figura 5.26(a) apresenta separadamente as variâncias de três grupos de pesos:

- **Variância-EC:** variância dos pesos da entrada comunicativa;



(a) Variância genética da população dividida por grupos de pesos relacionados à funções específicas de comunicação dos agentes: Entrada Comunicativa (EC); Saída comunicativa (SC); e pesos Não Comunicativos (NC). (b) Variância genética geral da população, ponderando todos os pesos.

Figura 5.26 Gráficos que mostram a tendência de queda na variância genética da população.

- **Variância-SC:** variância dos pesos da saída comunicativa;
- **Variância-NC:** variância dos pesos não comunicativos (entrada visual e saída motora).

Podemos perceber que a variância dos pesos da saída comunicativa se manteve praticamente constante durante toda a simulação; já a variância dos pesos da entrada comunicativa sofreu uma queda considerável em torno da geração 1000, se mantendo praticamente estável no restante da simulação; e os pesos não comunicativos, com algumas pequenas oscilações, manteve uma tendência de queda durante toda a simulação. Uma observação importante é que tanto a entrada quanto a saída comunicativa possuem quatro pesos em seus respectivos grupos, enquanto existem 22 pesos não relacionados diretamente à comunicação. Por isso as oscilações da variância nos dois primeiros grupos tendem a ser menos suaves que as do último. Porém, apesar de possuir menos pesos, os grupos de pesos comunicativos são os que mais influenciam no comportamento e, conseqüentemente no *fitness* da população. Em 5.26(b) é apresentado o gráfico com a média geral da variância, que é uma média ponderada dos três valores apresentados em 5.26(a), que apresenta também a tendência de queda ao longo das gerações.

Esses fatores parecem descrever de forma bastante apropriada o comportamento das três curvas analisadas. Há sempre uma pressão para que um sistema de comunicação confiável se estabeleça entre os grupos. Inicialmente isso ocorre relativamente rápido, visto que nesse período os agentes estão, através de contato visual, aprendendo a categorizar os objetos (cogumelos). Esse processo continua, porém mais lentamente que nas outras duas configurações de experimento, provavelmente devido à combinação de grupos diferentes para formar indivíduos que serão distribuídos em grupos aleatoriamente. Por

conta da reprodução e distribuição, também ocorrem oscilações esporádicas na produção e interpretação de signos, que se refletem no *fitness*. Mas a pressão para que todos os agentes compartilhem um repertório de signos comum, para que, independente do grupo de origem e do grupo de destino de cada indivíduo, possam cooperar na tarefa a ser realizada, faz com que a tendência seja de maior compartilhamento do sistema de comunicação, que aos poucos vai se tornando comum para todos os indivíduos. Com isso, ao longo das gerações, apesar das oscilações, a relação signo-objeto, as variâncias das ativações da camada intermediária e o *fitness* médio da população tendem a melhorar de maneira constante e relativamente lenta.

A dinâmica dessas simulações, assim como na simulação com interação entre agentes geneticamente similares, também favorece o aparecimento de boa produção e interpretação de signos, facilitando assim o surgimento de um sistema de comunicação confiável e duradouro na população, ainda que de forma mais lenta que nas simulações com interação entre indivíduos geneticamente iguais. Esse favorecimento ocorre devido a pressão que os indivíduos têm para colaborar. Por outro lado, a forma de seleção, reprodução, e redistribuição dos grupos, impede que esse mecanismo que favorece o bom processo de representação produza resultado de forma rápida e sem turbulências, só sendo possível alcançar valores de *fitness* muito altos após várias gerações.

5.4 ARQUITETURA SEM REPRESENTAÇÃO

A última modificação realizada no experimento foi de natureza arquitetural. Chamaremos a arquitetura do experimento original de arquitetura original, e a arquitetura proposta no experimento modificado de arquitetura modificada. O que pretendemos com a arquitetura modificada é separar influências visuais e auditivas, de forma que tanto a saída comunicativa quanto a saída motora sejam influenciadas apenas pelos fatores que têm ligação direta com cada uma, e assim separar a produção de signos do controle motor do agente, e com isso impossibilitar o processo de representação, tal como discutimos nos capítulos 2 e 4.

De acordo com o descrito na seção 4.4.1 e apresentado na figura 4.11, a arquitetura modificada possui a camada intermediária duplicada, com dois grupos de dois neurônios, que chamaremos de camada intermediária 1 e camada intermediária 2. A camada intermediária 1 só recebe ligações da camada de entrada visual, e por sua vez se conecta tanto à saída motora quanto à saída comunicativa. Já a camada intermediária 2 recebe ligações apenas da entrada comunicativa, e se conecta apenas à saída motora.

Com a arquitetura modificada, temos um grupo de neurônios intermediários para tratar cada conjunto de entradas (visuais ou comunicativas). Esses grupos, por sua vez, se conectam apenas com as saídas que devem ser influenciadas por determinado conjunto de entrada. Dessa forma, separamos as influências sofridas pelos agentes em cada um dos dois papéis que podem assumir. Quando está no papel de falante, ele visualiza o cogumelo, e os dados perceptivos irão para o grupo da camada intermediária 1, que se conecta tanto com a saída motora como com a saída comunicativa, uma vez que, nesse papel, a entrada visual é utilizada pelo agente para gerar a saída comunicativa. Quando está no papel de ouvinte, o agente recebe uma entrada comunicativa, e pode vir a visualizar o cogumelo.

Inicialmente ele recebe sinais comunicativos do agente falante, e depois de passar pela entrada comunicativa, segue para o grupo da camada intermediária 2, que se conecta apenas com a saída motora, visto que apenas a saída motora é influenciada pela entrada comunicativa. Caso ele chegue na vizinhança do cogumelo, poderá visualizá-lo, e então esses dados poderão influenciar também na decisão de locomover-se ou não. Logo, a camada intermediária 1, que recebe os dados vindos da entrada visual, também se conecta à saída motora.

Com essa arquitetura, então, é eliminada a possibilidade de ocorrer um processo representativo tal como descrevemos, pois as ativações da camada intermediária 1 não terão nenhum tipo de relação com as ativações da camada intermediária 2. Ao receber e processar entradas comunicativas através da camada intermediária 2, o agente não será capaz de gerar ativações similares à categorização dos objetos realizada pela camada intermediária 1, que recebe as informações de entrada visual. Com isso, enquanto o possível signo passa pela camada intermediária 2, a categorização dos cogumelos, através de sua visualização, ocorre através da camada intermediária 1. Dessa forma, os agentes não serão capazes de relacionar signo e objeto através da geração de ativações das camadas intermediárias que resultem no mesmo efeito interpretante para entradas diferentes (comunicativa e visual), para o mesmo tipo de cogumelo, não conseguindo assim estabelecer nenhuma relação simbólica entre o que vê e o que ouve. A impossibilidade que mencionamos aqui diz respeito à relação simbólica tal como inferimos no capítulo 2, e utilizando a camada intermediária da rede, como discutimos no capítulo 4. Nesse cenário, também não teremos como calcular medidas como as medidas de distância signo-objeto (e, conseqüentemente suas correlações e variâncias).

Foram realizadas 35 simulações com essa configuração, e os *fitness* médios das simulações com essa modificação foram significativamente inferiores aos valores obtidos pelas simulações com o experimento original (teste de Mann-Whitney, $p = 6.8 \cdot 10^{-8}$). O valor médio do *fitness* em todas as simulações ficou em aproximadamente 0.46. A figura 5.27 apresenta um gráfico do *fitness* médio da população em uma simulação com essa modificação. Apesar dos valores não serem uniformes entre as simulações, todos apresentam as características de possuir muitas oscilações, e variarem em torno de valores relativamente baixos. De maneira geral, as médias variam em torno de valores entre 0.4 e 0.6, aproximadamente.

No experimento em sua configuração original, o valor máximo de *fitness* que pode ser obtido pelos indivíduos sem o uso de comunicação pela população é de 0.55. Esse valor é alcançado quando o agente sempre vai até a vizinhança do cogumelo (gastando energia para isso), e sempre consome cogumelos comestíveis, e não consome cogumelos venenosos. Os valores de *fitness* próximos a essa média no experimento com a arquitetura modificada, parecem mostrar que a comunicação entre os indivíduos influencia relativamente pouco em seus comportamentos.

Com essa média de *fitness*, as criaturas do experimento com arquitetura modificada parecem não ser capazes de obter nenhum benefício da comunicação, precisando sempre se deslocar até a vizinhança do cogumelo para classificá-lo e consumi-lo ou não. Quando essa classificação não é feita apropriadamente o *fitness* cai abaixo de 0.55, e em alguns momentos nos quais coincidem que a maioria da vezes que os agentes ouvintes ficam

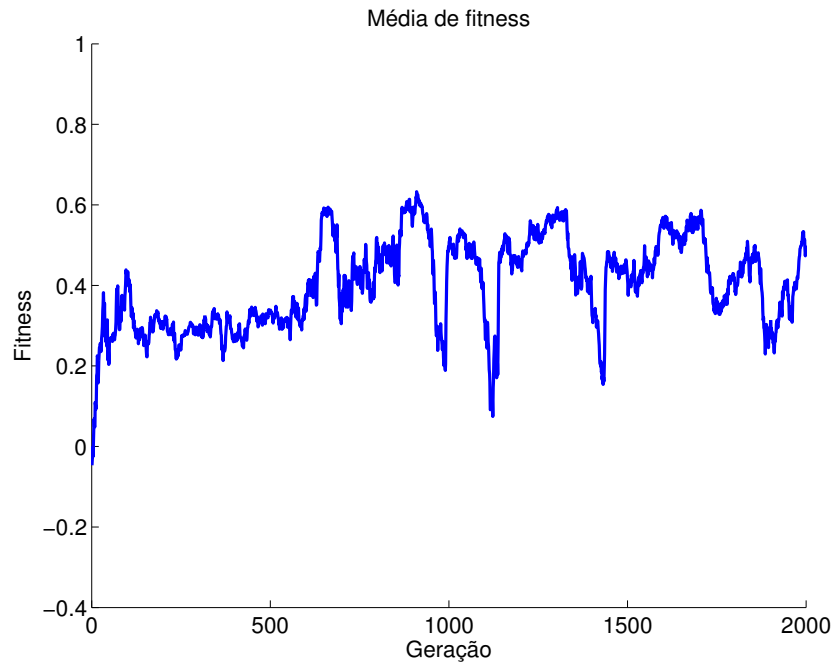
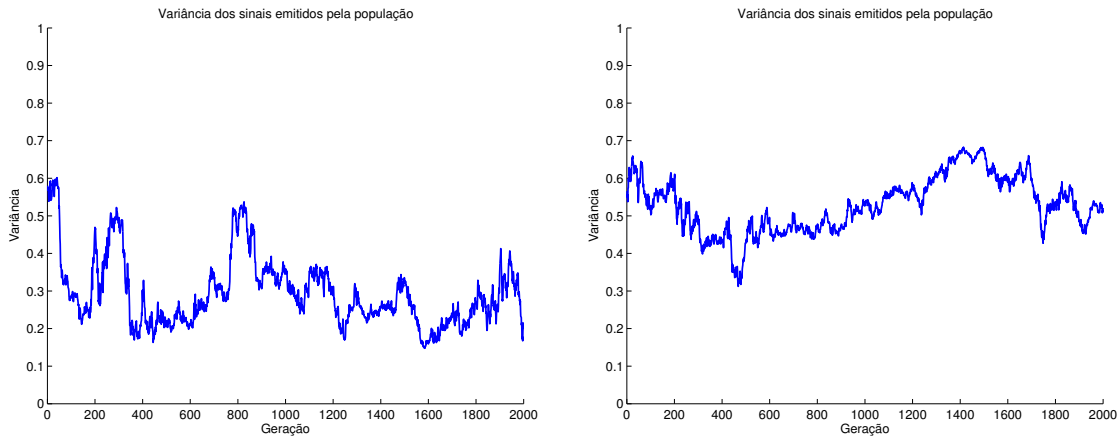


Figura 5.27 Exemplo de curva de *fitness* em uma execução de simulação com a arquitetura da rede neural modificada.

parados ao receberem alguma entrada comunicativa, e há um cogumelo venenoso no ambiente, o *fitness* chega a ficar um pouco acima de 0.55. Mas essas ocorrências são muito esparsas, e não ocorrem de forma generalizada e consistente na população, como podemos observar na figura 5.27.

Não podemos verificar ou correlacionar as medidas de relação signo-objeto, visto que essa arquitetura não permite. Porém, podemos utilizar as medidas de variância dos sinais emitidos para realizar comparações relacionadas à comunicação dos agentes. Essa medida indica o quão diferentes são os sinais produzidos pelos falantes para sinalizar o mesmo tipo de cogumelo. A figura 5.28 mostra uma comparação entre duas execuções nas duas configurações. A figura 5.28(a) mostra uma curva de variância dos sinais emitidos em uma execução do experimento original, enquanto a figura 5.28(b) mostra uma curva de variância dos sinais emitidos em uma execução com a arquitetura modificada. A variância dos sinais emitidos pelas simulações com a arquitetura modificada são significativamente maiores que a dos sinais emitidos com o experimento original (teste de Mann-Whitney, $p = 2.7 \cdot 10^{-6}$). Essa variância elevada com a arquitetura modificada (em torno de 40% maior que no experimento original), indica uma falta de condição para o surgimento de um bom sistema de comunicação, no qual há bons falantes e bons ouvintes (no sentido abordado na seção 4.2.1). Sem boa sinalização e boa interpretação, a comunicação pouco influencia na dinâmica do experimento. Note que essa situação é diferente de haver apenas agentes “enganadores”. De fato, por conta dos sinais emitidos serem tão variados (despadronizados), o que parece acontecer em alguns momentos é que os ouvintes simplesmente passam a “ignorar”, ou a sofrerem relativamente pouca influência dos sinais



(a) Exemplo de curva de variância dos sinais emitidos pelos falantes em um experimento com a configuração original. (b) Exemplo de curva de variância dos sinais emitidos pelos falantes em um experimento com a arquitetura modificada.

Figura 5.28 Variâncias dos sinais emitidos com a configuração original e com a arquitetura modificada.

comunicativos recebidos, e apresentam comportamento semelhante ao comportamento de um sistema sem comunicação, com a média do *fitness* ficando relativamente estável em valores inferiores a 0.55. Porém, estando sujeito a cruzamentos e mutações, em vários momentos podem surgir indivíduos que não ignoram os sinais recebidos, e podem começar a ser influenciados mais fortemente pelos sinais emitidos por outros agentes.

Essa modificação na arquitetura da rede neural, ao contrário das outras alterações realizadas no experimento, impossibilita o surgimento de processos representativos, impactando de maneira negativa o *fitness* médio da população nas simulações. Esse impacto negativo no *fitness* decorre justamente da falta da existência de representações, uma vez que as criaturas no papel de ouvinte não são capazes de relacionar o possível signo com seu objeto e, assim, não conseguindo interpretá-lo e conseqüentemente não conseguindo gerar a ação correta (de acordo com o cogumelo presente no ambiente) que derivaria dessa interpretação (efeito interpretante).

5.5 CONSIDERAÇÕES SOBRE OS PROCESSOS DE REPRESENTAÇÃO NAS SIMULAÇÕES

Juntando as simulações feitas com o experimento original e cada uma das variações realizadas sobre ele, ocorreram um total de 155 execuções. Com elas pudemos avaliar algumas condições que interferem no estabelecimento de um sistema de comunicação, que por sua vez influenciam no *fitness* médio da população. A concentração de esforços desse trabalho foi no sentido de avaliar a existência de processos representativos durante as simulações, analisando fatores que podem contribuir para sua ocorrência, e também avaliar qual o impacto sobre o experimento em algumas condições, caso exista algum tipo de processo representativo.

Em relação ao estabelecimento de processos representativos durante as simulações, adotamos as medidas de relação signo-objeto para quantificar e avaliar como esses processos estavam ocorrendo. Em algumas situações parece fácil afirmar que há de fato um processo de representação ocorrendo, segundo as definições de Peirce, como discutimos no capítulo 2. Por exemplo, nas simulações com interação entre agentes com similaridade genética, as medidas de relação signo-objeto, em geral, são bastante significativas (relação R_p bastante alta e relação R_a bastante baixa), assim como as respectivas variâncias (valores relativamente baixos). Isso resulta em uma interpretação dos signos praticamente sem equívocos por partes dos agentes, uma vez que a relação entre signo e objeto é plenamente estabelecida (e é característica espalhada pela população), fazendo com que os agentes consigam mapear as ativações da camada intermediária quando recebem uma entrada comunicativa, de maneira que o efeito interpretante quase sempre os leva a tomar as mesmas ações quando veem um tipo de cogumelo, e quando ouvem sobre ele, resultando em médias de *fitness* bastante elevadas.

Por outro lado, em algumas situações parece fácil afirmar que não existem processos de representação ocorrendo. Por exemplo, nas ocasiões em que os valores das medidas de relação signo-objeto com o objeto presente e com o objeto ausente estão muito próximas a zero, as respectivas variâncias são altas, e o *fitness*, com média baixa, parece variar sem muita influência de algum suposto sistema de comunicação. Nessas situações, as ativações originadas pelos supostos signos ficam bastante espalhadas, próximas da linha de decisão, e relativamente longe das ativações originadas pelas entradas visuais, correspondentes ao cogumelo presente no ambiente quando da recepção do signo (e ainda a uma distância relativamente próxima para as ativações originadas a partir das entradas visuais correspondentes ao cogumelo ausente no ambiente quando da recepção do signo). Dessa forma, os agentes não conseguem relacionar signo e objeto apropriadamente, a população não consegue estabelecer um repertório coerente de signos, e o *fitness* é afetado negativamente, ficando com valores de intermediários a baixos.

Entretanto, de maneira geral durante as simulações, principalmente do experimento original, as situações não são assim tão bem definidas, e os valores dessas medidas flutuam de forma intermediária entre esses dois cenários apresentados. Durante a análise dos dados da figura 5.11 e da tabela 5.1 pudemos ter uma percepção de que as medidas, principalmente as de relação signo-objeto e respectivas variâncias, se relacionam de maneira, até certo ponto, complementares, mas de forma não linear, e até certos limites de saturação. Por exemplo, se a variância da medida da diferença $R_p - R_a$ for muito grande, por maior que seja a medida média dessa diferença ($R_p - R_a$), o processo representativo será prejudicado. Logo, qualquer critério que estabelecêssemos deveria levar em conta essa relação. Porém, como dissemos, essa relação não é linear e obedece a certos limites extremos, o que dificulta (ou impossibilita) o estabelecimento de uma regra que possa ser adotada independentemente do contexto analisado. Por isso, achamos mais coerente e prudente fazer menção à qualidade do processo representativo, através dessas medidas. As informações fornecidas por essas medidas podem ser um forte indício da presença ou ausência de processos de representação, porém, a existência de fato de um processo representativo deve ser analisada em cada situação específica.

A boa correlação observada entre as medidas de relação signo-objeto e o *fitness* ocorreu

na maioria das execuções do experimento original e do experimento com grupos. De maneira geral, essa correlação não alcança valores significativos quando a medida de média da diferença $R_p - R_a$ é muito próxima a zero, ou a variância é relativamente grande. De fato, quando a média da diferença $R_p - R_a$ está próxima a zero, significa que boa parte das ativações da camada intermediária originadas a partir da entrada comunicativa está próxima da fronteira de decisão¹, o que aumentam as chances do agente realizar interpretações equivocadas. Variações pequenas dessa medida não farão muita diferença na taxa de interpretações equivocadas dos agentes, a menos que a variância seja extremamente baixa. Por outro lado, se a variância das relações signo-objeto é muito grande, a oscilação da média não é um dado tão significativo, e não podemos inferir como essa oscilação irá afetar o *fitness*.

Com exceção desses cenários, os valores significativos de correlação que ocorreram apontam para a existência de uma associação direta entre as medidas de qualidade do processo de representação e o *fitness*. Ou seja, aparentemente, a variação dos valores das medidas de relação signo-objeto, assim como de suas respectivas variâncias, produzem, com alguma proporção, efeitos sobre o *fitness*. Como argumentamos inicialmente, o simples fato dessas medidas covariarem não implica que uma variável é necessariamente dependente da outra (que esta varia em função daquela). Nesse caso, a hipótese seria que o *fitness* varia em função das medidas de qualidade de representação. Após essa argumentação, realizamos uma série de análises e discussões a fim de gerar mais subsídios e poder reavaliar a hipótese, procurando validá-la ou refutá-la. Essas análises e discussões em nenhum momento refutaram a hipótese, mas, pelo contrário, utilizando-a nas argumentações, as explicações se encaixam de forma coerente e consistente. Assim, mesmo que não tenhamos confirmado de maneira categórica, as evidências apontam fortemente para a validade dessa hipótese, na qual as nossas medidas de qualidade dos processos de representação impactam diretamente no *fitness* médio da população.

Os resultados obtidos pela simulação com a arquitetura modificada também corroboram a hipótese levantada. Nessa arquitetura que não permite a ocorrência de processos representativos, a média do *fitness* foi significativamente mais baixa do que em todas as outras simulações. Além disso, em alguns momentos de algumas simulações, os valores dos atributos avaliados pareciam indicar que os sinais emitidos pelos falantes estavam, se não sendo completamente ignorados, tendo pouca influência na tomada de decisão dos agentes ouvintes.

Em alguns momentos durante as execuções das simulações (e, como discutimos anteriormente, em algumas simulações mais do que em outras), as criaturas conseguem estabelecer uma relação entre signo e objeto de forma apropriada, conseguindo tomar as

¹Para que a diferença $R_p - R_a$ esteja próxima de zero, é necessário que a distância média das ativações da camada intermediária para entrada comunicativa esteja a distâncias similares das ativações originadas por entradas visuais tanto para o tipo de cogumelo presente no ambiente quanto para o tipo de cogumelo ausente. É característica das simulações (como podemos perceber em todos os gráficos que mostram as ativações da camada intermediária) que as ativações originadas a partir de entradas visuais se concentram ao longo de um extremo do plano cartesiano, em lados opostos para cada tipo de cogumelo. A fronteira de decisão passa necessariamente pela origem do plano. Logo, para que a distância das ativações originadas a partir de entradas comunicativas, tenha distância similar para ambos os grupos de ativações originadas a partir de entrada visual, estas devem estar concentradas em algum ponto próximo a fronteira de decisão.

ações de forma adequada em relação ao tipo de cogumelo presente no ambiente. Verificamos que essa associação é realizada, pelo menos em parte, através da aproximação das ativações da camada intermediária de quando a criatura ouve sobre determinado tipo de objeto, das ativações geradas nessa mesma camada, de quando a criatura vê esse mesmo tipo de objeto. Há dessa forma, ao menos nesses momentos, um processo de interpretação por parte da criatura, no qual ao ouvir sobre determinado tipo de objeto, suas ativações são tais que a ação tomada (o efeito gerado) tenderá a ser a mesma que tomaria se estivesse visualizando aquele mesmo tipo de cogumelo. Ou seja, a criatura consegue relacionar a entrada comunicativa com a entrada visual, de forma a tomar a mesma ação para tipos iguais de cogumelo (e diferentes para tipos diferentes de cogumelos). O intérprete consegue estabelecer relação entre signo e objeto, gerando um efeito interpretante adequado. O signo em questão, por sua vez, representa seu objeto em algum aspecto, nesse caso, o tipo do cogumelo, o que representa categorias, não um objeto de forma individual. Além disso ele não possui semelhança, nem relação espaço-temporal ou física com seu objeto. Portanto, o signo existente nesses momentos onde há processos representativos, de fato, só pode ser símbolo.

CONCLUSÃO

Representação (ou processo de representação) é um tema abordado na área da IA desde seu início. Porém, apesar de terem surgido várias abordagens diferentes ao longo do tempo, que tentaram lidar com representações em sistemas computacionais de forma apropriada, esse campo permanece em aberto, com muitas perguntas a serem ainda respondidas. Até mesmo a definição do termo representação, dentro da IA e Ciências Cognitivas, não é algo consolidado e amplamente aceito, existindo ainda muita confusão em sua utilização. Por outro lado, muitos experimentos sobre emergência de comunicação em agentes artificiais vêm sendo realizados a cada ano. A grande maioria desses trabalhos não demonstram interesse em possíveis processos de representação que possam existir na comunicação estabelecida pelos agentes, e assim não discutem o tema.

Entendemos que a inclusão da análise dos processos representativos nesses experimentos podem ser bastante úteis, e contribuir para avanços nas pesquisas dessa área. Nessas circunstâncias, o objetivo principal desse trabalho foi estudar e analisar possíveis processos representativos que possam ter ocorrido em um experimento já realizado, esperando assim contribuir com discussões, resultados complementares da análise do experimento estudado, e no desenvolvimento de uma metodologia que possa ser utilizada por outros pesquisadores que pretendam realizar algum tipo de análise sobre processos de representação em seus experimentos.

Com as análises realizadas no experimento de Mirolli e Parisi (2008), foi possível identificar a ocorrência de processos representativos (do tipo simbólico) no experimento replicado. Como comentamos na seção 5.5, a análise e a própria identificação dos processos representativos são facilitadas ou dificultadas tanto entre simulações, quanto em momentos distintos dentro da mesma simulação. Entretanto, com a metodologia que utilizamos, conseguimos coletar várias evidências de que, ao conseguir gerar ativações semelhantes na camada intermediária de sua rede neural para o mesmo tipo de cogumelo, e ativações diferentes para tipos de cogumelo diferentes, os agentes, ao receberem entrada comunicativa tendem a agir da mesma forma que agem quando recebem a entrada visual do tipo de cogumelo presente no ambiente. Ou seja, há uma interpretação por parte do agente, que relaciona signo e objeto, gerando um efeito pragmático, o que

caracteriza a tríade semiótica de Peirce, discutida na seção 2.6. A adequação desse efeito pragmático pôde ser medida, pelo menos em parte, pela correlação entre as medidas de relação signo-objeto e o *fitness*, e pela variância dessas medidas. Além disso, a vantagem adaptativa da utilização representação ficou evidente nas simulações realizadas, uma vez que as execuções da simulação com a arquitetura modificada para não permitir o uso de signos tiveram *fitness* médio significativamente menor que todas as outras simulações, que permitiam o uso de representações.

Para chegar a essas conclusões, iniciamos o trabalho com uma discussão sobre representação dentro do contexto da Ciência da Computação e da IA, a fim de deixar claro o que estávamos procurando. Apresentamos e discutimos três abordagens da IA, cuja diferença central entre elas passa pelo entendimento do que seja e de como pode ser utilizada a representação. A IA Clássica (ou IA Simbólica) defende que um sistema inteligente seja aquele que possa armazenar e manipular representações através de um algoritmo; para essa abordagem, representações são padrões físicos, como um conjunto de bits de uma palavra na memória de um computador. O conexionismo defende uma ideia de representação distribuída, na qual as ativações dos nós de uma rede, a força das conexões entre os nós, e os processamentos por eles realizados formariam padrões aos quais se poderia atribuir valor semântico. Por realizar processamentos em unidades que só poderiam possuir valor semântico se consideradas em conjunto, essa abordagem ficou também conhecida como abordagem sub-simbólica. Já a abordagem na Nova IA, refuta o uso da representação no sentido utilizado pela IA Clássica, e lida com aspectos até então ignorados pelas outras duas abordagens, como a incorporação e a situacionalidade de agentes robóticos, e acoplamento entre percepção e ação. Dessa forma, as abstrações que poderiam existir em sistemas inteligentes, segundo essa abordagem, seriam representações de informações fundamentadas no sensoriamento do agente, ao longo de sua história.

Apesar dessa última abordagem (e algumas de suas derivações) ter alcançado relativa popularidade, ainda há muita discussão sobre qual a abordagem mais apropriada, bastante discordância sobre a utilização do termo representação, e muitas questões em aberto. Dessas questões, podemos destacar o Problema da Fundamentação do Símbolo, que questiona como sistemas computacionais que processam símbolos baseados apenas em regras de manipulação, podem atribuir significado intrínseco a tais estruturas. Mesmo que aparentem agir como se estivesse entendendo determinadas estruturas simbólicas, na verdade estariam apenas seguindo as regras sintáticas de manipulação de tais estruturas, e para o próprio sistema computacional, o símbolo manipulado não possui nenhum vínculo com aquilo a que se refere, mas apenas observadores externos atribuem significado aos seus símbolos e comportamentos.

Levantamos mais discussões sobre o uso do termo, abordando desde defesas de algumas definições, a pesquisadores que preferem se abster dela e não entrar na discussão, passando por aqueles que preferem não o utilizar, temerosos de que seu uso atrapalhe mais do que ajude no entendimento de seus trabalhos. Por outro lado, há autores que indicam que a interação entre disciplinas com interesses em comum com representações pode ajudar a superar essa barreira, e apontam particularmente para a semiótica, campo que se dedica ao estudo dos signos. De fato, alguns autores na área da emergência da

comunicação já utilizaram a teoria semiótica de Peirce para fundamentar algum aspecto de seus trabalhos, especialmente relacionados à comunicação entre agentes artificiais. Dessa forma, utilizamos a semiótica de Peirce como *framework* teórico para estudar o fenômeno da representação em sistemas cognitivos, e assim fornecermos uma definição de representação robusta e genérica o suficiente para ser entendida e utilizada por todos os pesquisadores que têm interesse no tema, além de delimitar de maneira clara o que estávamos nos referindo quando falávamos em busca por representação.

Além disso, a simulação computacional vem se consolidando como uma importante ferramenta nas pesquisas sobre linguagem e comunicação, podendo contribuir em vários aspectos nessas pesquisas. Dentre essas simulações, encontram-se as simulações sobre emergência de comunicação. Porém, apesar do crescimento de trabalhos que utilizam experimentos sobre emergência de comunicação, e das possibilidades que eles oferecem, pouquíssimos trabalhos trazem alguma discussão sobre as representações possivelmente existentes nos experimentos, apesar de, na comunicação, sempre haver algo produzido e transmitido por um agente, que deve ser recebido e interpretado de alguma forma por outro agente.

Após a escolha de um experimento para ser replicado, e baseados na semiótica de Peirce, mapeamos os elementos de um processo de representação dentro do experimento replicado. Esse mapeamento tomou também como ponto de partida a ideia de que, se existe algum processo de representação, ele necessariamente passa pela camada intermediária da rede neural. A análise das ativações da camada intermediária de uma rede neural já havia sido utilizada como metodologia para a busca por representação em trabalho recente, porém, além das análises realizadas serem diferentes, também eram os propósitos dos trabalhos, uma vez que o experimento analisado em Mirolli (2012) não era relacionado com emergência de comunicação, e a definição de representação adotada é bastante diferente da que adotamos no presente trabalho. Essa análise consistiu essencialmente em avaliar as distâncias entre as ativações da camada intermediária de quando os agentes ouviam sobre determinado tipo de cogumelo, das ativações que ocorriam quanto as criaturas viam o mesmo tipo de cogumelo, e também para as ativações geradas quando os agentes viam o outro tipo de cogumelo. Dessa maneira conseguimos verificar se o agente consegue relacionar signo (sinal comunicativo recebido) ao que ele se refere, seu objeto (tipo de cogumelo, categorizado pelas entradas visuais), gerando um efeito interpretante adequado. Essa é a relação mais fundamental do signo, segundo a teoria de Peirce. Além das análises realizadas no experimento original, propomos modificações que nos permitiram verificar mais apropriadamente elementos que favorecem ou dificultam o surgimento de processos representativos, além de poder avaliar melhor se, e de que forma a emergência de representações influencia na capacidade de interpretação dos agentes ouvintes, e conseqüentemente, no *fitness* médio da população de criaturas artificiais.

As execuções da réplica do experimento original forneceram dados que nos permitiram mensurar covariância entre a relação signo-objeto e o *fitness*, sustentando a hipótese de que a existência de processos de representação podem contribuir para a melhor comunicação dentro de uma população de agentes artificiais. Nos casos em que essa covariância era insignificante, pudemos observar que com os dados relativos à variância das medidas

de relação signo-objeto, conseguimos explicar de forma complementar o comportamento das criaturas. Os experimentos modificados permitiram verificar e confirmar os indícios sobre quais características poderiam contribuir para uma melhora nos processos de representação, e quais impediriam completamente tal processo. Nos experimentos em que o processo de representação foi facilitado (interação entre agentes geneticamente similares, e interação e seleção por grupo), os agentes ouvintes conseguiam estabelecer relação entre signo e objeto mais facilmente e com menor risco de equívoco, aumentando assim o *fitness* da população, que alcançou valores significativamente mais altos que no experimento original. Enquanto no experimento em que esse processo foi impossibilitado (com arquitetura modificada), os agentes não conseguiram estabelecer relação entre signo e objeto, passando praticamente a ignorar os sinais recebidos em alguns momentos das execuções dessa simulação, o que causou queda nos valores de *fitness* médio dessas populações, os quais foram significativamente menores que a média de *fitness* do experimento original.

6.1 CONTRIBUIÇÕES

Como já mencionamos, mesmo em pouquíssimo número, alguns trabalhos que tratam de emergência da comunicação já abordaram a temática da representação em seus experimentos. Porém, até onde sabemos, esse é o primeiro trabalho que realiza essa análise em um experimento realizado previamente, de forma a complementar e avançar, e não sobre um novo experimento proposto no mesmo trabalho. Há dois principais fatores metodológicos nesse trabalho que podem contribuir tanto para outros trabalhos com propósitos semelhantes a esse, quanto para complementar análises de novos experimentos: a utilização da teoria semiótica de Peirce, e a análise e a busca por representações baseadas principalmente em comparações de ativações na camada intermediária da rede neural.

Alguns trabalhos relacionados a experimentos sobre emergência de comunicação e/ou sobre a fundamentação do símbolo já utilizaram, de algum modo, a teoria semiótica de Peirce. Porém, pudemos perceber que as notas sobre a teoria peirciana eram sempre pontuais e sem muita relação com o experimento propriamente dito, e/ou distorciam a teoria (inclusive atribuindo outros nomes e conceitos aos elementos da tríade semiótica). Aqui procuramos abordar a teoria semiótica de Peirce não de forma exaustiva, pois não era necessário, mas com todo o rigor necessário para não distorcê-la, e aplicá-la não de forma pontual, mas de modo que ela esteja diretamente relacionada com os resultados que encontramos. Dessa forma pretendemos ajudar a relacionar ainda mais a teoria de Peirce ao contexto dos experimentos em emergência de comunicação. A aproximação da semiótica dos estudos da IA e das ciências cognitivas, inclusive, foi apontado como um passo importante para o avanço no conhecimento de disciplinas que lidam com representações (ZIEMKE; SHARKEY, 2001).

Tendo em vista o grande potencial de gerar confusão que a utilização do termo representação possui, é primordial que se deixe explícito o que se quer dizer com esse termo, uma vez que ele pode ser interpretado de muitas maneiras. A utilização da semiótica de Peirce foi de fundamental importância não só para essa definição, mas ela oferece também uma teoria genérica e robusta, que pode ser utilizada em várias áreas de co-

nhcimento, sem nenhuma restrição. Por exemplo, não há restrições sobre quais tipos de criaturas poderiam utilizar signos, ou sobre algum tipo de subtrato neural necessário para que o intérprete pudesse sofrer o efeito interpretante. Por outro lado, faz restrições sobre fenômenos de simples covariação entre variáveis quaisquer, como temperatura e dilatação de um sensor, uma vez que essa relação não pode ser descrita pela tríade semiótica estabelecida por Peirce.

Até por conta da definição de representação (ou processo representativo) baseada na teoria semiótica de Peirce (algo bastante diferente da definição de representação utilizada por Mirolli e Parisi (2008), que avaliaram o sistema de comunicação apenas da perspectiva dos falantes), aqui analisamos o ouvinte como intérprete dos signos, oferecendo assim uma perspectiva diferente e complementar, uma vez que é na situação de ouvinte que as criaturas realizam a interpretação dos possíveis signos que chegam até elas.

A análise do funcionamento interno de uma rede neural pode não ser uma tarefa trivial, e nesse trabalho nos propusemos a avaliar vários dados relacionados às ativações da camada interna da rede neural, na procura por processos de representação. Coletamos dados para comparação entre as ativações da camada intermediária a partir de entradas diferentes, e realizamos então os cálculos e testes necessários para as análises pretendidas. Como citamos anteriormente, uma metodologia parecida com essa já foi utilizada, mas não no contexto de experimento de emergência de comunicação, e procurando por representações da maneira como a definimos.

Nos testes e cálculos que realizamos nesse trabalho, procuramos sempre utilizar métodos estatísticos apropriados à situação, como por exemplo, para saber se uma série de variáveis era significativamente maior (ou menor), que outra, ou para saber se duas variáveis sofriram modificações semelhantes ao longo das simulações (covariavam). O que poderia ser uma prática comum entre pesquisadores que lidam quase sempre com esse tipo de análise, acabou se tornando um diferencial desse trabalho, uma vez que de todos os trabalhos sobre emergência de comunicação que revisamos, apenas um apresentou os resultados obtidos com testes estatísticos associados (Floreano et al, 2007). Testes estatísticos agregam mais confiabilidade aos resultados apresentados, pois a utilização de ferramentas matemáticas adequadas podem ajudar a prevenir erros de avaliação. Especificamente no domínio das simulações sobre emergência de comunicação, ajuda a prevenir principalmente conclusões precipitadas sobre melhor rendimento de um determinado método sobre outro.

Com esses passos (definição de representação e análise detalhada da camada interna de rede neural) foi possível realizar uma análise bastante consistente do que ocorre durante as simulações, no que diz respeito aos processos de representação. Pretendemos assim oferecer à comunidade uma espécie de *framework* de trabalho que possa ser facilmente reutilizado, tomado como base e ajustado para outros experimentos. Dessa maneira espera-se ajudar no aumento do número de trabalhos que realizam análises sobre os processos de representação existentes, contribuindo assim para mais avanços nas pesquisas dessa área.

As alterações realizadas sobre o experimento original, nos permitiram confirmar a atuação de alguns fatores sobre a dinâmica do surgimento e desaparecimento do sistema de comunicação confiável da população, e a relacionar esse ciclo com a existência dos processos de representação envolvidos. O experimento com interação entre agentes gene-

ticamente similares mostrou a influência de fatores como similaridade genética e grau de parentesco entre as criaturas que interagem na transmissão de informações “honestas” entre elas, mesmo acarretando custo para isso. Nessas simulações, as medidas de relação signo-objeto se mostram significativamente melhores que em todas as outras simulações, se mantendo em níveis relativamente constantes durante todas as gerações da simulação, e não há o ciclo de aparecimento e destruição do sistema de comunicação. Isso facilita o processo de interpretação das criaturas, e se reflete diretamente na estabilidade e no alto valor do *fitness* médio da população nessa configuração do experimento. O experimento com grupos verifica a tendência à cooperação entre membros de um mesmo grupo, devido à pressão seletiva exercida pela seleção para reprodução por grupos, e não por indivíduos. Esse experimento mostrou que esse mecanismo, apesar de apresentar resultados menos expressivos que o experimento com interação entre agentes geneticamente iguais, tem influência bastante relevante nas simulações, as quais tiveram resultados relativos às medidas de relação signo-objeto e *fitness* significativamente maiores que o experimento original. As medidas citadas sobem mais lentamente, e possuem alguma instabilidade, mas em geral mantém uma tendência de alta, provavelmente proporcionada pela lenta convergência genética que vai ocorrendo na população. O experimento com arquitetura modificada procura mostrar os resultados de uma simulação com criaturas com arquitetura da rede neural sem a possibilidade de ocorrência de processo de representação, tal como definimos. Percebemos que, sem possibilidade de associar signo e objeto, e assim gerar o efeito interpretante adequado, as criaturas tomaram ações equivocadas com maior frequência nessa configuração, o que contribuiu para que o *fitness* das simulações com a arquitetura modificada fossem significativamente menores que todas as outras simulações, mostrando a influência negativa da falta de representação para a comunicação e, conseqüentemente, para o *fitness*.

Realizamos ainda uma revisão histórica (mas não exaustiva), relacionando a utilização e as principais diferenças conceituais e terminológicas em relação ao termo representação, entre algumas das principais abordagens da IA. Nesse levantamento, de maneira secundária, notamos ainda que outras definições no campo da IA, como conexionismo e computacionalismo, também são motivos de algumas controvérsias e confusões entre pesquisadores. Acreditamos que mais discussões sobre esses temas são necessárias para que se possa ter mais esclarecimento sobre esses assuntos, e a comunidade possa convergir no entendimento das terminologias empregadas em suas pesquisas. Acreditamos que uma das contribuições desse trabalho é justamente a discussão feita sobre representação e sobre algumas abordagens da IA, e esperamos também fomentar mais discussões como essa na comunidade.

6.2 TRABALHOS FUTUROS

Nesse trabalho foi analisada uma rede neural do tipo *feedforward*, na qual a informação flui sempre em uma única direção. Apesar de ser um tipo de rede neural artificial bastante comum, não representa a totalidade das simulações sobre emergência de comunicação em agentes artificiais. Por isso, pretendemos realizar trabalho semelhante com experimento que utilize redes neurais recorrentes, tanto para avançar na representatividade da me-

todologia, como para auxiliar no entendimento sobre processos representativos nesses experimentos, avaliando outras arquiteturas de redes neurais. Para isso, devemos desenvolver o método utilizado, para que abranja esse tipo de rede neural, também comum no tipo de experimento avaliado.

É necessário ainda uma reavaliação das análises realizadas. Acreditamos que é possível tanto encontrar novas medidas que quantifiquem novas informações, quanto melhorar algumas medidas que já fizemos. Tivemos alguma dificuldade, por exemplo, em, a partir das medidas calculadas nas simulações, definir categoricamente se existe processo de representação ou não. Esse é um ponto a ser mais discutido e amadurecido, tanto do ponto de vista teórico, com aprofundamento das discussões sobre a semiótica de Peirce, quanto do ponto de vista de revisão, melhoramento, e proposição de novas medidas, que possam refletir mais precisamente o que definimos por representação.

Dentro dessa mesma seara, há ainda a discussão, na qual podemos avançar, de como/quando as criaturas passam de um estado de não utilização de representações, para um estado de uso. Quais são as modificações cognitivas que essa passagem implica ou que é necessária, e como conseguir identificar tanto a passagem quanto as modificações ocorridas nas criaturas, que resultaram nessa mudança de comportamento em relação ao signo.

Aqui temos resultados iniciais de como a utilização de representação pode influenciar na comunicação e sucesso adaptativo dos indivíduos, do ponto de vista evolutivo. Por outro lado, a maior parte (se não todos) os fenômenos naturais e/ou com possíveis aplicações tecnológicas (incluindo o estudo da linguagem e comunicação humana, um dos campos de estudo que pode ser influenciado por esse trabalho), apresentam modelos mais complexos do que o modelo apresentado no experimento replicado aqui. Avaliar como o tipo de análise realizada aqui pode ser aplicada em modelos mais complexos pode representar muitos avanços para essa área de pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGMON, E.; BEER, R. D. The evolution and analysis of action switching in embodied agents. *Adaptive Behavior*, SAGE Publications, p. 3–20, 2013.
- AGRE, P. E.; CHAPMAN, D. Pengi: An implementation of a theory of activity. In: *AAAI*. [S.l.: s.n.], 1987. v. 87, n. 4, p. 286–272.
- AMPATZIS, C. et al. Evolution of signaling in a multi-robot system: Categorization and communication. *Adaptive Behavior*, SAGE Publications, v. 16, n. 1, p. 5–26, 2008.
- ANDERSEN, P. B. *A theory of computer semiotics: Semiotic approaches to construction and assessment of computer systems*. [S.l.]: Cambridge University Press, 1997.
- BEBIS, G.; GEORGIOPOULOS, M. Feed-forward neural networks. *Potentials*, IEEE, v. 13, n. 4, p. 27–31, 1994.
- BEDAU, M. A. *Philosophical content and method of artificial life*. [S.l.]: Citeseer, 1998.
- BEER, R. D. The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, Sage Publications, v. 11, n. 4, p. 209–243, 2003.
- BEER, R. D. Dynamical systems and embedded cognition. *The Cambridge handbook of artificial intelligence*, Cambridge University Press, p. 128, 2014.
- BEER, R. D.; WILLIAMS, P. L. Information processing and dynamics in minimally cognitive agents. *Cognitive science*, Wiley Online Library, 2014.
- BICKHARD, M. H.; TERVEEN, L. *Foundational issues in artificial intelligence and cognitive science: Impasse and solution*. [S.l.]: Elsevier, 1996.
- BOBROW, D. G. *Natural language input for a computer problem solving system*. Tese (Doutorado) — Massachusetts Institute of Technology, Cambridge, 1964.
- BROOKS, R. New approaches to robotics. *Science*, v. 253, n. 5025, p. 1227–1232, 1991a.
- BROOKS, R. A. Elephants don't play chess. *Robotics and autonomous systems*, Elsevier, v. 6, n. 1, p. 3–15, 1990.
- BROOKS, R. A. Intelligence without representation. *Artificial intelligence*, Elsevier, v. 47, n. 1, p. 139–159, 1991b.

BROOKS, R. A. Intelligence without reason. *The artificial life route to artificial intelligence: Building embodied, situated agents*, Hillsdale, NJ: Lawrence Erlbaum, p. 25–81, 1995.

BROOKS, R. A.; MAES, P. *Artificial Life IV: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*. [S.l.]: MIT Press, 1994b.

BROOKS, R. A.; STEIN, L. A. Building brains for bodies. *Autonomous Robots*, Springer, v. 1, n. 1, p. 7–25, 1994a.

CAMAZINE, S. et al. *Self-organization in biological systems*. Princeton, NJ: Princeton University Press, 2001.

CANGELOSI, A. Evolution of communication using symbol combination in populations of neural networks. In: IEEE. *Neural Networks, 1999. IJCNN'99. International Joint Conference on*. [S.l.], 1999. v. 6, p. 4365–4368.

CANGELOSI, A. Evolution of communication and language using signals, symbols, and words. *Evolutionary Computation, IEEE Transactions on*, IEEE, v. 5, n. 2, p. 93–101, 2001.

CANGELOSI, A. Solutions and open challenges for the symbol grounding problem. *International Journal of Signs and Semiotic Systems (IJSSS)*, IGI Global, v. 1, n. 1, p. 49–54, 2011.

CANGELOSI, A.; GRECO, A.; HARNAD, S. Symbol grounding and the symbolic theft hypothesis. In: *Simulating the evolution of language*. [S.l.]: Springer, 2002. p. 191–210.

CANGELOSI, A.; PARISI, D. The emergence of a 'language' in an evolving population of neural networks. *Connection Science*, Taylor & Francis, v. 10, n. 2, p. 83–97, 1998.

CANGELOSI, A.; PARISI, D. Computer simulation: A new scientific approach to the study of language evolution. In: *Simulating the evolution of language*. London: Springer, 2002. p. 3–28.

CANGELOSI, A.; RIGA, T. An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. *Cognitive science*, Wiley Online Library, v. 30, n. 4, p. 673–689, 2006.

CHALMERS, D. J. Subsymbolic computation and the chinese room. *The symbolic and connectionist paradigms: Closing the gap*, Hillsdale: Lawrence Erlbaum, p. 25–48, 1992.

CHRISTIANSEN, M. H.; KIRBY, S. Language evolution: Consensus and controversies. *Trends in cognitive sciences*, Elsevier, v. 7, n. 7, p. 300–307, 2003.

CLARK, A. Being there: putting brain, body, and world together again. 1997. *Cambridge, Mass.: MIT Press*, v. 19, n. 269, p. 4, 1995.

- COPELAND, B. J. *MYCIN*. 2015. <<http://http://global.britannica.com/EBchecked/topic/400270/MYCIN>>. Acessado em 10/01/2015.
- CORDESCHI, R.; FRIXIONE, M. Computationalism under attack. In: *Cartographies of the Mind*. [S.l.]: Springer, 2007. p. 37–49.
- DARPA. *Computer Learning of Linguistic Representations of the Physical World*. 2008. <<https://www.fbo.gov/index?s=opportunity&mode=form&id=c41a12e782331ebdd27d58ad755c2c11&tab=core&.cview=0>>. Defense Advanced Research Projects Agency. Acessado em 21/01/2015.
- DARPA. *Broad Operational Language Translation*. 2011. <<https://www.fbo.gov/utills/view?id=dfb543e2e101dd4288f538ea2af9da56>>. Defense Advanced Research Projects Agency. Acessado em 21/01/2015.
- DARPA. *Deep Exploration and Filtering of Text*. 2012. <<https://www.fbo.gov/utills/view?id=9645b2e015a728609daff66561339622>>. Defense Advanced Research Projects Agency. Acessado em 21/01/2015.
- DAVIS, R.; SHROBE, H.; SZOLOVITS, P. What is a knowledge representation? *AI magazine*, v. 14, n. 1, p. 17, 1993.
- DEACON, T. W. *The symbolic species: The co-evolution of language and the brain*. [S.l.]: WW Norton & Company, 1998.
- D'ERRICO, F.; HENSHILWOOD, C.; NILSSEN, P. An engraved bone fragment from c. 70,000-year-old middle stone age levels at blombos cave, south africa: implications for the origin of symbolism and language. *Antiquity*, Portland Press, v. 75, n. 288, p. 309–318, 2001.
- DIETRICH, E.; MARKMAN, A. B. Discrete thoughts: Why cognition must use discrete representations. *Mind & Language*, Wiley Online Library, v. 18, n. 1, p. 95–119, 2003.
- ELMAN, J. L. Learning and development in neural networks: The importance of starting small. *Cognition*, Elsevier, v. 48, n. 1, p. 71–99, 1993.
- EMMECHE, C. A semiotical reflection on biology, living signs and artificial life. *Biology and philosophy*, Springer, v. 6, n. 3, p. 325–340, 1991.
- EMMECHE, C. Um robô possui umwelt? reflexões sobre a biossemiótica qualitativa de jakob von uexküll. In: QUEIROZ, J.; LOULA, Â.; GUDWIN, R. (Ed.). *Computação, Cognição, Semiose*. [S.l.]: EDUFBA, 2007.
- FISHER, S. E. et al. Localisation of a gene implicated in a severe speech and language disorder. *Nature genetics*, v. 18, n. 2, p. 168–170, 1998.
- FLOREANO, D. et al. Evolutionary conditions for the emergence of communication in robots. *Current biology*, Elsevier, v. 17, n. 6, p. 514–519, 2007.

FLORIDI, L. Open problems in the philosophy of information. *Metaphilosophy*, Wiley Online Library, v. 35, n. 4, p. 554–582, 2004.

FODOR, J. A.; PYLYSHYN, Z. W. Connectionism and cognitive architecture: A critical analysis. *Cognition*, Elsevier, v. 28, n. 1, p. 3–71, 1988.

FRENCH, R. M. Moving beyond the turing test. *Communications of the ACM*, ACM, v. 55, n. 12, p. 74–77, 2012.

FROESE, T.; ZIEMKE, T. Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, Elsevier, v. 173, n. 3, p. 466–500, 2009.

GELDER, T. V. The dynamical hypothesis in cognitive science. *Behavioral and brain sciences*, Cambridge Univ Press, v. 21, n. 05, p. 615–628, 1998.

GLENBERG, A. M.; VEGA, M. de; GRAESSER, A. C. Framing the debate. In: VEGA, M. de; GLENBERG, A.; GRAESSER, A. (Ed.). *Symbols and Embodiment: Debates on Meaning and Cognition*. [S.l.]: Oup Oxford, 2008.

GONG, T.; SHUAI, L. Computer simulation as a scientific approach in evolutionary linguistics. *Language Sciences*, Elsevier, v. 40, p. 12–23, 2013.

GREEFF, J. D.; NOLFI, S. Evolution of implicit and explicit communication in mobile robots. In: *Evolution of communication and language in embodied agents*. [S.l.]: Springer, 2010. p. 179–214.

HABERMAS, J. Peirce and communication. In: KETNER, K. L. (Ed.). *Peirce and contemporary thought: philosophical inquiries*. [S.l.]: Fordham University Press, 1995.

HARNAD, S. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, Elsevier, v. 42, n. 1, p. 335–346, 1990.

HARVEY, I. Untimed and misrepresented: Connectionism and the computer metaphor. Citeseer, 1992.

HASELAGER, W. F. G. O mal estar do representacionismo: sete dores de cabeça da ciência cognitiva. São Paulo: Cultura Academica, 2005.

HINTON, G. E.; MCCLELLAND, J. L.; RUMELHART, D. E. *Distributed representations, Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*. [S.l.]: MIT Press, Cambridge, MA, 1986.

HOUSER, N. Introduction: Peirce as a logician. *Studies in the logic of Charles Sanders Peirce*, Indiana University Press, p. 1–22, 1997.

HURFORD, J. R. Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, Elsevier, v. 77, n. 2, p. 187–222, 1989.

- JÄGER, H. et al. What can mathematical, computational and robotic models tell us about the origins of syntax. *Biological foundations and origin of syntax*, v. 3, p. 385–410, 2009.
- LAHOZ-BELTRA, R.; PERALES-GRAVAN, C. A survey of nonparametric tests for the statistical analysis of evolutionary computational experiments. *INFORMATION THEORIES & APPLICATIONS*, p. 49, 2010.
- LEKVAM, T.; GAMBÄCK, B.; BUNGUM, L. Agent-based modeling of language evolution. In: *Proc. of 5th Workshop on Cognitive Aspects of Computational Language Learning (CogACL)@ EACL*. [S.l.: s.n.], 2014. p. 49–54.
- LIEBERMAN, P. *Eve spoke: Human language and human evolution*. [S.l.]: WW Norton & Company, 1998.
- LISZKA, J. J. *A general introduction to the semiotic of Charles Sanders Peirce*. [S.l.]: Indiana University Press, 1996.
- LOULA, A.; GUDWIN, R.; QUEIROZ, J. *Studying sign processes in the emergence of communication*. Quebec: Proceedings of 35th Annual Meeting of the Cognitive Science Society, 2013.
- LOULA, A. C. *Emergência de Comunicação e Representações em Criaturas Artificiais*. Tese (Doutorado) — Universidade Estadual de Campinas, Campinas, 2011.
- LYON, C.; NEHANIV, C. L.; CANGELOSI, A. *Emergence of communication and language*. Berlin: Springer, 2007.
- MAROCCO, D.; NOLFI, S. Emergence of communication in embodied agents evolved for the ability to solve a collective navigation problem. *Connection Science*, Taylor & Francis, v. 19, n. 1, p. 53–74, 2007.
- MAYO, M. J. Symbol grounding and its implications for artificial intelligence. In: AUSTRALIAN COMPUTER SOCIETY, INC. *Proceedings of the 26th Australasian computer science conference-Volume 16*. [S.l.], 2003. p. 55–60.
- MCCARTHY, J.; HAYES, P. *Some philosophical problems from the standpoint of artificial intelligence*. [S.l.]: Stanford University USA, 1968.
- MCCARTHY, J. et al. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI Magazine*, v. 27, n. 4, p. 12, 2006.
- MIROLLI, M. Representations in dynamical embodied agents: Re-analyzing a minimally cognitive model agent. *Cognitive science*, Wiley Online Library, v. 36, n. 5, p. 870–895, 2012.
- MIROLLI, M.; NOLFI, S. Evolving communication in embodied agents: Theory, methods, and evaluation. In: *Evolution of Communication and Language in Embodied Agents*. [S.l.]: Springer, 2010. p. 105–121.

MIROLLI, M.; PARISI, D. How can we explain the emergence of a language that benefits the hearer but not the speaker? *Connection Science*, Taylor & Francis, v. 17, n. 3-4, p. 307–324, 2005.

MIROLLI, M.; PARISI, D. How producer biases can favor the evolution of communication: An analysis of evolutionary dynamics. *Adaptive Behavior*, SAGE Publications, v. 16, n. 1, p. 27–52, 2008.

MONTANA, D. J.; DAVIS, L. Training feedforward neural networks using genetic algorithms. In: *IJCAI*. [S.l.: s.n.], 1989. v. 89, p. 762–767.

MORRIS, C. Signs, language and behavior. Prentice-Hall, 1946.

NEWELL, A.; SIMON, H. A. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, ACM, v. 19, n. 3, p. 113–126, 1976.

NOBLE, W.; DAVIDSON, I. Human evolution, language and mind. *CAMBRIDGE ARCHAEOLOGICAL JOURNAL*, Cambridge University Press, v. 7, p. 269–286, 1997.

NOLFI, S.; FLOREANO, D. Synthesis of autonomous robots through evolution. *Trends in cognitive sciences*, Elsevier, v. 6, n. 1, p. 31–37, 2002.

NOLFI, S.; MIROLLI, M. *Evolving communication in embodied agents: assessment and open challenges*. [S.l.]: Springer, 2010.

OLIVEIRA, E. S.; LOULA, A. C. *Symbol Interpretation in Neural Networks: an investigation on representations in communication*. Quebec: Proceedings of 36th Annual Meeting of the Cognitive Science Society, 2014.

PEIRCE, C. S. *The Collected Papers of Charles Sanders Peirce*, ed. C. Hartshorne and P. Weiss. [S.l.]: C. Hartshorne & P. Weiss). Harvard University Press, Cambridge, Mass, 1931–1935. Vols. I-VI.

PEIRCE, C. S. *The Collected Papers of Charles Sanders Peirce*, ed. A. W. Burks. [S.l.]: C. Hartshorne & P. Weiss). Harvard University Press, Cambridge, Mass, 1958. Vols. VII-VIII.

PEIRCE, C. S. *Annotated catalogue of the papers of Charles S. Peirce*, ed. Robin, Richard S. [S.l.]: University of Massachusetts Press, Amherst, 1967.

PESCHL, M. Autonomy vs. environmental dependency in neural knowledge representation. In: *Artificial Life IV*. [S.l.: s.n.], 1994. p. 417–423.

PFEIFER, R.; SCHEIER, C. *Understanding intelligence*. [S.l.]: MIT press, 1999.

PICCININI, G. Computationalism in the philosophy of mind. *Philosophy Compass*, Wiley Online Library, v. 4, n. 3, p. 515–532, 2009.

- PICEK, S.; GOLUB, M.; JAKOBOVIC, D. On the analysis of experimental results in evolutionary computation. In: IEEE. *MIPRO, 2012 Proceedings of the 35th International Convention*. [S.l.], 2012. p. 1064–1069.
- QUEIROZ, J. *Semiose segundo CS Peirce*. [S.l.]: EDUC São Paulo, 2004.
- QUEIROZ, J.; LOULA, Â.; GUDWIN, R. *Computação, cognição, semiose*. [S.l.]: EDUFBA, 2007.
- REITER, E.; DALE, R.; FENG, Z. *Building natural language generation systems*. [S.l.]: MIT Press, 2000.
- ROY, D. Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence*, Elsevier, v. 167, n. 1, p. 170–205, 2005.
- ROY, D.; REITER, E. Connecting language to the world. *Artificial Intelligence*, Elsevier, v. 167, n. 1, p. 1–12, 2005.
- RUMELHART, D. E. et al. A general framework for parallel distributed processing. *Parallel distributed processing*, v. 1, n. 2, 1986.
- RUTKOWSKA, J. *Reassessing piaget's theory of sensorimotor intelligence: a view from cognitive science*. [S.l.]: Citeseer, 1995.
- SAVAN, D.; CIRCLE, T. S. *An introduction to CS Peirce's full system of semeiotic*. [S.l.]: Toronto Semiotic Circle, 1988.
- SEARLE, J. R. Minds, brains, and programs. *Behavioral and brain sciences*, Cambridge Univ Press, v. 3, n. 03, p. 417–424, 1980.
- SEARLE, J. R. *Minds, brains and science*. [S.l.]: Harvard University Press, 1984.
- SEBEOK, T. A. *Signs: An introduction to semiotics*. [S.l.]: University of Toronto Press, 2001.
- SIMON, H. *The Shape of Automation, For Men and Management, 1965*. [S.l.]: Harper and Row, New York, 1965.
- SMITH, J. M. Honest signalling: the philip sidney game. *Animal Behaviour*, Academic Press, v. 42, n. 6, p. 1034–1035, 1991.
- SMOLENSKY, P. On the proper treatment of connectionism. *Behavioral and brain sciences*, Cambridge Univ Press, v. 11, n. 01, p. 1–23, 1988.
- STEELS, L. A case study in the behavior-oriented design of autonomous agents. In: *Proceedings of the third international conference on Simulation of adaptive behavior: from animals to animats*. [S.l.: s.n.], 1994. v. 3, p. 445–452.
- STEELS, L. The synthetic modeling of language origins. *Evolution of communication*, John Benjamins Publishing Company, v. 1, n. 1, p. 1–34, 1997.

STEELS, L. Language games for autonomous robots. *Intelligent Systems, IEEE*, IEEE, v. 16, n. 5, p. 16–22, 2001.

STEELS, L. Intelligence with representation. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, The Royal Society, v. 361, n. 1811, p. 2381–2395, 2003a.

STEELS, L. Evolving grounded communication for robots. *Trends in cognitive sciences*, Elsevier, v. 7, n. 7, p. 308–312, 2003b.

STEELS, L. The symbol grounding problem has been solved. so what's next. *Symbols and embodiment: Debates on meaning and cognition*, Oxford University Press USA, p. 223–244, 2008.

STEELS, L.; KAPLAN, F. Aibo's first words: The social learning of language and meaning. *Evolution of communication*, John Benjamins Publishing Company, v. 4, n. 1, p. 3–32, 2002.

STEELS, L.; VOGT, P. Grounding adaptive language games in robotic agents. In: *Proceedings of the fourth european conference on artificial life*. [S.l.: s.n.], 1997. v. 97.

SUN, R. Symbol grounding: a new look at an old idea. *Philosophical Psychology*, Taylor & Francis, v. 13, n. 2, p. 149–172, 2000.

SUN, R.; PETERSON, T. Some experiments with a hybrid model for learning sequential decision making. *Information Sciences*, Elsevier, v. 111, n. 1, p. 83–107, 1998.

TADDEO, M.; FLORIDI, L. Solving the symbol grounding problem: a critical review of fifteen years of research. *Journal of Experimental & Theoretical Artificial Intelligence*, Taylor & Francis, v. 17, n. 4, p. 419–445, 2005.

THOMAS, M. S.; MCCLELLAND, J. L. Connectionist models of cognition. *The Cambridge handbook of computational psychology*, p. 23–58, 2008.

TINBERGEN, N. On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, Wiley Online Library, v. 20, n. 4, p. 410–433, 1963.

TUCI, E. et al. An experiment on behavior generalization and the emergence of linguistic compositionality in evolving robots. *Autonomous Mental Development, IEEE Transactions on*, IEEE, v. 3, n. 2, p. 176–189, 2011.

TURING, A. M. Computing machinery and intelligence. *Mind*, JSTOR, p. 433–460, 1950.

VARSHAVSKAYA, P. Behavior-based early language development on a humanoid robot. Lund University Cognitive Studies, 2002.

VOGT, P. Bootstrapping grounded symbols by minimal autonomous robots. *Evolution of Communication*, v. 4, n. 1, p. 89–118, 2000.

VOGT, P. The physical symbol grounding problem. *Cognitive Systems Research*, Elsevier, v. 3, n. 3, p. 429–457, 2002.

WAGNER, K. et al. Progress in the simulation of emergent communication and language. *Adaptive Behavior*, SAGE Publications, v. 11, n. 1, p. 37–69, 2003.

WISCHMANN, S.; PASEMANN, F. The emergence of communication by evolving dynamical systems. In: *From Animals to Animats 9*. [S.l.]: Springer, 2006. p. 777–788.

YONG, C. H.; MIIKKULAINEN, R. Coevolution of role-based cooperation in multiagent systems. *Autonomous Mental Development, IEEE Transactions on*, IEEE, v. 1, n. 3, p. 170–186, 2009.

ZIEMKE, T. Rethinking grounding. In: *Understanding representation in the cognitive sciences*. [S.l.]: Springer, 1999. p. 177–190.

ZIEMKE, T.; SHARKEY, N. E. A stroll through the worlds of robots and animals: Applying jakob von uexkull's theory of meaning to adaptive robots and artificial life. *SEMIOTICA-LA HAYE THEN BERLIN-*, WALTER DE GRUYTER & CO, v. 134, n. 1/4, p. 701–746, 2001.

ZUBEN, F. J. V. Computação evolutiva: uma abordagem pragmática. *Anais da I Jornada de Estudos em Computação de Piracicaba e Região (1a JECOMP)*, v. 1, p. 25–45, 2000.