



UNIVERSIDADE NOVA DE LISBOA • FACULDADE DE CIÊNCIAS SOCIAIS E HUMANAS
DEPARTAMENTO DE FILOSOFIA

Paul Churchland e a Problemática da Semântica dos Estados Mentais

Dissertação para obtenção do Grau de Mestre em Filosofia Contemporânea pela
Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, sob a
orientação do Professor Doutor João Fonseca

Daniel José da Silva Ramalho

Lisboa, 2010

Índice

Agradecimentos	3
Introdução	4
I. Considerações Gerais sobre Paul Churchland	7
1.1. Percurso Académico de Paul Churchland	7
1.2. Naturalismo e Conexionismo	9
1.2.1. Naturalismo	9
1.2.2. Conexionismo	15
1.3. Programa Filosófico	17
1.4. Relevância da problemática da semântica dos estados mentais no contexto do programa neurofilosófico	18
II. O Paradigma Conexionista na Ciência Cognitiva	20
2.1. Origens da ciência cognitiva: o paradigma simbólico	20
2.2. Primeira geração de redes neuronais artificiais: origens e declínio do conexionismo	21
2.3. O “Novo Conexionismo”	26
2.4. Processamento Paralelo Distribuído (PPD)	27
2.4.1. Exemplo de uma rede de processamento paralelo distribuído	30
2.4.2. Espaços de activação	32
2.5. Plausibilidade neuronal do conexionismo	36
III. A Contenda Paul Churchland/Fodor & Lepore	41
3.1. Proposta inicial de Churchland	41
3.2. Primeiro conjunto de críticas de F&L	44
3.2.1. O problema da individuação de dimensões	45
3.2.2. O problema da relevância semântica	47
3.2.3. O problema da informação colateral	48
3.3. Segunda versão da proposta de Churchland	50
3.4. Segundo conjunto de críticas de F&L	51
3.5. Versão final da proposta de Churchland	53
3.6. O carácter reducionista da proposta de Churchland	59
3.7. Resumo sinóptico da teoria psicosemântica de Churchland	60
IV. Análise da Teoria Psicosemântica de Paul Churchland	62
4.1. Objecção 1: Homologia entre espaços de estado neuronais e semânticos	63

4.2. Objecção 2: Similaridade neuronal/similaridade semântica	65
4.3. Objecção 3: Composicionalidade	68
4.3.1. Composicionalidade lexical	69
4.3.2. Composicionalidade frásica	73
4.4. Nota conclusiva	75
Bibliografia	77

Agradecimentos

Gostaria de agradecer em primeiro lugar ao Instituto de Filosofia da Linguagem, e em particular aos seus Directores, os Profs. António Marques e João Sàágua, a quem devo o privilégio de poder exercer a minha paixão pela Filosofia todos os dias. Por isso, e pela aprendizagem que o acolhimento desta Instituição me possibilitou e continua a possibilitar (sem a qual a presente dissertação nunca poderia ter vindo à luz) estarei sempre grato.

Agradeço também aos meus pais, Ana Maria Ramalho e José Ramalho, pelo apoio incondicional que nunca deixaram de me dar ao longo de toda a minha vida, académica e não só, guiando os meus passos sempre que necessário sem nunca lhes ditarem um rumo. Por tudo o que me deram terão sempre a minha eterna gratidão e respeito, e o desejo de lhes retribuir tanto quanto possível com a felicidade que sei que têm em cada pequeno sucesso que seja meu continua a ser a minha mais importante fonte de motivação em tudo o que faço.

Estou também grato à minha namorada e futura mulher, Ana Gil, por mais coisas do que poderia enumerar aqui. Acima de tudo, agradeço-lhe pelo amor sem reservas que me dedica, manifestado nos vários sacrifícios que sempre fez por ter, por assim dizer, que me partilhar com a Filosofia, sem nunca vacilar no apoio que me dá em todos os aspectos possíveis, dia após dia. Finalmente, por ter aguentado o mundo à minha volta nas longas horas em que escrevia as linhas desta dissertação, agradeço-lhe, e continuarei a fazê-lo enquanto a minha vida durar.

Reservo para o fim um agradecimento especial ao João Fonseca. A minha dívida de gratidão para com ele transcende em muito a execução absolutamente irrepreensível da sua tarefa enquanto meu Orientador (que enfatizo com inicial maiúscula pelo facto de que um elogio suficiente às suas capacidades enquanto tal não caberia aqui). Pela sua competência enquanto filósofo, pelo seu carácter, pelas suas qualidades humanas, e por mais do que poderia enumerar aqui, constitui para mim um modelo a seguir e considerá-lo-ei sempre uma das melhores pessoas que tive o privilégio de conhecer e um dos melhores amigos que alguma vez terei.

A todos, muito obrigado.

Introdução

O foco temático da presente dissertação incidirá sobre a problemática da relação entre representações mentais e conteúdo semântico/intencional, nos moldes em que nos foi legada pela tradição, ramificada nas questões referentes: i) ao modo como os conceitos adquirem conteúdo semântico; ii) aos fundamentos da identidade conceptual na base da sua intersubjectividade; iii) à natureza dos elementos primitivos a partir dos quais conceitos complexos são formados; iv) às características do mecanismo epistémico subjacente às representações mentais que as vindica (ou não) enquanto portadoras de conteúdo semântico.

A motivação por detrás da escolha desta temática em específico liga-se directamente ao facto de o filósofo no qual se centrará a análise ser Paul Churchland, ou seja, foi secundária ao desejo de desenvolver um trabalho sobre este pensador em particular. Contudo, ainda que secundária não foi injustificada, na medida em que não obstante a obra de Churchland abarcar uma multiplicidade de questões filosóficas clássicas, a problemática da semântica dos estados mentais assume uma importância única na mesma porquanto, como se verá, a defensibilidade da teoria psicosemântica¹ que propõe é vital para que o seu programa filosófico como um todo possa reclamar legitimidade. A explicação do motivo por que assim é exigiria a articulação de um conjunto de noções que necessitariam de desenvolvimento prévio, pelo que será reservada para um momento posterior da presente dissertação.² Em todo o caso, a escolha da abordagem desta temática justifica-se ainda por três motivos mais genéricos: i) possibilita a introdução de todas as noções necessárias à compreensão sinóptica do projecto filosófico de Churchland; ii) torna evidente a metodologia mediante a qual Churchland procura dar resposta a problemas clássicos da filosofia, fazendo uso para tal do vasto manancial de informação experimental disponibilizado pela neurociência cognitiva, neurobiologia evolucionária, inteligência artificial, e outras ciências empíricas; iii) e permite verificar a inegável proficuidade desta metodologia, demonstrando que a teorização filosófica interdisciplinar não implica um afastamento da filosofia em benefício da ciência, e sim o estabelecimento de uma relação simbiótica mutuamente benéfica no contexto da qual ambas retêm a legitimidade no domínio que

¹ Uma teoria psicosemântica é, simplificada, uma teoria do modo como os estados mentais adquirem o seu conteúdo semântico/intencional sob a forma de conceitos.

² Cf. 1.4., Secção I

respectivamente lhes pertence – e, de facto, não obstante esta metodologia ir obviamente ao arripio de uma concepção milenar da teorização filosófica como uma tarefa estritamente racional, é interessante verificar que a teoria psicosemântica de Churchland, segundo as suas próprias palavras, revela ser no final surpreendentemente próxima da teoria das platónica das Formas do ponto de vista formal.

A análise a encetar dividir-se-á em quatro secções. Na primeira, tratar-se-á de se expor sumariamente o enquadramento de Paul M. Churchland (suas motivações filosóficas e influências) no contexto geral do pensamento filosófico e da problemática em questão em particular – tanto do ponto de vista histórico/sinóptico quanto em relação aos seus contemporâneos.

Na secção seguinte, expor-se-ão as noções fundamentais (históricas e técnicas) do modelo teórico subjacente a todo o projecto filosófico de Churchland, a saber, o *conexionismo*. Sucintamente, o conexionismo é caracterizável enquanto um paradigma explicativo do modo como a cognição se processa que nasceu do sucesso obtido pelas tentativas feitas a partir dos anos 40 de criar sistemas computacionais baseados nos cérebros biológicos enquanto redes de neurónios interconectados. O detalhe em que este modelo será descrito justifica-se pelo facto de ser vital para a compreensão da proposta semântica de Churchland a expor nas secções seguintes.

Embora desta feita o paralelismo seja menos apropriado, o desenvolvimento do pensamento de Churchland a respeito da semântica dos estados mentais tem ainda outra afinidade com a obra platónica na medida em que se deu mediante um processo *dialéctico*. Mais especificamente, a teoria psicosemântica que actualmente defende nasceu de uma proposta inicial mais rudimentar que foi sendo sucessivamente aperfeiçoada pelo confronto de cada nova versão da mesma com as objecções dos filósofos Jerry Fodor e Ernest Lepore – defensores de uma posição radicalmente oposta à de Churchland relativa aos mecanismos computacionais subjacentes ao processamento da cognição em geral –, numa troca de argumentos que se estendeu por quase uma década. A terceira secção será inteiramente dedicada a expor sinopticamente esta controvérsia, procurando-se assim evidenciar o processo de evolução do pensamento de Churchland relativo à temática em questão. A importância da exposição desta controvérsia advém ainda do facto de Fodor e Lepore, na diatribe que apontam a Churchland, darem voz a críticas de vários outros autores que seriam impossíveis de tomar em consideração individualmente neste espaço.

A quarta e última secção será de carácter mais especulativo. Aqui, o objectivo a cumprir será o de apresentar as objecções que Fodor e Lepore ergueram à versão final da proposta de Churchland e a que este último nunca chegou a responder. Subsequentemente, procurar-se-á avançar uma resposta possível para essas objecções partindo-se para tal sempre da base firme de textos de Churchland relevantes para a problemática em questão. Espera-se que desta analítica final resulte uma demonstração da robustez da teoria psicosemântica de Churchland que, *a fortiori*, conforme referido, constituirá um argumento a favor da viabilidade do seu programa filosófico em geral.³

³ Nota técnica: Todas as citações terão a ênfase em itálico no original. As legendas de todas as imagens foram traduzidas e adaptadas do original em Inglês.

I

Considerações Gerais sobre Paul Churchland

Os objectivos a cumprir nesta primeira secção serão três: i) expor em sinopse o percurso académico de Paul Churchland, evidenciando as influências de que o seu pensamento foi alvo ao longo da sua maturação; ii) introduzir os conceitos e noções necessários à subsequente exploração da sua teoria semântica; iii) apresentar genericamente o seu programa filosófico, situando-o no contexto global da história da Filosofia e explicitando a especial relevância de que a problemática da semântica dos estados mentais se reveste no mesmo (não obstante ser apenas uma parte num todo bastante mais abrangente). Naturalmente, a riqueza e extensão da obra de Churchland obrigam a que toda a tentativa de sumariá-la não deixe de ser superficial. A exposição que se segue não será excepção, conquanto se ressalve que todas as noções centrais para a presente dissertação cujo desenvolvimento resulte insuficiente no que se segue serão devidamente aprofundadas nas secções subsequentes.

1.1. Percurso académico de Paul Churchland

Paul M. Churchland (Canadá, 1942-...) licenciou-se na Universidade de British Columbia em 1964, tendo concluído o seu doutoramento em Filosofia na Universidade de Pittsburgh no ano de 1969 sob a orientação de Wilfrid Sellars. Nesses mesmos ano e universidade inicia a sua carreira de docente, após uma passagem de dois anos pela Universidade de Toronto, onde trabalhou como assistente. Ainda em 1969, aceita o cargo de Professor Assistente na Universidade de Manitoba, na qual permanece nos quinze anos seguintes.

Entre 1970 e 1978, publica uma série de artigos nos quais a influência de Sellars – defensor de um empirismo próximo do de Quine – é evidente. Nesta primeira fase, o seu trabalho é quase exclusivamente votado ao desenvolvimento da posição filosófica à qual ainda hoje o seu nome é mais comumente associado: o *materialismo eliminativista*⁴.

⁴ De acordo com esta posição, a rede de conceitos do senso comum que utilizamos comumente para nos referirmos a estados mentais como crenças, desejos e sentimentos (i.e. “atitudes proposicionais”) constitui uma teoria psicológica primitiva e radicalmente falsa que deve ser *eliminada* em benefício de um novo vocabulário informado pelos dados da neurociência contemporânea. Note-se que não obstante Paul Churchland ser desde há muito um dos principais defensores desta posição, não foi o primeiro a propô-la.

Durante esse período, Churchland acompanha de perto os desenvolvimentos recentes da neurociência, que passava então por um momento de revolução técnica e conceptual. A sua ligação às ciências do cérebro tornar-se-á progressivamente mais estreita nos anos seguintes, revelando-se determinante na consolidação das convicções que mantém até à actualidade. Igualmente determinante desde essa fase inicial da sua carreira até hoje tem sido a sua relação afectiva e intelectual com Patricia Churchland, esposa, colega, e co-autora de vários livros e artigos.

Em 1979, ainda em funções na Universidade de Manitoba, no cargo de Professor Associado, Churchland publica o seu primeiro livro – uma monografia intitulada *Scientific Realism and the Plasticity of Mind*, na qual sintetiza a posição filosófica do realismo científico ao mesmo tempo que lança as bases de uma noção a cujo aprofundamento dedicará parte significativa do seu trabalho ulterior, a saber, a da interconexão entre a epistemologia, as bases neurológicas da cognição e o processo cognitivo subjacente à teorização científica.

No ano de 1984, Churchland abandona o seu cargo anterior e torna-se docente na Universidade da Califórnia, San Diego (UCSD), onde ocupa hoje a *Valtz Chair of Philosophy*. Nessa instituição, o contacto com um grupo de investigação de ciência cognitiva dotou-o das ferramentas técnicas e conceptuais necessárias à maturação da teoria da mente que esboçara nos anos anteriores. A fecundidade dessa relação interteórica revelou-se nesse mesmo ano com a publicação de *Matter and Consciousness*⁵, o seu segundo livro, que ditará a tónica de toda a sua produção científica subsequente. Na sua maior parte, esta pequena obra assume os contornos de um manual introdutório às grandes linhas de pensamento na área da filosofia da mente e respectivo enquadramento histórico, dirigido a estudantes e a leigos em geral. A diferença que a singularizou entre as suas congéneres e a que se ficou a dever a sua popularidade imediata foi a de Churchland ter dedicado os seus capítulos finais à exposição de dados empíricos provenientes da ciência cognitiva, etologia, neuroanatomia e inteligência artificial e, concomitantemente, à demonstração da utilidade dos mesmos na resolução do confronto entre as várias hipóteses filosóficas tematizadas nos capítulos anteriores. Com esta obra, ao afirmar e fundamentar a viabilidade da abordagem de problemas tradicionalmente considerados do exclusivo âmbito da filosofia a partir de uma perspectiva científica, Churchland assumiu de

⁵ Revisto e actualizado em 1988, actualmente traduzido em seis línguas.

imediatamente um lugar de destaque na corrente filosófica da qual continua a ser um dos mais proeminentes defensores: o *naturalismo*.

1.2. Naturalismo e conexionismo

De modo a facilitar o exame da proposta de Churchland tanto no que tem de original quanto no que herdou dos seus predecessores (o que será necessário não só para a análise da sua troca de argumentos com Fodor e Lepore mas também para a compreensão das motivações filosóficas de ambos os lados da contenda) procurar-se-á no que se segue satisfazer dois propósitos: elucidar os pressupostos fundamentais da corrente *naturalista*, contextualizando sumariamente as suas origens e evolução na história da filosofia; e introduzir os princípios básicos do *conexionismo*, que desde há três décadas constitui o eixo científico do pensamento filosófico de Paul Churchland.

1.2.1. Naturalismo

O naturalismo caracteriza-se como a posição epistemológica de acordo com a qual não existe hierarquia ou hiato entre os conhecimentos filosófico e científico, e sim *continuidade*. Segundo a concepção tradicional (inaugurada por Platão), o domínio da filosofia é estritamente o do conhecimento *a priori*, sendo a Razão o único instrumento necessário e possível na abordagem dos problemas que lhe inerem. Todos os restantes pertencem ao âmbito da ciência e, não obstante o foco da filosofia poder em algumas instâncias incidir sobre o da prática científica, o contrário não se verifica. Na perspectiva naturalista, por outro lado, a filosofia não se caracteriza por um putativo estatuto epistemologicamente primordial das questões de que trata, para as quais seja metafisicamente impossível divisar um método de experimentação empírica com vista à obtenção de conhecimento que permita solucioná-las ou a iluminar o caminho a seguir na busca da sua solução. Os problemas considerados especificamente *filosóficos* distinguem-se dos científicos nesta perspectiva não pelo seu carácter intrinsecamente metafísico mas tão-somente por serem aqueles para os quais uma hipótese científica empiricamente testável *ainda* não foi desenvolvida.

Na concepção naturalista, a filosofia existe em relação simbiótica com a ciência – relação no âmbito da qual a informação flui livremente em ambas as direcções, sem que tenha cabimento a subsunção quer de certos tipos de questão, quer de conhecimento, às categorias independentes de “filosófico” e “científico”. A tarefa do filósofo naturalista

será, assim, a de providenciar a consolidação conceptual e teórica necessária a que o problema “filosófico” em causa atinja um ponto crítico de maturação a partir do qual possa transitar para o âmbito da ciência, num momento em que esta, por seu turno, esteja suficientemente maturada a nível teórico-técnico para abordá-la empiricamente. “Metafísico”, escreve Patricia Churchland, ‘é um rótulo que aplicamos a um *estádio* – um estágio de imaturidade, na realidade – do desenvolvimento científico de uma teoria, e não um campo distinto do saber com os seus próprios métodos’⁶.

Não obstante os primeiros filósofos auto-denominados naturalistas terem surgido apenas em meados do século XX, as origens do naturalismo *lato sensu* são bastante mais recuadas. Na verdade, identificam-se com as próprias origens da filosofia Ocidental, uma vez que o dealbar desta última é identificado com a tentativa por parte dos pensadores pré-socráticos de explicar a totalidade do mundo *natural* a partir de um número limitado de princípios fundamentais inferidos da experiência sensível, em alternativa ao mito e à intervenção divina. Hoje, os problemas sobre os quais incidiram os esforços especulativos daqueles que designamos de *filósofos* pré-socráticos (e.g. qual a natureza da matéria? qual a natureza dos astros? como se dá a reprodução dos seres vivos? por que razão a água congela e a madeira arde?) pertencem inteiramente ao domínio da ciência. Com base neste ensinamento histórico, o naturalismo propõe que assim como a especialização das várias disciplinas científicas se deu pela fragmentação do que outrora fora a *filosofia natural*, também as questões actualmente consideradas metafísicas pertencerão no futuro ao domínio da ciência.

O próprio conceito de “metafísica” deve a conotação supracientífica que ainda hoje lhe é atribuída por grande parte da comunidade filosófica à interpretação incorrecta por parte dos escoliastas latinos do verdadeiro intuito de um editor⁷ do *corpus aristotelicum* do primeiro século antes de Cristo, que terá forjado o termo para designar os tratados referentes à “filosofia primeira” – que Aristóteles deixara sem título –, como aqueles que deveriam ser estudados, literalmente, “*depois da Física*”⁸. Sendo os livros da “*Metafísica*” dedicados a expor os fundamentos de todas as ciências particulares, o iniciado deveria ter conhecimento prévio dos textos dedicados a essas ciências antes de começar o estudo dos de carácter mais genérico. Ademais, tudo indica que terá sido a maior abstracção dos tópicos que diziam respeito à “filosofia primeira” que levou

⁶ Churchland, Patricia, *Brain-Wise*, p. 40

⁷ Provavelmente Andrónico de Rodes.

⁸ O conjunto dos textos aristotélicos dedicados a questões referentes ao mundo natural.

Aristóteles a escolher esta última expressão para designá-los, e não o facto de os ter considerado epistemologicamente anteriores aos que diziam respeito ao mundo natural.

Segundo Paul Churchland, a perpetuação desta interpretação errónea na academia é uma das principais causas da fricção histórica entre a filosofia e a ciência, na medida em que tende a fundar a origem dessa fricção na autoridade aristotélica: ‘Tendencialmente aprendem-se a ética de Aristóteles, a sua lógica, e a sua teoria da percepção, por exemplo, em muito maior detalhe do que as suas cosmologia, biologia, o modo como entendia o espaço e o vácuo, ou a sua teoria do movimento [...]. Ao suprimirem-se ou ignorarem-se os elementos claramente proto-científicos de Aristóteles ignoramos alguns dos mais elevados momentos da filosofia, e emasculamos a nossa própria disciplina’⁹.

Ainda assim, são inúmeros os casos na filosofia pós-aristotélica em que a divisão entre conhecimento *a priori* e *a posteriori* não é claramente reconhecida, ou em que existe *de jure* mas não *de facto*. Os estudos empíricos de Berkeley sobre a visão¹⁰, por exemplo, tiveram um efeito indelével na teoria da percepção que viria a defender no contexto do mais radical idealismo. Kant, nas primeiras publicações da sua época pré-Crítica, dedicou-se quase em exclusivo ao campo da cosmologia e ao desenvolvimento de uma metafísica o mais possível adequada à teoria da matéria que lhe era dada a conhecer pelos mais recentes dados da ciência do seu tempo¹¹. Mesmo o dualismo de substâncias cartesiano não estava totalmente desligado da ciência: parte do motivo que levou Descartes a identificar a glândula pineal como o ponto de ligação entre o corpo e a alma terá sido sem dúvida o facto de ser uma das poucas componentes da anatomia cerebral que não se encontra duplicada e bilateralizada nos hemisférios cerebrais, como era exigido pela sua concepção unitária da alma.

Em todo o caso, a convicção no insulamento metodológico e epistemológico da filosofia permaneceu paradigmática até ao século XX e teve, de facto, a sua origem no racionalismo cartesiano. Descartes, ao encontrar a mais basilar certeza apodíctica – a da sua própria existência – no extremo metafísico do solipsismo em que culminara a radicalização da dúvida metódica, e ao propor a dedução de todas as outras verdades absolutas a partir dessa primeira, cristalizou a distinção entre conhecimento *a priori* e *a posteriori* (reificada na distinção entre *res cogitans* e *res extensa*), vindicando desse

⁹ Churchland, Paul (1986), “*The Continuity of Philosophy and The Sciences*”, *Mind and Language* 1, pp. 5-14

¹⁰ Berkeley, George (1709), *Ensaio para uma Nova Teoria da Visão*.

¹¹ Na verdade, toda a reflexão de Kant foi influenciada pelo seu estudo da física, que manteve ininterrupto ao longo de toda a vida.

modo a autonomia da tarefa e do objecto da filosofia. Concomitantemente, a sua proposta de fundação de um edifício de conhecimento inabalável sobre o *cogito* enquanto ‘*minimum quid [...] quod certum sit et inconcussum*’¹² originou a doutrina epistemológica que viria igualmente a tornar-se paradigmática no pensamento filosófico dos três séculos seguintes: o *fundacionalismo*¹³.

O legado cartesiano continuaria a influenciar a filosofia até ao século XX, estendendo-se, inclusivamente, ao próprio movimento que se revelaria determinante na afirmação do naturalismo: o *empirismo lógico*. Iniciado nas décadas de 20 e 30 em Viena e Berlim, o empirismo lógico propunha o abandono do então predominante idealismo hegeliano em benefício de uma concepção da epistemologia mais próxima do empirismo britânico. Duas razões principais motivaram esta reacção: os desenvolvimentos recentes na lógica e na teoria dos fundamentos da matemática trazidos por Wittgenstein, Russell e Whitehead; e o poder e alcance explicativos das ciências naturais tornados manifestos pela recente física einsteiniana e crescente consolidação da teoria evolucionária darwinista.

Ao lugar de destaque atribuído ao papel da ciência na teoria epistemológica adoptada pelos empiristas lógicos correspondeu um necessário cepticismo radical em relação à metafísica como um todo. O critério *verificacionista*, um dos seus pilares teóricos fundamentais, limitava os tipos de proposição cognitivamente significativa a dois únicos: as analíticas *a priori* – categoria em que se incluíam as tautologias lógicas e matemáticas; e as sintéticas *a posteriori*, ou seja, proposições sobre estados de coisas no mundo directamente fundadas nos dados dos sentidos, ou delas derivadas dedutivamente (reducionismo). Com o critério reducionista, os empiristas lógicos pretendiam fornecer uma matriz infalível de confirmação ou infirmação das proposições sintéticas, circunscrevendo os limites do discurso significativo acerca do mundo apenas ao que fosse passível de ser *reduzido* frase a frase¹⁴ à experiência imediata; ao passo que

¹² ('O mínimo que é certo e indubitável') Descartes, René (1641), *Meditações sobre Filosofia Primeira*, M.2, [1]

¹³ Designa-se de fundacionalista toda a teoria da justificação ou do conhecimento de acordo com a qual apenas sejam justificadas crenças básicas – fundadas numa base não-doxástica (axiomas) – e crenças dedutivamente derivadas de uma ou mais crenças básicas.

¹⁴ A tese de que o veículo básico de sentido seria a *frase* e não (como o era no empirismo clássico) o *termo*, foi originalmente importada da lógica fregeana para o empirismo lógico por Rudolph Carnap, na sua obra *Der Logische Aufbau der Welt* (1928).

a adopção de uma teoria *convencionalista*¹⁵ das linguagens formais preservaria a apodicticidade das proposições analíticas na impossibilidade (ditada pelo critério verificacionista) de aceitação de postulados idealistas ou racionalistas. Em consequência da estreiteza das limitações impostas por este modelo, conforme referido, todas as proposições metafísicas foram consideradas vazias de sentido e extirpadas da filosofia.

Em conjunto, o critério reducionista e a teoria convencionalista providenciariam, segundo acreditavam os empiristas lógicos, uma base inamovível a partir da qual a teorização científica poderia ser desenvolvida em segurança: a confirmação por confronto com os dados dos sentidos asseguraria a veracidade dos primitivos de toda a ciência que seriam as *proposições protocolo*¹⁶, e a irrevogabilidade *a priori* do núcleo da lógica e da matemática garantiria a conservação da verdade na construção de teorias por dedução a partir dessas proposições basilares. O empirismo lógico, como tal, não só resgatou o fundacionalismo da “filosofia primeira” cartesiana como é legítimo afirmar-se que constituiu sua forma maximamente depurada.

O movimento naturalista surgiu em meados do século XX precisamente a partir da reacção de um conjunto de filósofos que, concordando com a recusa da divisão epistemológica entre conhecimento filosófico e científico defendida pelos empiristas lógicos, não podiam aceitar os seus pressupostos fundacionalistas. Nessa reacção – que viria a ditar o fim do empirismo lógico –, a figura de maior destaque foi sem dúvida o filósofo W.V.O. Quine, em particular devido ao célebre artigo *Two Dogmas of Empiricism*, no qual apontou um ataque devastador aos dois “dogmas” centrais do empirismo lógico – a distinção entre proposições sintéticas e analíticas, e o reducionismo. Em termos sucintos, Quine demonstrou que a distinção sintético/analítico não é sustentável e que, dada a implicação directa dessa distinção na teoria da verificação de sentido, o reducionismo também o é, uma vez que se funda no critério verificacionista. Alternativamente à confirmação ou infirmação das proposições *frase a frase* do empirismo lógico, Quine propôs uma forma de holismo científico segundo o qual uma dada proposição sobre um determinado estado de coisas no mundo não derivaria o seu valor de verdade da sua confirmação pelos dados dos sentidos, mas sim da sua coadunação com o conjunto da ciência em geral: ‘O dogma do reducionismo

¹⁵ Segundo o convencionalismo, as verdades analíticas da lógica e da matemática são necessárias *a priori* não por serem fundamentais no sentido transcendente ou transcendental, mas por hipostasia das convenções imanentes ao próprio uso dessas linguagens. Por outras palavras, são necessárias *por convenção*.

¹⁶ Proposições pré-teóricas que exprimem o conteúdo da experiência imediata de um observador num determinado momento em termos elementares (não interpretativos).

sobrevive na suposição de que cada afirmação, considerada isoladamente, pode admitir confirmação ou infirmação. A minha contra-sugestão [...] é que as nossas afirmações sobre o mundo externo enfrentam o tribunal da experiência sensível não individualmente mas como um corpo colectivo'¹⁷.

Esta nova perspectiva epistemológica contrasta com a do empirismo lógico ainda pelo seu carácter distintamente anti-fundacionista: 'Enquanto empirista continuo a considerar o esquema conceptual da ciência, em última análise, como uma ferramenta para prever experiência futura à luz da experiência passada. Os objectos físicos são conceptualmente importados para a situação enquanto intermediários convenientes – não por definição em termos de experiência mas simplesmente como postulados irreduzíveis comparáveis, epistemologicamente, aos deuses de Homero'¹⁸. Em 1969, no seu não menos célebre artigo *Epistemology Naturalized*, Quine afirmou que a tarefa da epistemologia não deveria ser a de responder à questão *do que é* o conhecimento, e sim a de procurar explicar cientificamente *como adquirimos* conhecimento: 'A epistemologia, ou algo como ela, encaixa-se simplesmente no lugar de um capítulo da psicologia e, logo, da ciência natural. Estuda um fenómeno natural, viz., o sujeito humano físico'¹⁹. Por outras palavras, Quine propôs que a epistemologia perdesse o seu até então inquestionável estatuto de ramo independente da filosofia para que se tornasse algo como um acto introspectivo da ciência enquanto tal.

O projecto de Quine, com base no que foi dito, pode ser definido como uma tentativa de instauração de um diálogo entre filosofia e ciência onde até então havia existido dois monólogos, afirmando em simultâneo a primazia da ciência enquanto melhor forma de que dispomos de compreendermos não apenas o mundo natural, mas também o que significa compreender: '[...] a minha perspectiva é naturalista; vejo a filosofia não como uma propedêutica *a priori* ou trabalho de base para a ciência, mas sim como contínua com a ciência. Vejo a filosofia e a ciência como estando no mesmo barco – um barco que, para reverter à imagem de Neurath como tão frequentemente faço, apenas podemos reconstruir no mar, flutuando nele. Não há perspectiva externa privilegiada, não há filosofia primeira'²⁰. Esta última afirmação, talvez melhor do que qualquer outra, sintetiza aquele que foi sem dúvida o maior contributo de Quine para a filosofia em

¹⁷ Quine, W. V. O. (1951), "Two Dogmas of Empiricism", *The Philosophical Review* 60: 20-43, cap. V. *The Verification Theory and Reductionism*

¹⁸ Ibid., cap. VI. *Empiricism Without the Dogmas*

¹⁹ Idem (1969) "Epistemology Naturalized", in *Ontological Relativity & other Essays*, p. 82

²⁰ Ibid., "Natural Kinds", pp. 126-127

geral, e que justifica a um tempo o lugar de destaque que ocupa na sua história e a influência que exerce no pensamento de Paul Churchland: o facto de não se ter limitado a defender a perspectiva naturalista em alternativa à fundacionalista, mas ter demonstrado que *não existe* alternativa. Ao identificar falhas irresolúveis mesmo ao nível de extremo detalhe conceptual a que os empiristas lógicos refinaram o fundacionalismo, Quine deitou por terra em definitivo as aspirações de qualquer proposta futura de constituição de uma “filosofia primeira”. Na impossibilidade de nos alcandorarmos num patamar extra-teorético de neutralidade a partir do qual pudéssemos ajuizar comparativamente a adequação de uma dada teoria à realidade do mundo, a melhor teoria de que podemos dispor será apenas aquela que no momento presente mais se adequa aos dados empíricos, ficando permanentemente em aberto a possibilidade da sua substituição por uma outra de maior elegância e economia. Uma tal teoria, pois, será sempre necessariamente *científica*.

1.2.2. Conexionismo

A compreensão da especificidade da teoria epistemológica de Churchland e do modo como dá seguimento ao projecto filosófico iniciado por Quine conduz-nos ao segundo paradigma de importância central no seu programa filosófico, a saber, o *conexionismo*.

Nos anos 60, com o progressivo estabelecimento do paradigma cognitivista na psicologia²¹, desenvolveram-se dois modelos computacionais na área da inteligência artificial baseados em concepções distintas da natureza dos processos cognitivos, a saber, o lógico-linguístico (ou simbólico) e o conexionista²². A diferença radical entre ambos resulta do sistema de processamento de informação que inspirou o desenvolvimento de cada um – respectivamente, o computador digital e o cérebro. Assim, na concepção lógico-linguística, o processamento computacional cognitivo é considerado análogo ao de um programa informático: ambos funcionariam pela

²¹ O paradigma cognitivista instalou-se no contexto do movimento que ficou conhecido como “revolução cognitiva”, que transformou profundamente não só a psicologia mas também a inteligência artificial, a linguística, a neurociência, a antropologia e a filosofia, instituindo em simultâneo a área de investigação que agrega todas as anteriores num projecto comum: a ciência cognitiva. Esta última será considerada em maior detalhe na secção II da presente dissertação.

²² Não obstante ambos os modelos terem surgido quase em simultâneo, o “paradigma simbólico” foi dominante até aos anos 80, altura em que um rápido desenvolvimento de várias áreas da ciência cognitiva se reflectiu na acentuada consolidação teórica e técnica da proposta conexionista.

transformação de símbolos discretos (conceitos) de acordo com um conjunto fixo de instruções recursivas (sintaxe).

O conexionismo, por outro lado, propõe um modelo da cognição baseado no sistema nervoso central entendido como uma intrincada rede neuronal cujas unidades são por completo despojadas de conteúdo semântico (no sentido clássico do termo, de acordo com o qual os primitivos semânticos são conceitos) e cujo funcionamento global não obedece a qualquer tipo de matriz nomológica pré-definida.

A diferença entre os modelos lógico-linguístico e conexionista, naturalmente, implica diferentes concepções de qual deverá ser a metodologia da ciência cognitiva. Assim, segundo o modelo lógico-linguístico, o estudo empírico da microestrutura e funcionamento do cérebro pode revelar-se útil mas é, em última instância, secundário. Adoptar a perspectiva deste paradigma – segundo a qual a cognição humana está para o sistema nervoso central como um programa informático está para o computador digital que o executa –, implica assumir que o objectivo último da ciência cognitiva deverá ser o de isolar o “programa” da cognição humana em si e não o de explicar o modo como este se encontra especificamente implementado nos cérebros da nossa espécie. Enquanto sistema *abstracto* de manipulação de símbolos segundo operações lógicas, um tal “programa” poderá em princípio ser executado por qualquer sistema computacional suficientemente sofisticado²³, natural ou artificial. Não existe, pois, em estrita dependência do substrato físico do cérebro, pelo que o conhecimento deste último não será determinante para o estudo da cognição.

Já no modelo conexionista, a estrutura e funcionamento do cérebro diferem por completo dos de um computador digital, sendo a investigação neurobiológica de importância central na compreensão da actividade cognitiva. Segundo os Churchland, a analogia entre a actividade cognitiva e a computação informática é inaceitável na medida em que só pode subsistir na pressuposição de que o surgimento da espécie humana significou um salto *qualitativo* na escala filogenética, dado que a concepção simbólica da cognição exige que o funcionamento do nosso cérebro seja fundamentalmente diferente do das espécies incapazes de linguagem complexa – ou seja, todas as outras. Esta crença no carácter único do pensamento humano e a

²³ Este princípio, conhecido como argumento da *múltipla realização*, é um dos pilares centrais da teoria funcionalista da mente (inicialmente proposta nos anos 60 em alternativa ao behaviorismo e à teoria da identidade). De acordo com o funcionalismo clássico, os estados mentais definem-se exclusivamente pelo papel *funcional* que desempenham no sistema cognitivo, sendo o conhecimento do medium físico específico que em cada caso os realiza irrelevante para a sua compreensão.

secundarização das ciências empíricas do cérebro no seu estudo, concluem Paul e Patricia Churchland, são expressões de um atavismo filosófico que não é sustentável à luz da história que culminou no naturalismo: ‘[...] o funcionalismo avesso ao estudo do cérebro encontra-se metodologicamente próximo do Cartesianismo. No lugar da substância mental não-física de Descartes, o funcionalismo colocou o “*software*”²⁴.

1.3. Programa Filosófico

Em coerência com a sua posição naturalista, Paul Churchland defende que o conexionismo é o modelo da cognição mais profícuo, tanto por ser o que mais se adequa aos dados empíricos da neurociência quanto por ser o único em confluência com a teoria evolucionária, porquanto não implica a existência de uma distinção fundamental entre o mecanismo cognitivo humano e o de qualquer outra espécie – a diferença será somente *quantitativa*, ou seja, o cérebro humano será sem dúvida mais complexo do que o das restantes espécies mas emprega exactamente as mesmas estratégias computacionais que estas últimas. A adopção do conexionismo por parte de Churchland explica-se então pelo facto de vir consolidar sem contradição a afirmação naturalista de que a ciência constitui a melhor ferramenta de que dispomos para adquirir conhecimento, especificando exactamente *que* ciência no que concerne à investigação de todos os aspectos da natureza humana: a neurociência computacional²⁵.

O projecto desenvolvido por Paul Churchland nas últimas três décadas tem como objectivo aplicar os dados empíricos provenientes dessa ciência na resolução de problemas tradicionalmente considerados do exclusivo âmbito filosófico, como, por exemplo, os que dizem respeito à epistemologia, a ética, aos fundamentos da ciência ou à natureza da consciência (para referir apenas alguns dos que Churchland já abordou de modo profícuo). Conversamente, sendo a filosofia a disciplina da teorização e da integração conceptual por excelência, Churchland propõe a sua utilização na consolidação dos dados empíricos da neurociência computacional, iluminando caminhos futuros, criando hipóteses, aplicando a sua metodologia às tarefas de

²⁴ Churchland, Patricia (2002), *Brain-Wise: Studies in Neurophilosophy*, p. 27. No contexto original da citação, Patricia Churchland referia-se especificamente à teoria funcionalista de Jerry Fodor.

²⁵ O campo de investigação que aplica o modelo conexionista ao estudo dos cérebros biológicos. Conforme referido, o movimento conexionista originou no campo da inteligência artificial com base na ideia da criação de sistemas de computação a partir de unidades simples interconectadas *inspiradas* nos neurónios naturais, mas cujo funcionamento não tem que ser biologicamente realista.

clarificação e de síntese teórica e, acima de tudo, confrontando propostas científicas especulativas com o vasto repositório do pensamento filosófico acumulado ao longo de séculos. Hoje, vários outros filósofos estão envolvidos nesse programa de investigação interdisciplinar, conhecido genericamente pela designação cunhada por Patricia Churchland na sua obra seminal com o mesmo título: *Neurofilosofia*²⁶.

1.4. Relevância da problemática da semântica dos estados mentais no contexto do programa neurofilosófico

O objectivo da presente dissertação será estritamente o de avaliar os méritos e pontos de fragilidade da teoria semântica (conexionista) actual de Paul Churchland a partir da contenda que opôs a Jerry Fodor e Ernest Lepore, defensores de um modelo lógico-linguístico da cognição. Importa notar, contudo, que não obstante esta circunscrição temática da análise a desenvolver, o seu resultado final repercutir-se-á no programa neurofilosófico como um todo, devido ao facto de este ser fundamentalmente *reducionista*,²⁷ o que implica que a sua legitimidade depende directamente da validade da proposta neurosemântica, como ficará claro no que se segue.

Com base na sua concepção do mecanismo cerebral subjacente à representação mental e respectivo conteúdo semântico, Churchland aproxima-se da concretização integral da epistemologia naturalizada quineana (holista e anti-fundacionalista), com a excepção de que esta última sugeria a redefinição da epistemologia enquanto ramo da psicologia, ao passo que Churchland propõe a redução da epistemologia e da psicologia à neurociência. É importante relevar aqui que “redução” não equivale a “eliminação”. O materialismo eliminativista defendido por Paul Churchland tem como objecto de eliminação a psicologia do *senso comum*, não a *ciência* da psicologia. Em relação a esta última, diz Churchland: ‘[...] não deve presumir-se que a ciência da psicologia desaparecerá algures no processo [de redução], nem que o seu papel será limitado ao de alvo passivo de explicação neuronal. Pelo contrário, a química não desapareceu apesar da explicação das suas bases pela mecânica quântica; nem a ciência da biologia desapareceu apesar da explicação química das suas bases’.²⁸

²⁶ Churchland, Patricia (1986), *Neurophilosophy: Toward a Unified Science of the Mind-Brain*

²⁷ Como ficará claro nos parágrafos seguintes, o termo “reducionismo” refere neste ponto uma posição filosófica bastante diferente da que designava aquando do seu emprego mais acima no contexto do empirismo lógico.

²⁸ Churchland, Paul (1998), *On the Contrary*, p. 79

Churchland entende por “redução” o estabelecimento de uma relação de *equipotência* entre os conceitos de uma dada teoria científica e os de outra mais fundamental. Alguns casos exemplificativos de redução interteórica bem sucedida são os que enumera na citação acima. Trata-se, portanto, de um processo de correlação *semântica* por se tratar literalmente da tradução da linguagem de uma ciência para os termos de uma outra mais fundamental. Por esse motivo, se Fodor e Lepore estiverem correctos na sua convicção de que o mecanismo cognitivo humano é essencialmente simbólico, a psicologia será irreduzível à neurociência e reterá a sua autonomia teórica e conceptual. Nessa circunstância, a ciência mais fundamental para a qual os filósofos naturalistas deverão dirigir a sua atenção será, precisamente, a psicologia, de onde se segue necessariamente que o projecto neurofilosófico não poderá ser bem sucedido.

II

O Paradigma Conexionista na Ciência Cognitiva

A presente secção será inteiramente dedicada ao tópico do conexionismo, sendo três os objectivos a cumprir: i) expor em sinopse as origens históricas da ciência cognitiva e, nesse contexto, a evolução do paradigma conexionista; ii) introduzir a terminologia e princípios fundamentais do conexionismo; iii) apresentar sucintamente o problema da plausibilidade neuronal (i.e. a questão do grau de semelhança entre o funcionamento dos sistemas conexionistas artificiais e dos cérebros biológicos, ponto de importância fulcral para a viabilidade do projecto de Paul Churchland).

2.1. Origens da ciência cognitiva: o paradigma simbólico

Ainda que a afirmação da ciência cognitiva enquanto campo de investigação de direito próprio tenha sido gradual, a sua origem é frequentemente identificada com os grandes ciclos de conferências interdisciplinares que tiveram lugar em meados dos anos 50, que agregaram especialistas de todo o mundo de áreas como a psicologia, linguística e ciências computacionais,²⁹ entre outras. Desses encontros resultou a disseminação entre a comunidade científica de uma crença já esparsamente defendida por alguns pensadores da década anterior: a de que ‘[...] uma explicação adequada desses espaços em branco no mapa da ciência pode apenas ser feita por uma equipa de cientistas, cada qual especialista no seu próprio campo mas possuindo uma grande familiaridade, segura e treinada, com os campos dos seus vizinhos [...]’.³⁰ Noutros termos, gerou-se a consciência colectiva na academia de que a investigação dos fenómenos cognitivos só poderia progredir no contexto de uma cooperação inter e multidisciplinar, no âmbito da qual cada ramo científico individual contribuiria com as suas técnicas e métodos específicos para a construção de uma imagem sinóptica e coerente do objecto de estudo comum, em constante diálogo e confronto com as respectivas ciências adjacentes. A ciência cognitiva surgiu da reificação dessa consciência colectiva no estabelecimento de uma coligação entre disciplinas de cariz descritivo e empírico (psicologia cognitiva,

²⁹ Destes encontros, os mais significativos – ambos no ano de 1956 – foram o *Symposium on Information Theory* (M.I.T), e a *Dartmouth Summer Research Conference on Artificial Intelligence* (Dartmouth College), nos quais foram lançadas as bases, respectivamente, da ciência cognitiva e da inteligência artificial.

³⁰ Wiener, N. (1948), *Cybernetics*, p. 9

linguística, antropologia cognitiva e neurociência), especulativo (filosofia), e especulativo e aplicado (inteligência artificial) que, sem prejuízo da sua autonomia, passaram a interseccionar-se.

O eixo comum a todos os campos de investigação envolvidos neste esforço cooperativo – que assegura a coesão interna do mesmo não obstante as suas inúmeras fontes internas de fricção – é o de uma concepção da mente enquanto processador que recebe, armazena, recupera, transforma e transmite *informação*; ou seja, independentemente de quaisquer outras divergências, todas as disciplinas associadas à ciência cognitiva partem de uma perspectiva genericamente *computacionalista*. Conforme referido na secção anterior, os paradigmas conexionista e simbólico surgem como hipóteses explicativas radicalmente distintas do mecanismo *computacional* subjacente à actividade cognitiva humana. Não obstante ambos terem surgido quase em simultâneo, o paradigma simbólico – cuja acepção “clássica” é geralmente identificada com a *hipótese do sistema físico simbólico* de Newell e Simon,³¹ de acordo com a qual a cognição humana é um processo de manipulação de padrões de informação (segundo regras determinadas) que podem ser interpretados como *símbolos* (conceitos) – tornou-se rapidamente predominante. O surgimento dos primeiros computadores modelados segundo a arquitectura von Neumann³² foi um dos principais catalisadores dessa rápida ascensão (ainda que outros tenham existido). Dada a capacidade evidenciada pelo computador digital de execução de operações lógicas até então consideradas exclusivo apanágio do cérebro humano, a proposta de que a actividade cognitiva da nossa espécie teria na sua base um mecanismo de computação funcionalmente semelhante não pôde deixar de sair reforçada.

2.2. Primeira geração de redes neuronais artificiais: origens e declínio do conexionismo

O conexionismo opõe-se ao paradigma anterior fundamentalmente por consistir num modelo *assimbólico* de processamento de informação. Ao contrário dos sistemas clássicos de computação, cujo funcionamento se baseia na manipulação em série de

³¹ A. Newell, H. A. Simon (1976), “*Computer Science as empirical enquiry: Symbols and Search*”, Communications of the ACM, 19:113-126

³² A arquitectura von Neumann é o modelo “clássico” de computação característico do computador digital, com uma Unidade Central de Processamento (CPU) e uma unidade de armazenamento, ou memória, que contém em simultâneo a informação acessível pelo utilizador (data) e as instruções para a manipulação interna da mesma (programa).

símbolos discretos segundo um conjunto fixo de regras, os sistemas conexionistas transformam informação mediante o ‘processamento em paralelo de sub-símbolos, utilizando propriedades estatísticas em vez de regras lógicas’.³³ Os fundamentos deste paradigma foram lançados em 1943 por Warren McCulloch (neurocientista) e Walter Pitts (lógico) num célebre artigo³⁴ no qual propuseram um sistema de processamento de informação cuja unidade computacional básica fosse o “neurónio formal” (artificial),³⁵ modelado para emular as seis propriedades funcionais dos neurónios naturais:

1. *Receptiva* (capacidade de receber *inputs* de outros neurónios);
2. *Integrativa* (capacidade de integrar e de manipular esses *inputs*);
3. *Condutiva* (capacidade de veicular a informação integrada ao longo do espaço físico);
4. *Emissora* (capacidade de transmitir informação – *output* – a outros neurónios);
5. *Computacional* (capacidade de transformar um tipo de informação noutra);
6. *Representacional* (capacidade de contribuir para a formação de representações internas).

Neste modelo, os neurónios formais encontram-se divididos em *camadas* paralelas interconectadas (de onde provém a designação “conexionismo” ou, alternativamente, “redes neuronais artificiais”). Tendo formulado os princípios subjacentes ao funcionamento das redes neuronais artificiais na linguagem da lógica proposicional, McCulloch e Pitts demonstraram formalmente que sistemas com este tipo de configuração possuem a capacidade de executar as operações lógicas da conjunção, disjunção e negação. Adicionalmente, demonstraram que qualquer processo passível de ser executado num número finito de passos mediante as operações lógicas dos três tipos referidos pode, em princípio, ser executado por um sistema conexionista – ou, por outras palavras, que tais sistemas possuem, em teoria, o poder computacional de uma máquina de Turing universal.

O enfoque de Pitts e McCulloch no seu artigo conjunto seguinte incidiu sobre a capacidade cognitiva de *reconhecimento de padrões* e a possibilidade de emulá-la em

³³ Medler, David (1998), “A Brief History of Connectionism”, *Neural Computing Surveys*, 1 (2), p. 21

³⁴ McCulloch, W.S., Pitts, W. (1943), “A logical calculus of the ideas immanent in nervous activity”, *Bulletin of Mathematical Biophysics*, 5

³⁵ Este primeiro neurónio artificial recebeu o nome de Threshold Logic Unit (TLU).

sistemas conexionistas³⁶. Com base nos fundamentos teóricos lançados neste trabalho, Frank Rosenblatt, um investigador norte-americano na área das ciências computacionais, desenvolveu o primeiro sistema artificial capaz de formar representações canónicas espontaneamente: o *Perceptron*.³⁷

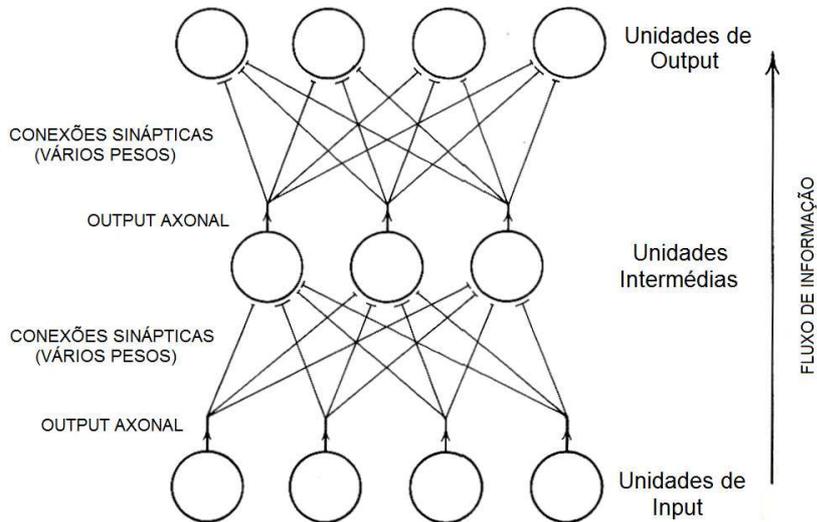


Fig. 1 Esquemática da arquitectura computacional do *Perceptron*.³⁸

O *perceptron* comum é constituído por três camadas de unidades computacionais simples denominadas *neurónios formais*: uma camada sensória, ou de *input* (receptora de estímulos externos), uma ou mais camadas intermédias (também designadas de camadas “escondidas”), e uma camada de resposta, ou de *output*. A comunicação entre neurónios neste modelo é estritamente unidireccional – da camada neuronal de *input* para a de *output*.³⁹

Ao activar-se, um neurónio envia um sinal ao longo do seu *axónio* (conexão de *output*), que se ramifica para estabelecer *sinapses*⁴⁰ com os neurónios da camada seguinte, sendo possível mas não necessário que essa conectividade seja *massiva*, isto é,

³⁶ Pitts, W., McCulloch, W.S. (1947), “How we know universals: the perception of auditory and visual forms”, Bulletin of Mathematical Biophysics, 9

³⁷ Rosenblatt, F. (1958), “The Perceptron: A probabilistic model for information storage and organization in the brain”, Psychological Review, 65:386-408

³⁸ Churchland, Paul (1992), “On the Nature of Theories: A Neurocomputational Perspective”, in *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, p. 162

³⁹ As arquitecturas conexionistas unidireccionais designam-se de *arquitecturas feedforward*.

⁴⁰ Zona terminal do axónio, onde este contacta com o neurónio da camada seguinte. A transmissão de sinais neuronais dá-se ao nível da sinapse.

que cada neurónio de uma camada esteja conectado a *todos* os da seguinte. A activação ou desactivação dos neurónios da camada inicial depende do estímulo que receberem do exterior. Nas camadas seguintes, essa activação ou desactivação é determinada pelo somatório dos valores dos sinais enviados pelos neurónios da camada imediatamente anterior a que estejam ligados, modificados pelo *peso*⁴¹ das conexões (inibitórias ou excitatórias)⁴² que esses sinais atravessam.

Esta estrutura funcional básica, note-se, estava já presente na proposta original de McCulloch e Pitts. As principais inovações introduzidas por Rosenblatt ao modelo dos seus antecessores foram: a criação de conexões entre neurónios com pesos *contínuos* (ou seja, ao contrário das conexões binárias, os valores das conexões deste modelo podiam variar no seu grau);⁴³ e a introdução de procedimentos matemáticos para o ajustamento do valor do peso de cada conexão individual. Uma vez que é a configuração dos pesos que dita o comportamento da rede como um todo, esta última inovação tornou possível *treinar* a rede para executar tarefas tais como, por exemplo, reconhecer padrões.

O “treino” de uma rede consiste na apresentação repetida de um mesmo padrão de *input* à primeira camada de neurónios e subsequente ajustamento dos pesos das conexões cujos neurónios receptores tenham produzido um *output* incorrecto, ou seja, que tenham contribuído para afastar o comportamento global do sistema do desejado. Repetido este processo vezes suficientes, a rede atinge a configuração global óptima dos vários pesos, o que lhe permite gerar o *output* desejado. Com base neste procedimento de treino, Rosenblatt provou o *Teorema da Convergência do Perceptron*, demonstrando que um *perceptron* que incorpore um tal processo de correcção de erro tem a capacidade teórica de resolver infalivelmente qualquer problema de reconhecimento de padrões desde que: i) uma solução exista; ii) cada padrão seja apresentado à rede pelo menos duas vezes, e iii) as conexões entre a primeira camada de neurónios e a camada intermédia permaneçam fixas.

⁴¹ ⁴¹ O “peso” de uma conexão, situado na sua sinapse, modifica o valor do sinal que a atravessa, influenciando o modo como este afectará o neurónio receptor.

⁴² Se a conexão entre um neurónio e outro for *inibitória*, a activação do primeiro contribuirá para a desactivação do segundo. Se a conexão for *excitatória*, contribuirá para a sua activação. Em ambos os casos, a medida da influência do sinal depende do peso da conexão.

⁴³ No caso das unidades binárias, como o eram as do modelo de McCulloch e Pitts, os pesos admitem apenas dois valores: 1 (conexões excitatórias), e -1 (conexões inibitórias). Unidades contínuas admitem um número arbitrário de valores intermédios possíveis entre 0 e 1.

Ainda que teoricamente este modelo possua um grande poder computacional, sofre de uma grave limitação no que se refere ao ponto iii): não fornece por si só um método fiável de ajustamento dos pesos fixos das conexões entre as camadas neuronais inicial e intermédia. Como tal, na prática, o *perceptron* só possui capacidade de aprendizagem nos casos em que se compõe de apenas duas camadas neuronais – uma inicial e uma de resposta – o que implica que o seu poder computacional real seja extremamente reduzido, porquanto o reconhecimento de certo tipo de padrão por sistemas conexionistas é matematicamente impossível sem a existência de camadas neuronais intermédias.

De modo a evitar este tipo de limitação, Oliver Selfridge, pioneiro da inteligência artificial, desenvolveu um novo tipo de sistema conexionista: o *Pandemonium*.⁴⁴ Selfridge obviou aos referidos problemas de computação em redes com camadas intermédias propondo uma arquitectura neuronal na qual os neurónios não fossem funcionalmente idênticos ao longo de toda a rede, como o eram no *perceptron*, mas que desempenhassem tarefas distintas consoante a camada em que se encontravam.⁴⁵ Porém, ainda que este sistema se tenha revelado eficaz na execução de tarefas de maior complexidade das que estavam ao alcance do modelo de Rosenblatt e demonstrado a viabilidade de redes neuronais com múltiplas camadas, revelou-se não menos limitado do que o *perceptron* na sua capacidade de reconhecimento de padrões pela rigidez que era necessariamente imposta às conexões entre as camadas neuronais, na sua grande maioria ajustadas *a priori* consoante a tarefa a realizar.

Uma vez que nesta primeira fase (anos 50 e 60) todos os sistemas conexionistas de aprendizagem existentes se baseavam, *grosso modo*, nos paradigmas determinados pelos modelos propostos por Rosenblatt e Selfridge, a investigação dedicada ao conexionismo entrou numa fase de declínio generalizado. Apenas dois tipos de rede neuronal artificial eram viáveis, ambos com grandes limitações: por um lado, redes pequenas, sem camadas intermédias, matematicamente equivalentes a máquinas de Turing universais na sua capacidade de aprendizagem mas computacionalmente deficientes na prática (e.g. *perceptrons*); e por outro, redes de grandes dimensões, computacionalmente poderosas mas com muito limitadas nas suas capacidades de

⁴⁴ Selfridge, O. G. (1959), "*Pandemonium: A paradigm for learning*", Proceedings of the Symposium on Mechanisation of Thought Processes, pp. 511-529

⁴⁵ De modo a enfatizar a diferença entre os típicos neurónios formais indiferenciados e as unidades computacionais especializadas do seu sistema, Selfridge adoptou o termo "demónio" em vez de "neurónio" (tipos diferentes de "demónio" desempenhavam funções diferentes), o que justifica o nome "*Pandemonium*".

aprendizagem (e.g. *Pandemonium*). Sem uma forma de combinar ambos os tipos de rede, o conexionismo não podia afirmar-se como um candidato a modelo explicativo da actividade cognitiva e foi progressivamente preterido em benefício de sistemas computacionais simbólicos.

A partir de 1969, com a publicação de *Perceptrons*, de Minsky e Papert, a investigação dedicada aos sistemas conexionistas tornou-se praticamente inexistente. Nesta obra, os autores demonstraram que as limitações práticas do modelo de Rosenblatt eram bastante mais extensas do que inicialmente se julgara. O exemplo central nessa demonstração foi o caso da operação da *disjunção exclusiva* (ou “função XOR”), que provaram ser matematicamente incomputável por uma rede neuronal sem camadas intermédias, o que implicava que o mesmo se verificaria em relação a um grande número de outras operações lógicas com exigências computacionais semelhantes. Tendo conjecturado que o problema se manteria para operações mais complexas ainda que se acrescentasse um número arbitrário de camadas intermédias (uma vez que seria impossível especificar um processo fidedigno de treino do sistema), Minsky e Papert concluíram que toda investigação em sistemas neuronais de múltiplas camadas estava à partida condenada à esterilidade. Desse momento em diante, o enfoque da ciência cognitiva passou a incidir quase em exclusivo sobre o paradigma simbólico,⁴⁶ o que mudou apenas na década de 80 com a demonstração de que a referida conjectura de Minsky e Papert estava incorrecta.

2.3. O “Novo Conexionismo”

As redes neuronais do período do “Novo Conexionismo”⁴⁷ caracterizam-se por conjugarem um grande poder computacional e algoritmos de aprendizagem altamente desenvolvidos. Destes últimos, o mais importante (na verdade, o catalisador do ressurgimento do conexionismo na ciência cognitiva) foi o da “*regra delta generalizada*”, ou “algoritmo da *retropropagação de erro*”, comumente designado apenas por “algoritmo da retropropagação”. Ainda que proposto independentemente por três fontes distintas, a sua popularização ficou a dever-se ao trabalho de Rumelhart,

⁴⁶ Note-se que a investigação em ciência cognitiva dedicada ao conexionismo nesta fase não cessou por completo, tornando-se apenas a centrada nos mecanismos cerebrais subjacentes à *memória*, e não já ao *raciocínio*.

⁴⁷ Medler, David, *ibid.*, p. 41

Hinton e Williams.⁴⁸ A regra delta generalizada, como o nome indica, foi concebida como uma generalização da *regra delta Widrow-Hoff*⁴⁹, um algoritmo de aprendizagem criado em 1960 especificamente para a rede neuronal *Adaline*.⁵⁰ Ao contrário deste último algoritmo, cuja aplicação se limitava a sistemas de duas camadas neuronais, a regra delta generalizada não impõe limites teóricos ao número de camadas intermédias. Este algoritmo possibilita a aprendizagem gradual do sistema propagando a medida de erro (i.e. a discrepância entre o *output* global do sistema e o *output* desejado, dado o estímulo apresentado à camada inicial) calculada nas unidades de *output* no sentido inverso do fluxo de informação inicial, ou seja, da camada final em direcção às intermédias (*retropropagação*), e ajustando o valor dos pesos das conexões nessas camadas para um valor que gradualmente aproxime o comportamento da rede daquele que é desejado. Em termos mais simples, o algoritmo da retropropagação permite calcular numericamente em que medida cada unidade de cada camada contribuiu para o erro global do sistema, e corrigir a diferença.⁵¹

Esta inovação foi de importância fulcral no desenvolvimento do paradigma que caracteriza a maioria dos sistemas conexionistas actuais, a saber, o *processamento paralelo distribuído*, assim designado pelo facto do processamento de informação nestes sistemas se dar ao longo de um número variável de camadas neuronais *paralelas*, e de as representações ou itens semânticos que incorporam se encontrarem *distribuídos* pela rede,⁵² especificamente na configuração dos pesos das conexões neuronais.

2.4. Processamento Paralelo Distribuído (PPD)

As principais componentes deste tipo de arquitectura são i) unidades computacionais simples (neurónios formais); ii) conexões entre essas unidades com polaridade inibitória ou excitatória e pesos e variáveis; iii) equações que determinam o valor total da activação de cada unidade a cada momento; iv) regras de aprendizagem

⁴⁸ Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986), "*Learning representations by back-propagating errors*", *Nature*, 323:533-536

⁴⁹ Widrow, B., Hoff, M.E. (1960), "*Adaptive switching circuits*", 1960 IRE WESCON, Convention Record, pp. 96-104

⁵⁰ *Adaptive Linear Neuron*, ou *Adaptive Linear Element*.

⁵¹ Os sistemas que incorporam o algoritmo da retropropagação opõem-se aos sistemas *feedforward* (unidireccionais), e designam-se de sistemas *interactivos* ou *bidireccionais*.

⁵² Estes sistemas, designados de redes de representação *distribuída*, opõem-se às redes neuronais híbridas de representação *localista*, nas quais cada neurónio formal codifica uma representação unitária (e.g. Pandemonium).

que permitem a alteração do comportamento global do sistema pela alteração gradual dos pesos das suas conexões.

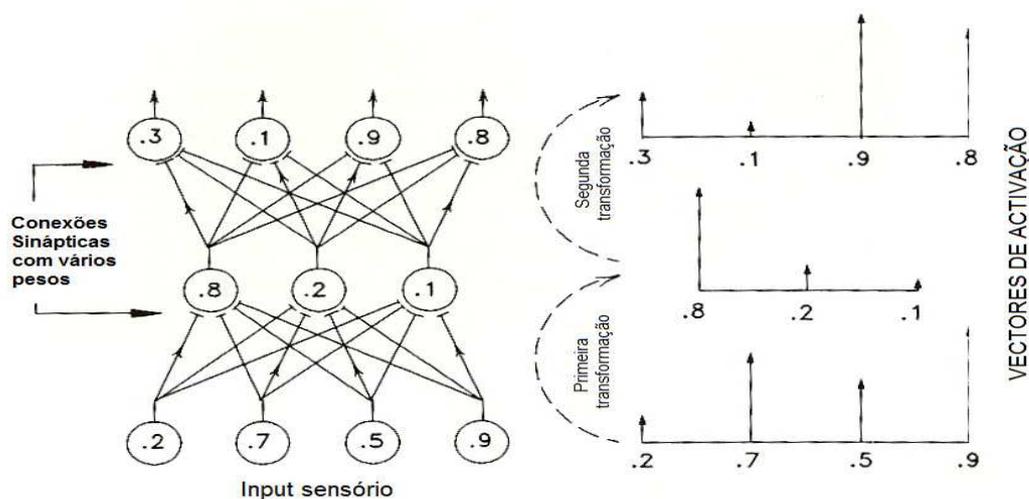


Fig. 2 Esquemática da arquitectura computacional uma rede de processamento paralelo distribuído e respectivos vectores de activação.⁵³

À excepção da regra de aprendizagem que incorporam, as redes de PPD são funcionalmente idênticas ao *perceptron*. Enquanto tal, nos restantes aspectos, a exposição que se segue pode ser considerada como um aprofundamento das noções previamente apresentadas a respeito da arquitectura proposta por Rosenblatt.

Na rede da figura 2, como no *perceptron*, os círculos da camada inferior representam os neurónios sensoriais (i.e. de *input*). Cada um destes projecta um axónio individual que se ramifica de modo a estabelecer sinapses directamente com o “corpo celular” de cada neurónio na segunda camada. O nível de activação dos neurónios da camada inicial é uma função do estímulo externo que lhe é apresentado. Do mesmo modo, o nível de activação da camada intermédia corresponde ao somatório dos valores dos sinais recebidos de todas as unidades da camada inicial. O valor do sinal emitido individualmente por cada neurónio é calculado por uma equação com três variáveis: i) o nível de activação desse neurónio, que ditará a *força* do sinal;⁵⁴ ii) o *peso* da conexão que o sinal atravessa; e iii) a *polaridade* dessa conexão (de acordo com a qual o valor numérico do sinal contará como negativo ou positivo para efeitos do somatório operado no neurónio receptor).

⁵³ Churchland, Paul (1998), *On The Contrary*, p. 58

⁵⁴ Comumente, a força de um sinal equivale à sua frequência, ou seja, ao número de vezes que é emitido pelo neurónio activo numa dada unidade temporal.

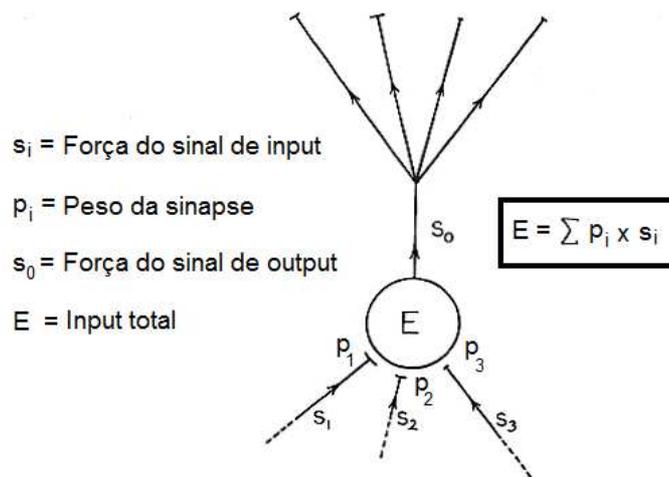


Fig. 3 Um neurónio artificial.⁵⁵

O conjunto global dos níveis de activação dos neurónios de uma dada camada neuronal constitui a *representação* dessa camada para o estímulo apresentado. Esta configuração de níveis de activação, uma vez que se trata de um conjunto ordenado de magnitudes numéricas entre 0 e 1, pode ser matematicamente descrita como um *vector*, sendo o padrão produzido pelo total de activações dos neurónios de uma dada camada o seu *vector de activação*. Por exemplo, o vector de activação da camada inicial da rede neuronal na figura 2 é { .2, .7, .5, .9}, sendo essa a representação dessa camada do estímulo recebido. Uma vez calculado, o vector é então propagado para a camada seguinte ao longo das várias conexões e modificado consoante o peso das mesmas, o valor total resultante determinando o nível de activação de cada neurónio receptor.

O vector de activação da camada neuronal intermédia, por sua vez, é enviado pelo conjunto seguinte de conexões e produz um outro na camada de resposta. O vector de activação desta última camada, sendo o *output* final da rede, é a representação global do sistema do estímulo apresentado à camada inicial. Dada esta arquitectura computacional, Churchland descreve as redes de PPD como ‘mecanismos de transformação sucessiva de um vector de activação sensorio inicial numa sequência de vectores de activação, incorporados numa sequência descendente de populações neuronais’.⁵⁶ Não obstante a sua simplicidade teórica, este tipo de rede suplanta vastamente os sistemas simbólicos em poder computacional, capacidade de

⁵⁵ Churchland, Paul (1992), “On the Nature of Theories: A Neurocomputational Perspective”, in *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, p. 160

⁵⁶ Churchland, Paul (2008), “What Happens to Reliabilism Liberated from Propositional Attitudes?”, in *Neurophilosophy at Work*, p. 98

discriminação categorial espontânea, resistência à danificação e rapidez de processamento de informação.

2.4.1. Exemplo de uma rede de processamento paralelo distribuído

De modo a elucidar algumas propriedades de sistemas conexionistas de maior complexidade, bem como o funcionamento da regra delta generalizada, será utilizado um exemplo recorrente de Paul Churchland: o de uma rede neuronal artificial à qual é dada a tarefa de distinguir entre os ecos de sonar produzidos por rochas e por minas explosivas subaquáticas. Como Churchland refere, este exemplo é de particular interesse na medida em que a tarefa apresenta duas dificuldades à rede: os padrões acústicos produzidos pelas rochas e pelas minas são muito semelhantes; e existe uma grande variação de padrões acústicos mesmo dentro de cada categoria, uma vez que as rochas, assim como as minas, produzem imagens de sonar muito distintas consoante as suas formas, dimensões e orientações relativamente ao observador.

A rede neuronal em causa possui três camadas: uma de *input* com treze neurónios, uma intermédia com sete, e uma de *output*, com dois. Uma vez que as categorias entre as quais a rede deverá aprender a discriminar são apenas duas (rocha/mina), pretende-se que no final do processo de treino os vectores de activação dos neurónios da camada final se reduzam apenas a dois: $\{1, 0\}$ se o estímulo inicial apresentado corresponder ao eco de uma mina, e $\{0, 1\}$ se corresponder ao de uma rocha.

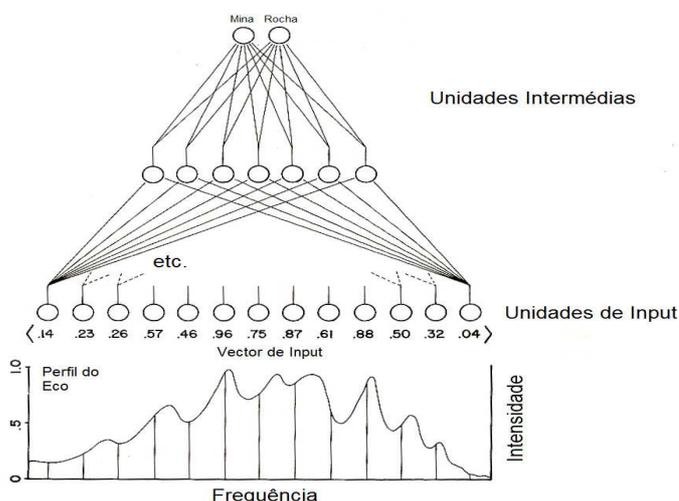


Fig. 4 Rede treinada para discriminar entre padrões de sonar de rochas e de minas.⁵⁷

O processo de treino consiste na apresentação de um conjunto significativo de ambos os tipos de estímulo à camada inicial, que deverá aprender a categorizá-los correctamente. No caso presente, cinquenta padrões acústicos de minas e cinquenta padrões acústicos de rochas (de diferentes tamanhos, formas e orientações) previamente

⁵⁷ Churchland, Paul (1992), "On the Nature of Theories: A Neurocomputational Perspective", in *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, p. 165

gravados foram digitalizados por um analisador de frequências, que registou cada padrão acústico individual de acordo com os níveis de energia que apresentava em treze frequências diferentes. Subsequentemente, cada uma dessas frequências foi apresentada a um neurónio na camada de input. Assim, o nível de activação de cada um dos treze neurónios da camada sensória da rede dependerá da frequência que unicamente lhe corresponde.

Uma vez que os pesos das conexões de toda a rede possuem valores de partida aleatórios, os *outputs* na fase inicial do treino resultantes reflecti-lo-ão, sendo igualmente caóticos. De modo a corrigir esta situação, o valor da divergência entre o *output* desejado e *output* real da rede é computado através da regra delta generalizada para cada padrão acústico apresentado. Este processo permite identificar quais as conexões que contribuíram em maior medida para o erro global do sistema e ajustar gradualmente os valores dos seus pesos de modo a aproximar o comportamento global da rede tanto quanto possível do *output* desejado para cada caso individual de rocha ou mina. Repetido este processo vezes suficientes, a rede adquire a capacidade de distinguir com uma grande medida de fiabilidade entre padrões acústicos de rochas e de minas, incluindo padrões que não lhe tenham sido apresentados durante o processo de treino. Por outras palavras, a rede aprende a categorizar correctamente estímulos que nunca antes encontrara, extrapolando a partir da sua “experiência passada”.

Segundo Paul Churchland, a adopção de um tal processo de correcção de erro equivale a dotar o sistema de um “professor”,⁵⁸ na medida em que fornece à rede as ferramentas para se auto-corrigir passo a passo até um ponto em que consiga executar a tarefa adequadamente, sem necessidade da “supervisão” que tivera na fase de treino. Contudo, note-se que esta analogia não é perfeita porquanto a rede não foi explicitamente “ensinada” a distinguir entre os dois tipos estímulo que lhes foram apresentados. O processo de treino consistiu em exclusivo no fornecimento de *inputs* e ajustamento dos pesos das conexões consoante o *output* desejado para cada caso. A identificação das diferenças entre as frequências sonoras de rochas e minas foi inteiramente operada pela rede, espontaneamente. Na verdade, não havia sequer garantia inicial alguma de que os ecos de minas divergissem de modo sistematicamente

⁵⁸ O processo é inteiramente operado por computador, sem qualquer intervenção humana directa, desde a apresentação dos estímulos à rede até ao cálculo da medida de erro do peso de cada conexão e respectivo ajustamento.

detectável das rochas. O conhecimento dessa distinção não guiou o processo de treino, mas resultou dele.

2.4.2. Espaços de activação

Importa neste ponto justificar a afirmação de que uma rede neuronal artificial treinada incorpora, de facto, *conhecimento*, ou seja, que esta desenvolve efectivamente a capacidade de representar determinadas características do mundo através de uma *rede conceptual*.

Considere-se novamente a rede neuronal de reconhecimento de minas e de rochas submarinas. Conforme referido, cada um dos treze neurónios da camada sensória da rede codifica exclusivamente uma determinada frequência de um padrão de eco dividido, precisamente, em treze frequências. Sendo os pesos das conexões sinápticas e os sinais que as atravessam quantificáveis, essa camada neuronal inicial pode ser geometricamente representada como um (hiper)espaço de treze dimensões – denominado “*espaço de activação*” – cujas fronteiras são abstractamente delimitadas pela totalidade dos vectores de activação *possíveis* para essa camada. Nesse hiperespaço, cada um dos treze neurónios da camada é um *eixo* (ou *dimensão*), e o nível de activação de cada neurónio individual é uma *coordenada*. Logo, o vector de activação resultante da actividade conjunta dos treze neurónios determina um *ponto* específico no interior do espaço abstracto da camada em causa.

Adoptando esta perspectiva, a importância das camadas intermédias no processamento de informação das redes neuronais artificiais torna-se evidente. Após gerado um vector de activação na *primeira* camada de neurónios, cada um emite um sinal para todos os da camada seguinte, sendo o seu valor modificado de acordo com o peso das conexões que atravessa. Uma vez que a camada intermédia, sendo composta por treze neurónios, não possui a capacidade de codificar exhaustivamente as treze frequências originais apresentadas à primeira camada (o que requereria treze neurónios), a classificação das mesmas passa a ter que ser feita na base de relações de *similaridade*. Noutros termos, o facto de a camada intermédia não ter a capacidade de codificar com exactidão os vectores de activação enviados pela camada inicial força-a a gerar vectores de algum modo “semelhantes” a esses que, como esboços, retêm somente as características mais salientes dos originais.

O objectivo do processo de treino é precisamente o de possibilitar a cada camada a aquisição de uma configuração de pesos que lhes permita identificar quais as características mais manifestas dos vectores de activação recebidos das camadas anteriores. Em termos abstractos, a camada intermédia da rede da figura 4, possuindo apenas sete neurónios, tem um espaço de activação correspondente de sete dimensões. Sendo-lhe impossível codificar um vector de activação produzido por uma camada com treze neurónios, o espaço de activação da rede divide-se espontaneamente em subvolumes, ou espaços de similaridade – no caso presente, em dois, consoante o tipo de vector de activação recebido da camada anterior: *semelhante a mina*, ou *semelhante a rocha*.

Terminado o processo de treino, os vectores de activação da camada final (cujo espaço abstracto é bidimensional, uma vez que possui apenas dois neurónios) não poderão deixar de reflectir nos seus *outputs* a bipartição em subvolumes da camada anterior, pelo que os vectores de activação que produzirá serão sempre *aproximadamente* $\{1, 0\}$ ou $\{0, 1\}$. Nesse momento, é legítimo afirmar que a rede incorporou uma rede conceptual simples (de apenas dois conceitos) que aprendeu a utilizar para representar o “mundo” populado exclusivamente por minas e rochas que lhe é dado a conhecer pelos seus “sentidos”.

O maior ou menor grau de aproximação aos valores *ideais* para rocha e mina na camada final depende em cada caso da posição do vector de activação no seu espaço abstracto. Desse modo, vectores próximos dos ideais localizarão o respectivo ponto de activação perto do *centro* do subvolume correspondente ao tipo de estímulo inicial em causa (mina ou rocha), ou seja, esse estímulo será categorizado com um maior grau de fiabilidade. Por esse motivo, a região central de cada subvolume designa-se de *ponto-protótipo* do tipo de estímulo a que corresponde. Parafraseando Churchland, essa região é algo como a “forma platónica” desse estímulo, em torno da qual orbitam imagens imperfeitas tanto mais dissemelhantes da imagem ideal quanto mais periféricas no espaço abstracto. Assim, os casos mais próximos das fronteiras de cada subvolume resultarão de vectores de grande ambiguidade (e.g. $\{6, 4\}$), produzidos por estímulos atípicos.

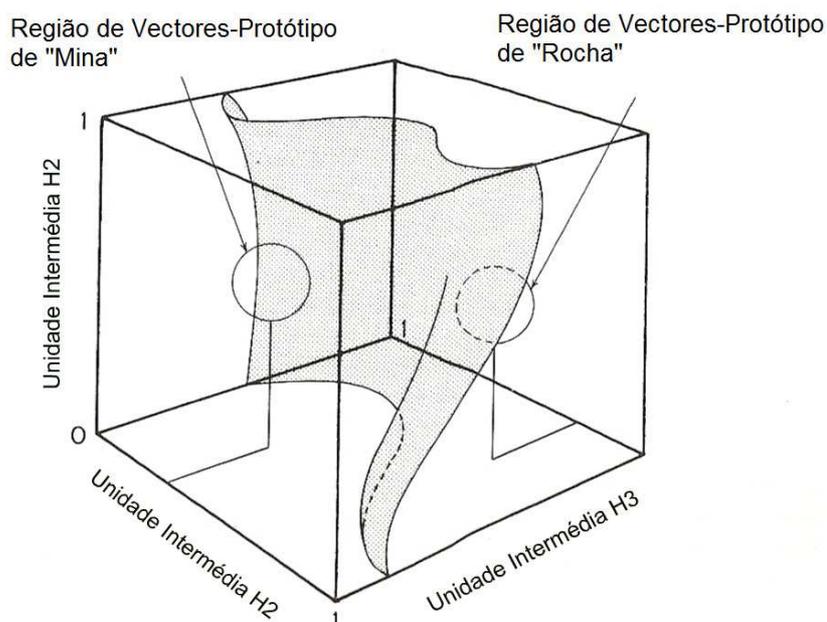


Fig. 5 Divisão em subvolumes do espaço de activação da camada intermédia da rede representada na figura 4, e respectivos pontos-protótipo. Para efeitos de simplificação gráfica, apenas três dos seus sete eixos se encontram aqui representados.⁵⁹

No caso presente, em que a rede em causa discrimina apenas entre dois conceitos, o subvolume encontra-se meramente bipartido. Porém, no caso de redes mais complexas, com uma maior quantidade de neurónios na camada final e necessidade de distinguir entre um maior número de tipos de estímulo, o número de subvolumes no espaço de activação multiplica-se. Nestas redes, os pontos-protótipo de cada conceito podem encontrar-se distribuídos por todo o espaço abstracto, sendo mesmo possível existir subcategorização (por exemplo, uma rede treinada para identificar rostos humanos pode aprender não só a distinguir entre “rosto” e “não-rosto”, mas também entre “rosto masculino” e “rosto feminino”).

Importa notar que o grau de eficiência de uma rede na tarefa de categorizar diferentes tipos de estímulo não é directamente proporcional ao número de camadas neuronais que possui. De facto, é possível que uma rede falhe numa dada tarefa por ter *demasiadas* camadas. Conforme anteriormente referido, a divisão do espaço abstracto de cada camada neuronal em subvolumes só é possível se a informação recebida pela camada intermédia exceder o seu poder representacional, o que a obrigará a *generalizar* a partir dos vectores de activação recebidos, ou seja, a *conceptualizar*. Se a quantidade de neurónios e de camadas intermédias de uma rede for vastamente superior ao

⁵⁹ Ibid., p. 169

requerido dada a quantidade de informação apresentada durante o processo de treino, aquela já não necessitará de discriminar entre *tipos* de estímulo, pois não haverá sobrecarga de informação que a tal a obrigue. Poderá limitar-se a criar uma representação para cada estímulo individual, disseminando a memória da totalidade da informação recebida ao longo das suas múltiplas camadas neuronais, “lembrando-se” com precisão de todo e cada um dos vectores de activação produzidos na camada inicial durante a fase de treino, o que impossibilitará a criação de pontos-protótipo suficientemente indeterminados para constituírem *conceitos*. Como tal, para cada tarefa apresentada a uma rede neuronal existe um número óptimo de neurónios e de camadas neuronais: ‘Fazer uso de recursos limitados é, claro, uma virtude em si, ainda que, é certo, uma virtude meramente pragmática. Mas esta não é a principal virtude aqui demonstrada. A superior capacidade de generalização é uma virtude genuinamente epistémica, e é regularmente demonstrada por redes forçadas [...] a encontrar as hipóteses mais simples relativas a quaisquer estruturas que possam estar escondidas no interior ou por detrás dos seus vectores de *input*’.⁶⁰

Com base na descrição acima do funcionamento dos sistemas de arquitectura computacional conexionista, Churchland formula a seguinte definição de conceito: ‘Ter um conceito é ter a *capacidade* de representar cada um de uma variedade de casos particulares relacionados de modo relevante como pertencendo ao mesmo estreitamente limitado subvolume no espaço de activação, um subvolume que se encontra num conjunto relativamente fixo de relações de distância com um grande número de outros subvolumes. Em termos simplificados, um conceito não é uma imagem, mas um leque isolado e gradativo de imagens *possíveis*. E *ter* um conceito é *controlar* esse leque bem informado de representações possíveis’.⁶¹

Importa referir que as vantagens explicativas que Churchland encontra no modelo conexionista não se restringem ao âmbito da linguagem. Com efeito, Churchland defende que o processo do progressivo “treino” neuronal que nos cérebros humanos conduz à criação de pontos-protótipo em espaços abstractos de activação correspondentes a *conceitos linguísticos* é o mesmo que, *mutatis mutandis*, subjaz à aprendizagem da categorização dos estímulos sensoriais em geral,⁶² da estruturação de

⁶⁰ Ibid., p. 181

⁶¹ Churchland, Paul (2008), “*Neurosemantics: On the Mapping of Minds and the Portrayal of Worlds*”, in *Neurophilosophy at Work*, p. 145

⁶² Por exemplo, no espaço abstracto de activação do sistema neuronal responsável pela percepção cromática, cada ponto-protótipo corresponde a uma cor.

sistemas de relações sociais, e à formação do código moral de cada um. Por este motivo, Churchland não considera o conexionismo meramente um modelo explicativo dos fenómenos *linguísticos*, mas da cognição *tout court* (humana e animal).

2.5. Plausibilidade neuronal do conexionismo

Naturalmente, uma vez que qualquer proposta de explicação dos mecanismos computacionais subjacentes à actividade cognitiva a partir da perspectiva do conexionismo parte do sucesso demonstrado pelas redes neuronais artificiais na execução de tarefas tipicamente operadas por cérebros biológicos, a legitimidade deste paradigma na ciência cognitiva enquanto alternativa real ao paradigma simbólico depende inteiramente da premissa de que os sistemas construídos com base no modelo do processamento paralelo distribuído retêm de facto as características funcionais mais relevantes do sistema nervoso central. Do ponto de vista global não existem dúvidas de que qualquer cérebro é um sistema de processamento de informação composto por uma rede de neurónios divididos em camadas e interconectados por ligações sinápticas com diferentes pesos, mas não é claro até que ponto os sistemas conexionistas simulam fielmente as suas heurísticas computacionais básicas. No que se segue, alguns pontos de relevo da extensamente debatida questão da plausibilidade neuronal dos sistemas conexionistas serão brevemente apresentados. A exposição, contudo, incidirá apenas sobre os aspectos mais fundamentais das redes neuronais artificiais e o modo como se encontram (ou não) espelhados no funcionamento cerebral à *microescala*. Reserva-se para a secção seguinte a discussão desse paralelismo à *macroescala*, especificamente do modo como a formação de redes conceptuais em sistemas conexionistas artificiais poderá espelhar a formação de redes conceptuais em cérebros biológicos.

De modo a que o estrito âmbito do que foi exposto até ao presente momento não seja ultrapassado, e em simultâneo para que a abordagem da questão – não obstante a sua inevitável superficialidade – seja tão abrangente quanto possível, a análise do paralelismo existente entre redes neuronais artificiais e biológicas será feita pelo confronto sucessivo das quatro características fundamentais do modelo de processamento paralelo distribuído anteriormente listadas com os dados relevantes da neurociência contemporânea. Sucintamente, os pontos a abordar serão os seguintes, seguindo a mesma ordem da referida listagem: i) neurónios; ii) conexões; iii) cálculo dos níveis de activação neuronal; e iv) algoritmos de aprendizagem.

Ao nível funcional mais básico – o da unidade computacional/neurónio – nada indica que a muito maior complexidade intrínseca e tipologia⁶³ dos neurónios naturais por comparação com os artificiais acarrete diferenças substanciais do ponto de vista do comportamento global das respectivas redes. As duas principais características de qualquer neurónio são a sua *morfologia* (i.e. genericamente, tripartida em ligação de *input*, corpo e ligação de *output*, da qual dependem as suas propriedades computacionais) e a *função* que desempenha na rede, que varia consoante a camada em que o neurónio em causa se encontra. No que respeita à morfologia, neurónios formais e biológicos não divergem radicalmente nas suas características relevantes. Do mesmo modo, assim como acontece com as redes neuronais artificiais, os neurónios biológicos dividem-se funcionalmente em três tipos, consoante a posição que ocupam na rede: *sensoriais* (receptores de informação exterior situados na periferia do organismo, equivalentes aos neurónios da camada de *input* nas redes de PPD); *interneurónios* (conectados entre si, equivalente aos neurónios das camadas intermédias nas redes de PPD); e *motores* (directamente conectados às fibras musculares, equivalentes aos neurónios da camada de *output* nas redes de PPD). A este nível, portanto, o paralelismo entre redes neuronais biológicas e artificiais é indiscutível.

Ao nível das conexões existem algumas diferenças mas, como no caso anterior, não particularmente problemáticas. Por exemplo, os neurónios naturais estabelecem frequentemente ligações “horizontais” com neurónios da mesma camada, ao passo que nas redes de PPD todas as conexões são projectadas “verticalmente”, de uma camada para a seguinte. Mas também ao nível da conectividade vertical existem diferenças: ao passo que numa rede de PPD o axónio de um dado neurónio projecta ramificações para todos, ou quase todos, os neurónios da camada seguinte, os axónios dos neurónios biológicos tipicamente estabelecem contacto sináptico apenas com uma percentagem relativamente pequena dos neurónios da população neuronal alvo, e apesar de testes realizados com redes neuronais artificiais com conectividade reduzida terem demonstrado que retêm ainda assim um elevado potencial de aprendizagem, verifica-se que há um inegável decréscimo do mesmo por comparação com redes massivamente conectadas. Esta perda de poder cognitivo pode, à primeira vista, constituir uma fragilidade na proposta conexionista, uma vez que o poder computacional dos cérebros reais aparentemente não é truncado por esta limitação. Em todo o caso, é necessário

⁶³ Existem mais de cem tipos de células cerebrais altamente especializadas, ao passo que nas redes de PPD todas as unidades computacionais são estruturalmente idênticas.

manter em mente que um neurónio natural de importância média recebe sinais de aproximadamente 10.000 outros neurónios (e alguns neurónios do cerebelo⁶⁴, cerca de 150.000), sendo que um cérebro humano é composto por aproximadamente 10⁹ neurónios. Logo, é expectável que as limitações nas capacidades de aprendizagem de uma rede artificial (de proporções comparativamente bastante mais modestas) cuja conectividade neuronal média seja apenas *percentualmente* idêntica à de um cérebro humano, não se verifiquem neste último. Assim, o paralelismo artificial/biológico mantém-se sólido também ao nível das conexões.

Relativamente ao cálculo dos níveis de activação neuronal existem diferenças algo mais vincadas entre sistemas naturais e artificiais. Nestes últimos, por exemplo, é essencial que um mesmo axónio projecte ramificações de polaridades tanto excitatória quanto inibitória para os da camada seguinte, e que essas polaridades possam inverter-se no decurso do processo de treino. Sem esta possibilidade, os modelos de PPD revelam-se incapazes de aprender. Porém, as ramificações dos axónios naturais têm uma única polaridade que permanece imutável ao longo de toda a sua existência, sem prejuízo da manifesta capacidade de aprendizagem dos cérebros biológicos. Uma solução possível para esta dificuldade pode advir da vasta conectividade horizontal existente entre neurónios naturais referida acima: um axónio exclusivamente excitatório pode contribuir para inibir um outro indirectamente, excitando um interneurónio inibitório da mesma camada que, por sua vez, inibirá um outro na camada seguinte. Contudo, até à presente data, esta hipótese permanece inteiramente especulativa.⁶⁵

Aquela que é talvez a maior dificuldade enfrentada pelo conexionismo prende-se com o facto de não haver indícios de que o cérebro empregue a regra delta generalizada como algoritmo de aprendizagem. A retropropagação exige que o sistema satisfaça duas condições: i) que possua a capacidade de calcular a medida de erro para cada neurónio na camada de *output*, bem como a correcção a efectuar nos neurónios das camadas anteriores; ii) e que incorpore um método de propagação dessa correcção no sentido inverso do fluxo inicial de sinais neuronais, ajustando os pesos dos neurónios das camadas anteriores individualmente. Dadas estas exigências, dois problemas se apresentam. Em primeiro lugar, não existem indícios de que as fibras axonais descendentes nos cérebros biológicos (que, de facto, existem em quantidade comparável

⁶⁴ Região cerebral ligada ao controlo das funções motoras.

⁶⁵ Churchland, Paul (1992), "On the Nature of Theories: A Neurocomputational Perspective", in *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, p. 184

às ascendentes, como seria expectável numa rede que empregasse a retropropagação) influenciem os pesos dos neurónios das camadas inferiores a que se ligam. A sua influência parece ser directamente no nível de activação dos neurónios em si. Em segundo, para que haja correcção da medida de erro de cada vector de activação na camada de *output* é necessário que o sistema tenha conhecimento prévio de qual seria o vector de activação correcto. Nas redes conexionistas, esta fonte de informação ideal (o “professor” do sistema) é comumente o computador no qual a rede é simulada. Contudo, os cérebros biológicos não possuem uma semelhante garantia externa de certeza, ou seja, ao contrário das redes simuladas, não dispõem do conhecimento do que seria em cada caso o *output* ideal. Ambos estes factores parecem indicar que a regra delta generalizada, não obstante ser o mais eficaz algoritmo de aprendizagem conhecido para as redes conexionistas em geral, não foi evolutivamente incorporado no cérebro biológico.

Uma alternativa possível é a da aprendizagem hebbiana, proposta pelo neuropsicólogo canadiano Donald Hebb⁶⁶ e resumida na formulação conhecida como “postulado de Hebb”: *‘Quando um axónio de uma célula A está suficientemente próximo para excitar a célula B e repetida e persistentemente influencia o seu disparo, algum processo de crescimento ou de mudança metabólica tem lugar numa ou nas duas células de tal modo que a eficiência de A, enquanto uma das células envolvidas no disparo de B, é aumentada’*.⁶⁷ Noutros termos, Hebb propôs que o peso de uma dada conexão num neurónio natural aumentaria em função da frequência do disparo do neurónio emissor. Apesar de menos eficaz do que a regra delta generalizada, este processo de aprendizagem tem duas vantagens enquanto modelo da cognição biológica: não exige a existência de um “professor” onisciente externo ao sistema; e não implica um aumento do tempo de computação proporcional ao tamanho da rede (a regra delta generalizada, sendo um processo *global* de correcção simultânea da medida de erro em cada sinapse individual, implica maior lentidão de aprendizagem a escalas maiores, ao passo que a aprendizagem hebbiana, enquanto processo não-supervisionado, localizado inteiramente ao nível das próprias sinapses, proporciona sempre a mesma rapidez de aprendizagem independentemente das dimensões da rede). Em todo o caso, não obstante este algoritmo se ter revelado eficaz em redes artificiais, também não existem ainda

⁶⁶ Hebb, Donald O. (1949), *The Organization of Behaviour*

⁶⁷ *Ibid.*, p. 62

indícios seguros de que a teoria hebbiana seja uma explicação adequada do processo de aprendizagem dos cérebros biológicos.

Não é certo, portanto, até que ponto as redes neuronais artificiais reflectem fielmente os mecanismos computacionais dos cérebros biológicos. Contudo, seria prematuro inferir desta incerteza a inadequação do paradigma conexionista enquanto modelo explicativo da actividade cognitiva cerebral. Uma atitude mais prudente, sustentada pelo conhecimento dos aspectos estruturais e funcionais em que as redes artificiais são indubitavelmente semelhantes às redes naturais, obriga a assumir a maior plausibilidade da hipótese de que as discrepâncias verificadas actualmente entre ambas sejam efeito da relativa imaturidade dos campos da inteligência artificial e neurociência, bem como da necessidade de incrementar a cooperação interdisciplinar na ciência cognitiva em geral. Esta é, noutras palavras, a expressa convicção de Paul Churchland: ‘Se mesmo redes artificiais pequenas são capazes de executar [...] sofisticadas tarefas cognitivas [...], não é um mistério que as redes reais possam fazer o mesmo ou melhor. O que o cérebro apresenta em termos de *hardware* não é radicalmente diferente do que os modelos contêm, e as diferenças convidam à exploração, não ao desapontamento’.⁶⁸

⁶⁸ Churchland, Paul, *ibid.*, p.187

III

A Contenda Paul Churchland/Fodor & Lepore

A presente secção tratará na íntegra de expor a controvérsia em torno do problema da semântica dos estados mentais que opôs Paul Churchland a Jerry Fodor e Ernest Lepore (F&L) numa sequência de réplicas e trélicas que se prolongou por um período de aproximadamente uma década (1992-2001). O objectivo será o de demonstrar a evolução da teoria semântica de Churchland no decurso da sua sucessiva depuração motivada pelas sucessivas objecções erguidas por F&L de que foi alvo, culminando na proposta que actualmente defende – uma proposta fundada no esteio do conexionismo, localizada, como se explicitará, na conjugação de três eixos teóricos essenciais: o *internalismo*, o *holismo* e o *reducionismo* semânticos.

3.1. Proposta inicial de Churchland

A contenda em questão teve início com a reacção de F&L à teoria psicosemântica avançada por Churchland originalmente na obra *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science* (1991). Esta teoria é resumida por Churchland da seguinte forma: ‘A ideia básica [...] é a de que o cérebro representa diversos aspectos da realidade através de uma *posição* num *espaço de estado*⁶⁹ apropriado, e executa computações sobre tais representações por via de *transformações de coordenadas* de um espaço de estado para outro’.⁷⁰ Como é evidenciado pela terminologia empregue nesta definição, a noção genérica que Churchland pretende defender é a de que o processo de formação de conceitos em cérebros biológicos é análogo ao de formação de pontos-protótipo nos sistemas conexionistas: a “transformação de coordenadas de um espaço de estado para o outro” a que Churchland se refere na citação acima não é mais do que a descrição abstracta da transformação de sinais ao longo de conexões de pesos variáveis em camadas neuronais.

O capítulo da obra referida acima no qual Churchland funda as bases da sua teoria tem como propósito essencial blindar duas teses fundamentais para a mesma com dados

⁶⁹ “Espaço de estado” é a designação genérica atribuída às representações gráficas de sistemas dinâmicos do tipo das que Churchland utiliza para apresentar em abstracto estados de activação de redes neuronais (e.g. fig. 1.5). Assim, no presente contexto, “espaço de estado” e “espaço abstracto de activação” são sinónimos.

⁷⁰ Churchland, Paul, (1992) “Some Reductive Strategies in Cognitive Neurobiology”, in *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, pp. 78-79

recentes provenientes da inteligência artificial e da neurociência. Em primeiro lugar, Churchland pretende evidenciar o ‘impressionante poder *computacional* das transformações de coordenadas de espaços de estado e da possível implementação neuronal dessa actividade’⁷¹; e em segundo, procura demonstrar que uma das principais virtudes dos modelos de espaço de estado é o facto de preservarem nas relações de distância entre pontos-protótipo as relações de *similaridade* existentes entre os itens representados: a dois pontos-protótipo muito próximos num espaço de estado corresponderão dois itens semelhantes no domínio que esse espaço de estado representa.

O exemplo que Churchland utiliza para demonstrar o poder representacional das redes conexionistas e a proficuidade explicativa do modelo dos espaços de estado é o da percepção visual. Considere-se a figura representada abaixo: o “espaço de estado cromático” originalmente proposto por Edwin Land⁷² e reproduzido por Churchland na sua obra.

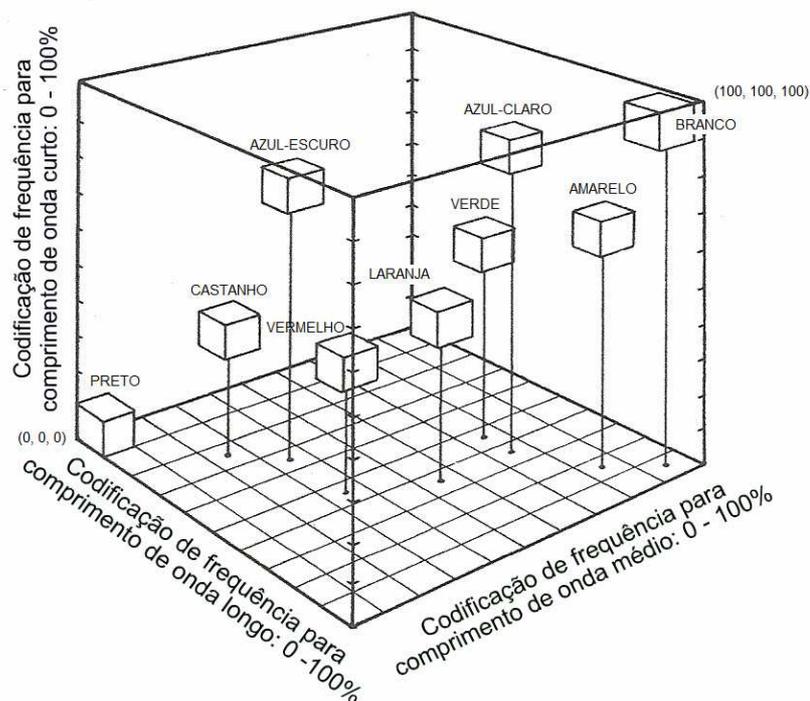


Fig. 6 Espaço de estado cromático⁷³

A totalidade dos vectores de activação que compõem o espaço de estado cromático corresponde ao conjunto total das cores passíveis de serem discriminadas pelo sistema visual humano. Dado que todas as cores humanamente perceptíveis são codificadas

⁷¹ Ibid., p. 102

⁷² Land, Edwin (1977), “The Retinex Theory of Color Vision”, Scientific American, Dezembro, pp. 108-128

⁷³ Churchland, Paul, *ibid.*

pelos *inputs* recebidos por apenas três tipos de células retiniais (ou seja, cada célula do mesmo tipo codifica os *inputs* recebidos com exactamente o mesmo valor de activação), cada eixo/dimensão do espaço de estado cromático corresponde a um *tipo* de célula e não a uma célula em particular, como acontece no caso de redes em que cada célula codifica um dado *input* de modo idiossincrático (e.g. a rede treinada para discriminar entre padrões de sonar de rochas e minas). O espaço *tridimensional* da figura 6, portanto, representa exhaustivamente a sua dimensionalidade.⁷⁴

No espaço de estado cromático, cada coordenada num dos seus eixos/dimensões (determinada pelos valores de activação de células de um mesmo tipo na recepção de um mesmo *input*) constitui a representação ocular/cerebral do padrão luminoso reflectido pelo objecto percebido num de três comprimentos de onda. Cada padrão luminoso, portanto, é redutível a três valores numéricos (correspondentes aos valores de activação de cada um dos três tipos de célula). Tal significa que cada cor percebida pode ser matematicamente descrita como um vector de activação (i.e. o total dos valores de activação produzidos por um dado *input* luminoso ao longo dos três eixos) e, assim, ser identificada com um ponto (coordenada) no espaço de estado cromático. Importando o exemplo ilustrativo que Churchland utiliza: ‘o cor-de-rosa “inefável” da sensação visual presente de alguém poderá ser rica e precisamente expresso como um “acorde de 95Hz/80Hz/80Hz” no sistema cortical triúnico relevante’.⁷⁵

Este modelo, como não deixa de ser enfatizado por Churchland na citação acima, demonstra o poder explicativo do paradigma conexionista ao mesmo tempo que fundamenta as suas convicções naturalistas, na medida em que i) oferece uma metodologia para descrever matematicamente aquilo que à partida mais resistência oferece a qualquer tentativa de quantificação – a experiência fenomenal *qualitativa* de uma cor; e ii) revela uma via pela qual o carácter qualitativo da experiência fenomenal possivelmente poderá transitar do domínio da filosofia (ao qual pertence tradicionalmente) para o da ciência.

Expressando matematicamente os *inputs* luminosos captados na retina e representando os respectivos valores sob a forma de coordenadas específicas no espaço

⁷⁴ Importa atentar ao facto de que a referida possibilidade de divisão das células do sistema visual humano em tipos aplica-se somente à discriminação *cromática*. No que concerne a funções mais complexas como, por exemplo, o reconhecimento de formas (e.g. rostos), o contributo de cada um dos milhões de células retiniais para a formação da representação do objecto visionado é único, pelo que o respectivo espaço de estado terá um número de dimensões igual ao número das células retiniais envolvidas.

⁷⁵ Churchland, Paul, *ibid.*, 106

de estado cromático, as relações “objectivas” de *similaridade* entre cores tornam-se evidentes: ‘Duas cores são muito semelhantes apenas se as suas posições no espaço de estado forem muito próximas. E duas cores são dissemelhantes apenas se as suas posições no espaço de estado forem distantes’.⁷⁶ De acordo com esta noção, o facto de ser consensual que a experiência subjectiva do cor-de-rosa é “mais semelhante” ao vermelho do que ao azul torna-se explicável pelo facto objectivo de haver uma maior proximidade entre os pontos-protótipo correspondentes às duas primeiras cores no espaço de estado cromático (partilhado por todos os humanos com um sistema visual normal) do que entre os correspondentes à primeira e à terceira. *Mutatis mutandis*, o mesmo sucederá com as percepções auditivas, olfactivas, gustativas e tácteis, bem como com funções de reconhecimento de padrões complexas como a identificação de rostos semelhantes.

A capacidade de explicar os fenómenos cognitivos associados à percepção sensorial ‘de um modo biologicamente realista’⁷⁷ evidenciada por este modelo representacional motiva Churchland a lançar as bases da teoria que virá a expandir em várias publicações subsequentes: ‘[...] terá [esta proposta] recursos para explicar as chamadas actividades cognitivas superiores, representadas pelo *uso da linguagem*, por exemplo, e pelo conhecimento proposicional do mundo em geral? É concebível que sim. Poderemos tentar encontrar, por exemplo, uma forma de representar o “espaço linguístico anglófono” de tal modo que todas as frases gramaticais residam numa hiper-superfície dentro desse hiper-espaço, com as relações lógicas entre as mesmas reflectidas enquanto relações espaciais de algum tipo’.⁷⁸

3.2. Primeiro conjunto de críticas de F&L

O foco da crítica de F&L incide precisamente sobre a teoria de uma semântica dos estados mentais baseada na *similaridade de conteúdo*⁷⁹ que Churchland deixa assim esboçada. Não obstante Churchland não ter definido explicitamente os contornos que uma tal teoria assumiria, F&L, com base no texto citado e noutros anteriores, formulam-na este provavelmente a teria formulado: ‘O que Churchland propõe é a ideia de que

⁷⁶ Ibid., p. 103

⁷⁷ Ibid., 108

⁷⁸ Ibid., p. 109

⁷⁹ Em oposição a teorias anteriores de *identidade* de conteúdo, de acordo com as quais dois conceitos são o mesmo só e apenas se forem exactamente *idênticos* do ponto de vista semântico.

dois conceitos são semelhantes desde que ocupem posições (relativamente) semelhantes *no mesmo espaço de estado*.⁸⁰

3.2.1. O problema da individuação de dimensões

A primeira objecção que erguem consiste no designado problema da *individuação de dimensões*. Uma teoria semântica baseada em espaços de estado exige que as dimensões do mesmo sejam dimensões *semânticas*, uma vez que serão as coordenadas ao longo das mesmas que ditarão o conteúdo de um conceito. Logo, a questão imediatamente que se coloca é a de como estabelecer um critério que permita identificar dimensões em espaços de estado distintos como sendo *uma mesma dimensão semântica*, condição necessária para que se possa dizer que um mesmo conceito é partilhado por dois espaços de estado distintos (i.e. dois indivíduos diferentes).

Uma solução possível seria adaptar a proposta da similaridade semântica ao empirismo clássico, de acordo com o qual os conceitos se dividem em dois tipos, consoante o modo como obtêm o seu conteúdo semântico: os conceitos na “periferia” da rede conceptual, designados de *conceitos observacionais* (sintéticos), derivam o seu conteúdo semântico das propriedades observáveis do mundo que expressam directamente, ao passo que os *conceitos abstractos* (analíticos, não-observacionais) derivam-no da posição em que se encontram na rede relativamente a todos os outros conceitos abstractos e aos conceitos observacionais aos quais se encontram sempre ancorados por um número maior ou menor de passos inferenciais.⁸¹

F&L e Churchland concordam que esta forma de empirismo foi desacreditada, na medida em que se o conteúdo semântico de um conceito for determinado pela posição única que ocupa na estrutura da rede *como um todo*, apenas redes *idênticas* poderão partilhar um mesmo conceito, e uma vez aceite a demonstração quineana da inexistência de distinção entre juízos analíticos e sintéticos, este critério torna-se impossível de

⁸⁰ F&L (1992), “Paul Churchland: State Space Semantics”, in *Holism: A Shopper's Guide*, p. 197

⁸¹ Este tipo de empirismo designa-se de *externalismo semântico*, na medida em que funda o conteúdo de todos os conceitos de uma linguagem directa ou indirectamente na relação do vocabulário observacional com propriedades externas do mundo. Uma vez que nesta perspectiva os conceitos abstractos derivam o seu conteúdo da sua posição relativamente a *todos os outros*, esta posição é também *holista*. Contudo, é-o apenas parcialmente na medida em que esses conceitos abstractos continuam a estar indirectamente ligados aos conceitos observacionais e, logo, a propriedades externas. As teorias integralmente holistas (como a que Churchland defenderá num momento mais tardio), isto é, que não supõem qualquer ligação causal entre os conceitos e o mundo exterior e, como tal, não distinguem fundamentalmente entre conceitos abstractos e observacionais, designam-se de *internalistas*.

satisfazer. F&L, contudo, consideram a hipótese de Churchland contornar o “obstáculo quineano” optando por defender uma perspectiva igualmente holista embora baseada na noção de similaridade e não já na de identidade, de acordo com a qual conceitos abstractos em redes distintas poderiam ser considerados o mesmo se ocupassem lugares apenas *semelhantes* nas respectivas redes, e não já exactamente *idênticos*. Em simultâneo, Churchland poderia atribuir a determinadas dimensões (por exemplo, ligadas aos espaços de estado dos órgãos sensoriais) o papel de primitivos semânticos de tal modo que o conteúdo de todos os conceitos não directamente ligados a propriedades observáveis do mundo estaria sempre indirectamente radicado nestas “dimensões sensoriais” da rede, equivalentes no papel que desempenhariam ao vocabulário observacional do empirismo clássico. Deste modo, Churchland reteria na sua teoria todas as vantagens do empirismo (em particular, o critério externalista de fixação de conteúdo semântico), evitando os seus problemas tradicionais através do critério de similaridade semântica possibilitado pelo modelo conexionista.

Todavia, apesar da sua aparente robustez teórica, uma resposta empirista claramente não solucionaria o problema de determinar o conteúdo semântico de representações abstractas. Como F&L apontam, ‘simplesmente não parece verdade que as dimensões de conteúdo ao longo das quais as palavras (/conceitos) possam ser similares sejam reduzíveis às várias maneiras como possam estar ligadas a observáveis’⁸². Importando o exemplo que utilizam, não parece possível que o conjunto de propriedades semânticas partilhadas entre os conceitos “tio” e “tia” (quaisquer que sejam) e que os tornam semelhantes possa ser expresso pelas propriedades observáveis de tios e de tias.

Em todo o caso, Churchland sempre se negou explicitamente a adoptar o empirismo em qualquer das suas formas (ainda que F&L considerem que de certo modo o supõe implicitamente): ‘Linguagens, e a rede de crenças que incorporam, têm uma identidade que transcende e pode permanecer constante não obstante variações nos canais sensoriais aos quais possam estar ligadas, e nas ligações particulares dentro da linguagem onde as ligações sensoriais possam estar estabelecidas’⁸³. A teoria da similaridade semântica de Churchland, assim, funda-se inteiramente na noção de que dois espaços de estado partilharão um mesmo conceito se esses espaços forem holisticamente semelhantes, sem recurso a propriedades observacionais. Logo, e por

⁸² F&L., *ibid.*, p. 191

⁸³ Churchland, Paul (1979), *Scientific Realism and the Plasticity of Mind*, p. 65

oposição ao empirismo clássico, a teoria que propõe é *internalista*, ou seja, as dimensões de um espaço de estado podem corresponder a qualquer tipo de propriedade representável pelo cérebro sem distinção (observável ou abstracta). Ora, segundo F&L, ao procurar evitar os problemas do empirismo, recusando-o, Churchland ficou desprovido de um método que possibilite determinar se dois espaços de estado são de facto o mesmo (pois deixa de haver um referente que permita estabelecer uma correspondência entre as suas respectivas dimensões) e, *a fortiori*, tornou impossível determinar se dois espaços de estado partilham ou não um mesmo conceito.

F&L rejeitam igualmente o contra-argumento possível de que não é necessário que conceitos ocupem lugares semelhantes em espaços de estado idênticos mas apenas em espaços de estado também eles apenas semelhantes. Esta resposta, dizem, implicaria um eterno retorno, pois não é possível determinar a similaridade entre quaisquer dois itens em geral sem que se especifique a propriedade *fixa* em relação à qual essa semelhança é avaliada, o que se aplica sem distinção a conceitos e a espaços de estado.

Por fim, F&L afirmam que ‘possuímos uma noção robusta de similaridade semântica apenas se tivermos um critério para a identidade de espaços de estado’ e ‘possuímos um critério para a identidade de estados de espaço apenas se possuímos um critério para a identidade das dimensões dos espaços de estado’⁸⁴. Ora, uma vez que qualquer critério de identidade supõe sempre a inaceitável distinção analítico/sintético, F&L concluem que a teoria semântica baseada na similaridade de conteúdo proposta por Churchland na verdade não se distingue fundamentalmente das teorias de identidade das quais se procura distanciar, dado que se limita a transpor o problema quineano da identidade do domínio *conceptual* para o *dimensional*.

3.2.2. O problema da relevância semântica

O problema seguinte identificado por F&L é o da *relevância semântica* das dimensões. Considere-se o espaço de estado cromático representado na figura 6. As células retiniais envolvidas no reconhecimento cromático são, como vimos, de três tipos (especializadas na recepção de *inputs* luminosos em comprimentos de onda curto, médio e longo). Assim, as dimensões que compõem o respectivo espaço de estado são, concomitantemente, somente três, dado que o contributo de cada célula retinal (i.e. o seu valor de activação) na codificação do vector de activação correspondente a uma

⁸⁴ F&L, *ibid.*, p. 198

dada cor é idêntico ao de qualquer outra de um tipo funcionalmente idêntico ao seu. Neste caso, assim como em outros semelhantes em que o número de dimensões do espaço de estado em causa é reduzido e inequívoco, a determinação das dimensões que são relevantes para a comparação de espaços de estado distintos não oferece dificuldades (por exemplo, no caso do espaço cromático, *todas* são igualmente relevantes para a determinação da cor a que corresponde um dado *input* luminoso).

Contudo, nos espaços em que o número de dimensões é exponencialmente mais vasto, afirmam F&L, nem todas as dimensões terão a mesma relevância na determinação do conteúdo semântico de um conceito. Ao contrário da representação da cor “vermelho” pelas células retiniais, na qual cada valor ao longo de três dimensões do espaço de estado cromático é indispensável, o conceito de “cão”, por exemplo, será codificado por valores em dimensões que em princípio não serão igualmente fundamentais (por exemplo, ‘o espaço no qual o conceito “cão” reside deve possuir uma dimensão correspondente à propriedade de ser um animal, mas não precisa de ter uma dimensão correspondente à propriedade de tipicamente ter o nariz molhado’⁸⁵). Exige-se, portanto, um critério que permita distinguir que dimensões pertencem *necessariamente* a um conceito e quais lhe pertencem apenas *contingente* ou *secundariamente*, de modo a que seja possível identificar exactamente que dimensões têm que ser partilhadas por dois espaços de estado para que seja possível determinar se são ou não idênticos e, logo, se partilham ou não um mesmo conceito. Mas como F&L referem, esta última exigência suscita imediatamente o mesmo problema quineano de anteriormente: ‘Se estivermos convencidos de que não existe forma de estabelecer uma distinção analítico/sintético, a que princípio recorreremos de modo a distinguir as dimensões que são relevantes das que não o são?’⁸⁶ Novamente, com esta objecção, F&L enfatizam a afirmação de que a proposta de Churchland não traz qualquer vantagem explicativa sobre as teorias semânticas de identidade, preservando todas as dificuldades que estas acarretam.

3.2.3. O problema da informação colateral

A terceira objecção de F&L, que designam de problema da *informação colateral*, procede directamente da anterior. Segundo argumentam, ainda que fosse possível obviar

⁸⁵ F&L, *ibid.*, 202

⁸⁶ *Ibid.*

ao problema da distinção analítico/sintético e determinar exactamente que dimensões semânticas devem ser partilhadas por dois espaços de estado para que possam ser considerados idênticos, ainda assim seria impossível determinar a partilha de um mesmo conceito por ambos os espaços. De acordo com a proposta de Churchland (tal como F&L a interpretam), dois espaços de estado semânticos partilham um mesmo conceito se esse conceito ocupar uma posição semelhante nos mesmos. A dificuldade que F&L encontram nesta noção é que ainda que se admitisse a possibilidade de determinar o isomorfismo dimensional de ambos os espaços, a posição de, por exemplo, o conceito “cão” em cada um provavelmente seria muito diferente, pelo facto de que ‘muito do que qualquer pessoa sabe sobre cães é idiossincrático; é “informação colateral”’.⁸⁷ Noutros termos, dado que o conteúdo semântico de um conceito reflecte em grande medida a biografia de quem o detém (e.g. o conteúdo semântico do conceito “cão” não será o mesmo para um veterinário que será para um leigo), a posição de um mesmo conceito em espaços de estado distintos, ainda que estes partilhem exactamente as mesmas dimensões, variará de acordo com a experiência subjectiva que em cada caso conduziu à formação do conceito. Numa palavra, ao recusar o critério externalista do empirismo, Churchland evitou os problemas que este acarreta, mas por não ter proposto uma alternativa ficou sem uma matriz objectiva que constrinja espaços de estado idênticos a situar um mesmo conceito em posições similares. Sem um tal critério, nada obsta a que a posição de um mesmo conceito em espaços de estado distintos varie drasticamente e que, como tal, o torne incomparável a outro num outro espaço de estado.

Uma forma possível de responder a esta objecção sem recurso ao empirismo seria afirmar que dois conceitos são idênticos em dois espaços de estado distintos somente se ocuparem posições semelhantes nos mesmos ao longo de dimensões que não variem idiossincraticamente, isto é, que sejam absolutamente necessárias na determinação do conteúdo semântico do conceito em causa. Este argumento, todavia, redundaria no problema anterior da relevância semântica, uma vez que supõe a existência de um critério que permitisse a distinção entre dimensões primárias e secundárias e um tal critério, qualquer que fosse, não poderia deixar de invocar a distinção analítico/sintético.

⁸⁷ F&L, *ibid.*, 203

3.3. Segunda versão da proposta de Churchland

Na sua tréplica,⁸⁸ Churchland escusa-se a responder à argumentação de F&L ponto por ponto. Ao invés, procura aprofundar a teoria semântica que inicialmente deixara apenas esboçada e assim demonstrar que todo o conjunto das objecções erguido por F&L carece de fundamento por partir de uma concepção errónea do modo como se processa a formação de representações nas redes neuronais. Num primeiro momento, Churchland concorda com a argumentação de F&L, reconhecendo que ‘a identidade de conteúdo torna-se exponencialmente mais difícil de alcançar quando a dimensionalidade das representações atinge os milhões, e quando não há uma correspondência de um-para-um nas populações neuronais entre indivíduos’.⁸⁹ Essa correspondência de um-para-um, como vimos, é estabelecida sem dificuldades no caso de redes em que os neurónios possam ser divididos em *tipos* e que por isso formem dimensões *colectivamente* (como acontece com as células retiniais envolvidas na codificação de cores, como exposto no exemplo anterior), o que implica que independentemente do número de neurónios que constitua a rede em causa, o respectivo espaço de estado terá sempre um número reduzido de dimensões, facilmente identificáveis. O mesmo, porém, não acontece no caso de espaços de estado multidimensionais. Churchland, portanto, concede que não existe um método que permita determinar o isomorfismo de espaços de estado de redes que possuam uma população neuronal na ordem dos milhões, cujos neurónios contribuam idiossincriticamente para a formação das respectivas representações. Ao fazê-lo, não recua na sua rejeição do empirismo (que na sua teoria, corresponderia à sugestão de que todo o conteúdo semântico se funda em ‘combinações vectoriais de significados básicos, significados básicos esses que seriam fixados pelas sensibilidades causais dos neurónios sensoriais do indivíduo’⁹⁰), ao mesmo tempo que reafirma a posição internalista de acordo com a qual o significado dos termos observacionais não é determinado pela ligação causal dos mesmos ao mundo exterior, de tal modo que ‘duas criaturas poderiam partilhar essencialmente a mesma rede conceptual mesmo no caso extremo em que não tivessem quaisquer órgãos sensoriais em comum’.⁹¹

⁸⁸ Churchland, Paul, (1996) “Fodor and Lepore: State-Space Semantics and Meaning Holism”, in *The Churchlands and their Critics*, pp. 272-277

⁸⁹ Ibid., p. 273

⁹⁰ Ibid., 274

⁹¹ Ibid.

Como fica claro por esta linha de argumentação, Churchland não faz qualquer tentativa de formular o critério de individuação de dimensões que F&L consideram vital para que a proposta semântica conexionista possa reclamar legitimidade. Parece, aliás, aprofundar ainda mais o problema inicial ao afirmar que a impossibilidade de se encontrar um método que possibilite a identificação de espaços de estado dimensionalmente isomórficos não é consequência de não haver um método seguro de individuação de dimensões em alternativa ao externalismo semântico, e sim do facto de esse isomorfismo entre espaços de estado multidimensionais simplesmente *não existir*. Esta aparente concessão, todavia, serve meramente o propósito de demonstrar que, contrariamente à argumentação de F&L, o isomorfismo dimensional não é uma condição necessária para que exista similaridade conceptual entre espaços de estado.

3.4. Segundo conjunto de críticas de F&L

A versão modificada da proposta original de Churchland em resposta à primeira série de críticas de F&L é notoriamente frágil e insatisfatória. Sucintamente, Churchland propõe que quando um cérebro humano forma a região-protótipo de um dado conceito num espaço de estado apropriado (mediante um processo de “treino” análogo ao que foi descrito relativamente à rede neuronal artificial treinada para a discriminação de estímulos correspondentes a rochas e a minas), os vectores de activação que codificam posições no espaço de estado internas a essa região-protótipo – ou seja, vectores que codificam o *mesmo* conceito –, originam um conjunto de ‘sequelas causais e computacionais’⁹² que produzem no indivíduo em causa expectativas perceptivas e comportamentos *próprios desse conceito*. Por outras palavras, de acordo com esta perspectiva, as regiões-protótipo (conceitos) são determinadas não já pela sua posição no espaço de estado que ocupam, como era o caso na proposta original de Churchland, mas pelo modo como a activação dos vectores que lhes correspondem se repercute na ‘economia cognitiva e comportamento motor da criatura em questão’⁹³. Se indivíduos de culturas distintas tiverem comportamentos e expectativas perceptivas completamente diferentes quando confrontados com um mesmo objecto, os seus respectivos conceitos desse objecto serão também diferentes. Conversamente, se duas criaturas cognitivas possivelmente com órgãos sensoriais totalmente distintos tiverem

⁹² Ibid., 276

⁹³ Ibid.

expectativas e comportamentos semelhantes perante um dado objecto, ambas terão o mesmo conceito desse objecto. Numa palavra, o conteúdo semântico dos conceitos passa a depender do seu *papel funcional*.

Apesar de esta nova versão da proposta de Churchland obviar *prima facie* às críticas de F&L (não por lhes dar resposta mas por lhes retirar a pertinência, eliminando a importância da posição das regiões-protótipo nos espaços de estado para a determinação do conteúdo semântico dos conceitos), verifica-se que na verdade não traz benefícios explicativos relativamente à teoria original.

Como foi referido, na sua argumentação crítica inicial, F&L enfatizaram que a teoria da similaridade semântica de Churchland não evita o problema quineano da distinção analítico/sintética implicado nas teorias semânticas da identidade, na medida em que deixa de requerer um critério para a identificação de *conceitos* entre espaços de estado (basta que sejam similares) mas continua a exigí-lo para as suas *dimensões*. Churchland, ao argumentar que o conteúdo semântico é determinado pelo papel *funcional* dos conceitos e não pela posição relativa dos mesmos nos espaços de estado em que se inserem, contornou de certo modo a objecção. Não obstante, F&L afirmam⁹⁴ que tudo o que Churchland conseguiu com este desvio de curso da sua proposta original foi redireccionar o problema quineano uma vez mais, desta feita do âmbito da dimensionalidade para o da *funcionalidade* conceptual. Ao sugerir que a similaridade conceptual não depende já da posição de regiões-protótipo em espaços de estado distintos e sim da similaridade *funcional* dos mesmos, ou seja, da putativa semelhança entre os eventos mentais e motores despoletados por um mesmo conceito em indivíduos que o partilhem, Churchland, como se explicitará abaixo, permaneceu vulnerável a todas as três objecções originais de F&L; e uma vez que todas radicavam no problema de que qualquer tentativa de solucioná-las invocaria sempre necessariamente a distinção analítico/sintético, essa mesma crítica, *mutatis mutandis*, mantém a sua pertinência relativamente à nova versão da teoria.

Assim, no lugar do problema original da individuação de dimensões, a proposta de Churchland suscita agora a questão de como individuar os papéis funcionais a partir dos quais se poderá avaliar os comportamentos/expectativas de dois indivíduos de modo a determinar se ambos partilham ou não o mesmo conceito. Se é impossível que o confronto com o mesmo objecto origine sempre previsivelmente

⁹⁴ F&L, (1996) "Reply to Churchland", in *The Churchlands and their Critics*, pp. 159-162

comportamentos/expectativas idênticos em dois indivíduos, qual o critério que permitirá determinar se os seus comportamentos/expectativas são de facto similares? Do mesmo modo, mantém-se o problema da relevância semântica, uma vez que Churchland não sugere método algum que permita determinar que papéis funcionais dos conceitos são relevantes na determinação da similaridade intersubjectiva e quais não são; e, por fim, o problema da informação colateral, porquanto mesmo admitindo a possibilidade de se especificar que papéis funcionais são relevantes para a determinação do conteúdo semântico de um conceito, os mesmos inevitavelmente teriam expressões muito diversas em indivíduos diferentes.

3.5. Versão final da proposta de Churchland

Subsequentemente, Churchland propôs uma teoria semântica bastante mais robusta que, ao contrário da anterior, procura responder directamente aos problemas colocados por F&L. Dado que esta proposta se manteve fundamentalmente inalterada desde que foi inicialmente esboçada⁹⁵ até aos dias de hoje, exceptuando nos pontos em que foi expandida e aprofundada, para fins de facilitação da exposição será aqui apresentada na sua forma final, isto é, nos moldes em que Churchland a concebe actualmente, ignorando-se para os presentes efeitos o processo de maturação que atravessou ao longo dos anos.

A nova versão da teoria semântica de Churchland pretende ‘defender um critério para a *identidade de significado*, e uma medida para a *similaridade de significado*, entre redes conceptuais distintas, um critério que não faz apelo a quaisquer ligações causais ou informacionais que os conceitos em causa, ou as suas dimensões neuronais idiossincráticas, possam ter com propriedades no ambiente circundante’.⁹⁶ Com esta proposta, portanto, Churchland recupera ao seu propósito original de desenvolver uma teoria semântica inteiramente *internalista*, tomando como ponto de partida a constatação de que não existe isomorfismo dimensional entre espaços de estado de indivíduos distintos: ‘[...] somos todos idiossincráticos no número de células que temos numa dada população neuronal. Os nossos respectivos espaços de activação, logo, não têm a mesma dimensionalidade [...]. Do mesmo modo, se não podemos fazer sentido da

⁹⁵ Churchland, Paul (1998), “*Conceptual Similarity across Sensory and Neural Diversity: The Fodor-Lepore Challenge Answered*”, in *On the Contrary*, pp. 81-112

⁹⁶ Churchland, Paul (2001), “*Neurosemantics: On the Mapping of Minds and the Portrayal of Worlds*”, in *Neurophilosophy at Work*, pp. 126-127

noção de “mesmo espaço de activação” entre indivíduos distintos, é óbvio que não poderemos fazer sentido da noção de “mesma posição em tal espaço” entre dois indivíduos, ou mesmo de “posições aproximadas”⁹⁷.

A nova teoria de Churchland baseia-se no procedimento matemático desenvolvido por Laakso e Cottrell⁹⁸ para quantificar similaridade em redes que não partilhem o mesmo número de neurónios ou de camadas neuronais. De acordo com este modelo, a similaridade entre pontos em espaços de estado distintos não é determinada pelas suas posições individuais nos mesmos, pela sua ligação a um referente externo comum, ou por qualquer critério que dependa de alguma forma de um critério de individuação de dimensões. É, sim, determinada pela *similaridade geométrica dos sólidos configurados holisticamente por todos os pontos protótipos em espaços de estado distintos*. No que se segue explicitar-se-á este procedimento em maior detalhe. Considere-se o exemplo ilustrado na figura abaixo:

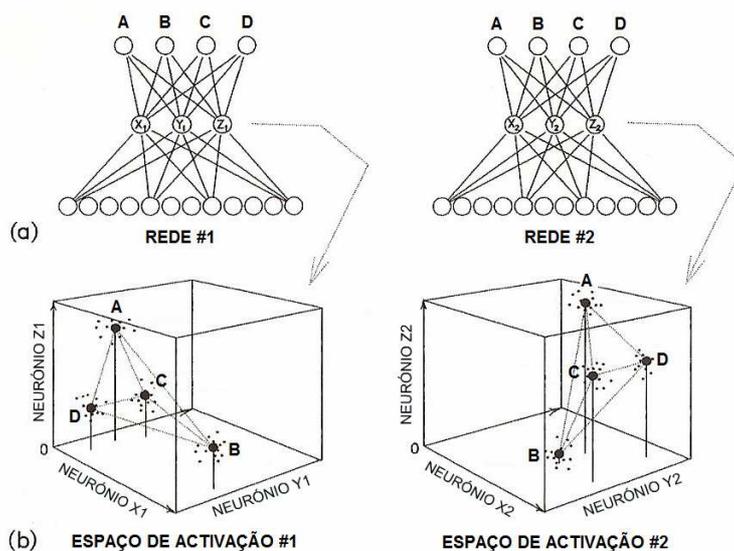


Fig. 7 a) Duas redes treinadas para discriminar fotografias de rostos como pertencendo a uma de quatro famílias (famílias A, B, C, D); b) os espaços de activação das respectivas camadas intermédias das duas redes.⁹⁹

⁹⁷ Churchland, Paul (1998), “*Conceptual Similarity across Sensory and Neural Diversity: The Fodor-Lepore Challenge Answered*”, in *On the Contrary*, p. 84

⁹⁸ Laakso, A., & Cottrell, G. (1998), “*How can I know what you think?: Assessing representational similarity in neural systems*”, Proceedings of the 20th Annual Cognitive Science Conference; Idem (2000), “*Content and cluster analysis: Assessing representational similarity in neural systems*”, in *Philosophical psychology*, 13, pp. 47-76

⁹⁹ Churchland, Paul (2001), “*Neurosemantics: On the Mapping of Minds and the Portrayal of Worlds*”, in *Neurophilosophy at Work*, p. 128

Como referido na legenda, cada ponto-protótipo nos espaços de activação em b) corresponde genericamente a uma família. Cada um desses pontos, por seu turno, é circundado por uma constelação de pontos menores correspondentes aos rostos dos membros da família em questão. Assim, uma dada imagem de *input* será interpretada pela rede como sendo o rosto de um membro de uma das quatro famílias se originar um vector de activação que codifique um vector próximo de um dos quatro pontos-protótipo do espaço. Ora, como se vê, não obstante ambos os espaços de activação partilharem os pontos-protótipo de todas as quatro famílias, cada ponto ocupa individualmente um lugar completamente díspar daquele que o seu homólogo ocupa no outro espaço. A similaridade semântica com base no critério da posição dos pontos-protótipo relativamente às *dimensões* do espaço de activação, portanto, não é viável. No entanto, verifica-se que apesar das suas posições variarem, as *distâncias* relativas entre os pontos-protótipo são preservadas em ambas as redes. Logo, se uma linha imaginária for traçada que una entre si os pontos-protótipo de cada espaço de estado, estes formarão colectivamente dois *sólidos isométricos*. Com base neste facto, Churchland propõe um método de verificação da similaridade de conceitos inteiramente baseado nas propriedades geométricas dos referidos sólidos.

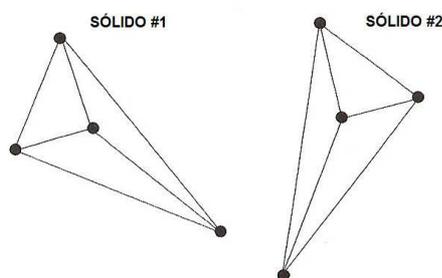


Fig. 8 As configurações dos pontos-protótipo faciais para cada uma das redes treinadas.¹⁰⁰

O primeiro passo desse método será o de extrair os sólidos resultantes do referido traçado de linhas entre pontos-protótipo dos seus respectivos espaços de activação. Note-se que na sua demonstração deste procedimento (reproduzida na figura 8), Churchland optou por retirar a cada ponto-protótipo a referência da letra que identificava a sua família correspondente, de modo a enfatizar o facto fundamental de que o conhecimento do *conceito* individual a que cada ponto-protótipo corresponde é irrelevante para determinar se ambas as redes o partilham. Deste procedimento inicial, então, como ilustrado na figura 8, resultam dois sólidos irregulares desprovidos de

¹⁰⁰ Ibid., p. 130

ligação a qualquer referência externa que permita determinar o conteúdo semântico que os seus vértices possam representar.

O passo seguinte consiste em sobrepor o lado mais longo de ambos os sólidos na orientação que permita encontrar pontos díspares *coplanares*, conforme demonstrado na figura 9.

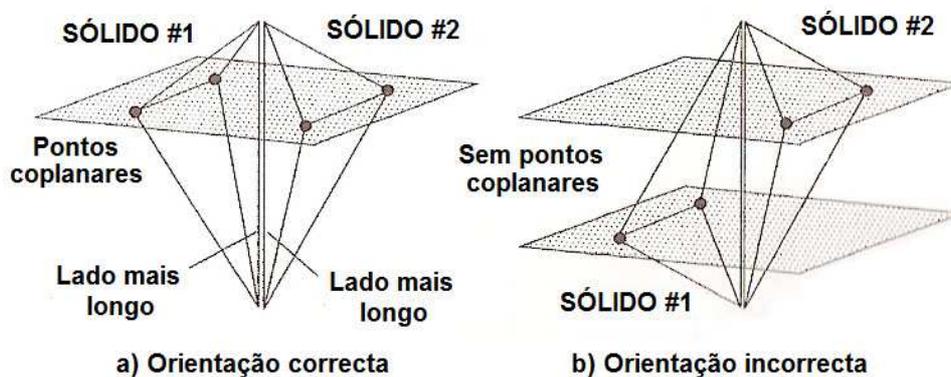


Fig. 9 Dois possíveis alinhamentos dos dois sólidos em questão, que sobrepõem o lado mais longo de cada um.¹⁰¹

Por último, devem rodar-se um dos sólidos ao longo do eixo do lado comum sobreposto, como indicado na figura 10, de modo a que a distância entre os pontos coplanares se reduza a um valor próximo do zero. Como afirma Churchland, ‘uma vez que ambos os sólidos são metricamente idênticos, deverá existir pelo menos uma orientação mútua que produza congruência. E uma vez que os sólidos são irregulares [...], deverá existir no máximo uma tal orientação mútua’.¹⁰²

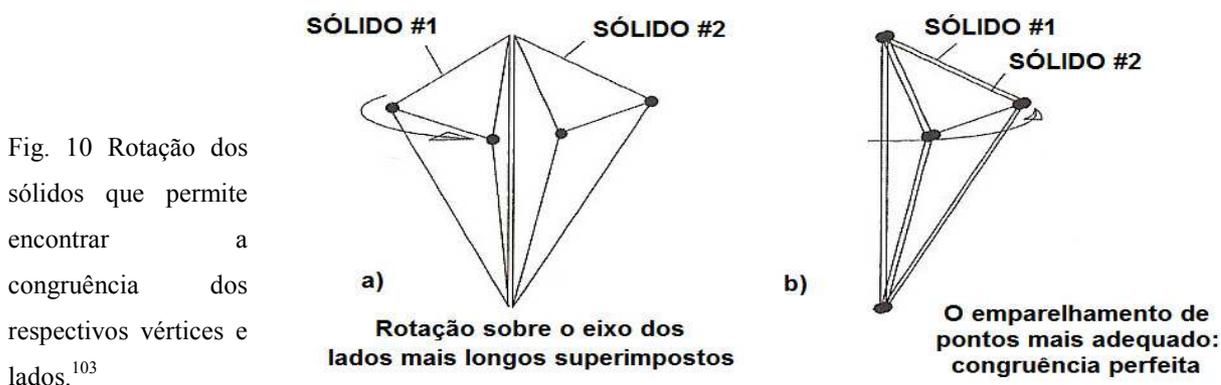


Fig. 10 Rotação dos sólidos que permite encontrar a congruência dos respectivos vértices e lados.¹⁰³

¹⁰¹ Ibid.
¹⁰² Ibid., p. 131
¹⁰³ Ibid.

No final deste processo, os pontos-protótipo que se encontrarem sobrepostos (ou muito próximos) serão *semanticamente idênticos*, isto é, corresponderão ao mesmo conceito.

Não obstante o facto de a introdução desta metodologia tornar a versão final da proposta semântica de Churchland notoriamente mais robusta do que ambas as anteriores – nomeadamente por possibilitar a determinação da similaridade conceptual entre espaços de estado distintos sem implicar qualquer critério de identidade –, o internalismo radical que a caracteriza deixa em aberto duas questões incontornáveis, intimamente ligadas:

a) Admitindo que não existe uma relação *causal* entre a posição de um ponto-protótipo/conceito num espaço de estado e o seu referente externo, como justificar que redes conceptuais de indivíduos diferentes formem sólidos congruentes, se conceitos com um mesmo referente podem ocupar posições completamente díspares em espaços de estado distintos?

b) Como pode uma teoria radicalmente internalista preservar a intencionalidade dos conceitos? Noutros termos, parafraseando Churchland, como é possível que uma teoria que nega a existência de uma relação causal entre os conceitos e os seus referentes preserve a capacidade tradicionalmente considerada essencial aos conceitos que é a de “apontarem para fora de si mesmos”, isto é, de constituírem representações de propriedades do mundo exterior?

A resposta de Churchland,¹⁰⁴ dada sob a forma de uma analogia, é a mesma para ambas as questões: as redes conceptuais representam o mundo como um “*mapa*” das suas características mais salientes. Um mesmo domínio de elementos do mundo pode ser representado/mapeado com igual sucesso de um grande número de formas diferentes (e.g. fotografias, pinturas, plantas arquitecturais, hologramas, etc.). O que determinará se constituem ou não representações do *mesmo* domínio será o facto de preservarem ou não internamente as relações de distância significantes existentes entre os elementos constituintes desse domínio. Um mapa rodoviário, por exemplo, será uma representação

¹⁰⁴ Note-se que estas questões não constituem objecções erguidas por F&L à proposta de Churchland, mas problemas cuja pertinência Churchland reconhece antecipadamente e que procura resolver no seguimento da exposição da sua teoria, aprofundando a especificidade do internalismo semântico que esta supõe.

correcta do seu domínio-alvo se espelhar fielmente nas distâncias dos seus elementos gráficos as distâncias entre as características relevantes da estrada real (entendendo-se por características relevantes, neste caso, aquelas cujo conhecimento seria necessário para uma navegação bem sucedida no percurso da estrada). Ora, dado que os pontos-protótipo emergem da exposição repetida de uma rede neuronal a características salientes *fixas* do ambiente externo, redes neuronais distintas preservarão sempre as relações de distância existentes entre essas características na configuração dos pontos-protótipo que lhes correspondem pois independentemente do modo como recebem os *inputs* exteriores (i.e. do seu aparato sensorial), os padrões estatísticos salientes contidos nesses *inputs* serão idênticos para ambas as redes. Logo, redes neuronais expostas a um mesmo domínio de elementos externos produzirão sempre de modo previsível famílias de pontos-protótipo que, como um todo, serão geometricamente congruentes, *ainda que possuam um número diferente de dimensões/neurónios*: ‘O ponto mais importante a reter é que a componente básica desta medida de similaridade – a *distância* – tem as mesmas unidades de medição qualquer que seja a dimensionalidade do espaço no qual essa distância é medida’.¹⁰⁵

A analogia estabelecida por Churchland entre espaços de estado e mapas permite igualmente responder à questão b), na medida em que ‘os mapas podem ser identificados como este, aquele ou aqueloutro tipo de mapa independentemente de qualquer ligações causais que possam (ou não) ter com o mundo externo’¹⁰⁶. Assim, o radical internalismo da teoria de Churchland não constitui obstáculo à preservação da intencionalidade conceptual na medida em que a intencionalidade, de acordo com esta proposta, não é já definida como a propriedade dos conceitos “apontarem para fora de si mesmos”, mas como a capacidade de uma rede conceptual como um todo representar correctamente as relações de distância relevantes entre os elementos do domínio representado. A relação causal com o mundo externo é relevante apenas no momento inicial do processo de aprendizagem da rede neuronal, na medida em que são necessários *inputs* sensoriais para que uma rede neuronal adquira a configuração sináptica necessária à formação em abstracto dos pontos-protótipo correspondentes aos padrões salientes inclusos nesses *inputs*. Uma vez obtida essa configuração, a rede passa

¹⁰⁵ Churchland, Paul (1998), “*Conceptual Similarity across Sensory and Neural Diversity: The Fodor-Lepore Challenge Answered*”, in *On the contrary*, p. 89

¹⁰⁶ Churchland, Paul (2001), “*Neurosemantics: On the Mapping of Minds and the Portrayal of Worlds*”, in *Neurophilosophy at Work*, p. 156

a incorporar um mapa detalhado do mundo externo que, como qualquer mapa, constitui uma representação bem sucedida do seu domínio-alvo *sem que a ele esteja ligado causalmente*.

Importa observar que a analogia do mapa que representa características *objectivas* do mundo utilizada por Churchland para ilustrar o modo como o conteúdo intencional/semântico dos conceitos é assegurado numa perspectiva internalista aplica-se com igual propriedade a conceitos abstractos, isto é, que não representem elementos sensorialmente captáveis (e.g. “felicidade”, democracia”). Na referência que faz a este facto, Churchland lembra que o método que propõe para determinar a identidade semântica de redes conceptuais (unicamente pelas suas propriedades *geométricas*) dispensa por completo o conhecimento do termo linguístico que corresponde a cada ponto-protótipo. Por este motivo, é indiferente que os pontos-protótipo de dois espaços de estado distintos correspondam a conceitos de elementos observáveis do mundo ou a conceitos abstractos. Do ponto de vista da determinação da similaridade entre os espaços de estado, tal informação é irrelevante: são meramente *vértices*. O exemplo de identificação semântica referido anteriormente (figura 10), no qual cada vértice corresponde a uma família, seria igualmente bem sucedido se cada vértice/conceito de ambos os sólidos correspondesse, por hipótese, a um tipo de sistema político, ou a uma emoção. Numa palavra, o conhecimento das posições dos pontos-protótipo de espaços de estado distintos é quanto basta para determinar se são semanticamente idênticos ou não – o seu significado individual é indiferente.

3.6. O carácter reducionista da proposta de Churchland

A noção de que o conhecimento dos termos associados a cada ponto-protótipo é desnecessária para a determinação da identidade semântica de espaços de estado distintos é particularmente enfática do carácter *reducionista* da teoria semântica de Churchland, a que F&L se opõem por privilegiarem o paradigma lógico-linguístico/simbólico na explicação dos mecanismos cognitivos de acordo com o qual, por definição, os primitivos semânticos são unidades simbólicas/conceitos. Com a proposta de um método inteiramente sublinguístico para a identificação de redes conceptuais, Churchland, ao arrepio da tradição filosófica, relega o plano psicológico para uma posição secundária na determinação do conteúdo semântico dos estados mentais, colocando a tónica na neurobiologia computacional para esse fim. Esse

método, segundo defende, tem a vantagem acrescida de se coadunar melhor com a teoria evolucionária do que as teorias semânticas baseadas no paradigma simbólico, na medida em que não implica que o surgimento da capacidade linguística humana tenha significado um salto *qualitativo* na cognição animal.¹⁰⁷ A linguagem, afirma Churchland, é apenas o resultado da maior complexidade nos cérebros humanos dos mesmos mecanismos neuronais presentes em todos os animais possuidores de um sistema nervoso. Trata-se meramente uma “competência” a par de outras como a motora e a perceptiva, sendo que o facto de ser especificamente humana não se justifica por uma putativa propriedade exclusiva dos cérebros da nossa espécie – hipótese esta que, na verdade, não é mais do que uma adaptação do conceito de “alma” à terminologia moderna –, mas pelo facto de o desenvolvimento da capacidade comunicativa (ubíqua no reino animal) ter sido mais acentuado nos humanos em particular.¹⁰⁸

De facto, Churchland sustenta que a linguagem pode mesmo ser considerada uma extensão das faculdades perceptiva e motora. A faculdade perceptiva tem o propósito de permitir representar internamente o ambiente exterior e a motora o de possibilitar a navegação nesse ambiente e de manipulá-lo. A linguagem, de certo modo, retém ambas estas características, por ser ‘uma competência adquirida de *percepção* (opaca, é certo) e de *manipulação* (novamente, opaca) das *actividades cerebrais* dos elementos da mesma espécie [...]’.¹⁰⁹ A aptidão linguística, então, terá tido origem na adaptação (recente) dos mesmos mecanismos cognitivos existentes na Terra há centenas de milhões de anos, não sendo, nesse sentido, fundamentalmente diferente de qualquer outra aptidão evolutivamente adquirida.

Churchland considera a possibilidade de se erguer conta este último ponto o argumento de que ainda que o mecanismo neuronal subjacente à cognição humana e animal fosse o mesmo, o *conhecimento factual* não deixará ainda assim de ser apanágio

¹⁰⁷ Este ponto evidencia uma vez mais a importância anteriormente referida da teoria semântica no contexto geral do programa neurofilosófico de Churchland (cf. 1.2.2., 1.4.).

¹⁰⁸ Importa notar que esse desenvolvimento da capacidade comunicativa também não terá sido despoletado pela emergência de algo fundamentalmente original nos mecanismos cognitivos do homínido algures no decurso da sua evolução. Com efeito, crê-se que o motivo pelo qual apenas os humanos são capazes de linguagem complexa residirá no facto de a nossa ter sido a única espécie com um cérebro de grandes dimensões a tornar-se inteiramente *bípede*. Esta última característica tornou possível a execução de actividades com as mãos que anteriormente exigiam o uso da boca (e.g. caça, capacidade preênsil, etc.), o que por sua vez possibilitou a redução dos músculos e ossatura das maxilas e o concomitante desenvolvimento de uma maior capacidade de vocalização de sons distintos – de onde a comunicação fonética na origem da linguagem articulada se tornou viável, tendo o desenvolvimento subsequentemente desta última sido privilegiado pela selecção natural dadas as óbvias vantagens práticas que traz.

¹⁰⁹ *Ibid.*, p. 159

exclusivo da nossa espécie pelo facto de ser a única capaz de linguagem complexa. Em resposta, Churchland afirma: '[...] esta conclusão chauvinista é desmentida pela manifesta sofisticação perceptiva e conceptual dos animais, muitos dos quais possuem um mais extenso conhecimento factual dos seus nichos ambientais peculiares do que quaisquer forasteiros alguma vez possuirão. E pretende extrair uma indefensável distinção Cartesiana entre humanos e “bestas”, indefensável porque – linguagem aparte – as bestas demonstram (pelo menos ocasionalmente e até certo ponto) todas as capacidades cognitivas demonstradas pelo humano primitivo médio’.¹¹⁰

3.7. Resumo sinóptico da teoria psicosssemântica de Churchland

Em resumo, a teoria psicosssemântica de Churchland, conforme referido no parágrafo inicial da presente secção, caracteriza-se por ser: i) *internalista* – i.e. dispensa qualquer ligação causal ao mundo externo na determinação do conteúdo intencional dos conceitos; ii) *holista* – i.e. o conteúdo de cada conceito é determinado pela sua posição relativa a todos os outros conceitos da mesma rede conceptual/espaco de estado que ocupa; iii) e *reducionista* – i.e. a explicação do modo como os estados mentais adquirem o seu conteúdo semântico encontra-se ao nível neuronal e não ao nível linguístico, o que implica o corolário de que as redes conceptuais não são algo de exclusivamente humano mas vastamente difundido entre o reino animal. Centrando-se na intersecção destes três eixos teoréticos, a teoria final de Churchland prova ser resistente a todas as objecções iniciais de F&L, demonstrando em simultâneo: a) a possibilidade da formulação de um método fiável de determinação e comparação intersubjectiva de conteúdo semântico baseado num critério de *similaridade* (que não supõe implicitamente um critério de identidade como acontecia com as versões iniciais da proposta, segundo F&L); e b) a potencialidade explicativa do paradigma conexionista, pelo facto de ser suficientemente robusto para sustentar uma tal teoria e de fazê-lo em confluência com a teoria evolucionária.

Finda a exposição do processo de maturação da proposta semântica de Churchland, torna-se necessário avaliar a robustez *tout court* da mesma na sua forma ultimada (ou, pelo menos, actual). Será esse tópico da análise a encetar na próxima secção.

¹¹⁰ Ibid., p. 160

IV

Análise da Teoria Psicosssemântica de Paul Churchland

Havendo sido determinada a resistência da versão final da teoria psicosssemântica de Churchland às múltiplas objecções movidas por F&L no decurso da sua evolução ao longo da década que perdurou a controvérsia, procurar-se-á agora, nesta secção final, avaliar a sua robustez quando confrontada com a nova classe de problemas que suscita. Tal como na secção precedente, a avaliação da proposta de Churchland terá como ponto de partida um conjunto de objecções erguidas por F&L. Contudo, a análise a encetar não seguirá já o modelo meramente expositivo de apresentação do argumento de F&L e subsequente exposição do contra-argumento de Churchland, pelo motivo que será explicitado no que se segue.

A teoria psicosssemântica de Churchland foi apresentada na sua forma final em dois momentos – primeiramente em 1998,¹¹¹ e em maior profundidade mas sem alterações fundamentais em 2001.¹¹² Ora, no período intercalar (1999), F&L publicaram um artigo no qual expuseram um conjunto de críticas à versão final da proposta de Churchland que este publicou em 1998.¹¹³ O artigo Churchland de 2001, porém, não obstante o facto de expor a sua teoria em maior detalhe, deixa essas objecções sem resposta. Enquanto tal, o escopo da presente secção será o de apresentar as objecções que F&L avançaram no seu artigo de 1999 e procurar determinar se partindo estritamente do manancial teórico da proposta final de Churchland é possível extrapolar respostas para as mesmas. Naturalmente, ressalva-se que uma tal metodologia implicará que todas as conclusões não poderão deixar de ser meramente especulativas e assume-se à partida que não é certo que Churchland teria utilizado os mesmos argumentos caso tivesse abordado directamente as objecções em causa.

¹¹¹ Churchland, Paul (1998), "Conceptual Similarity across Sensory and Neural Diversity: The Fodor-Lepore Challenge Answered", in *On the Contrary*, MIT Press, Cambridge, Massachusetts, pp. 81-112

¹¹² Idem (2001), "*Neurosemantics: On the Mapping of Minds and the Portrayal of Worlds*", in K. E. White (ed.), *The Emergence of Mind: Proceedings of the international Symposium*, pp. 117-147

¹¹³ F&L (1999), "*All at Sea in Semantic Space: Churchland on Meaning Similarity*", *Journal of Philosophy* 96, n. 8, pp. 381-403

4.1. *Objecção 1: Homologia entre espaços de estado neuronais e semânticos*

A primeira objecção erguida por F&L à proposta final de Churchland é resumida pelos próprios da seguinte forma: ‘[...] mesmo que se tivesse uma noção de homologia transdimensional que [...] preservasse a similaridade (identidade) entre *estados neuronais*, ainda assim não se teria o direito de presumir que a similaridade (ou identidade) *semântica* seria igualmente preservada’.¹¹⁴

Como é patente na citação acima, F&L consideram que Churchland não fornece argumento algum que justifique a inferência de que o conteúdo semântico das representações mentais de dois indivíduos distintos será semelhante se os seus estados de activação neuronal concomitantes forem semelhantes, especialmente porque o método de comparação de redes conceptuais de Laakso e Cottrell na base da proposta Churchland foi originalmente concebido apenas para comparar padrões de activação neuronal, não sendo feita qualquer correspondência entre os mesmos e espaços de estado *semânticos* no seu contexto original: ‘Churchland julga que Laakso e Cottrell forneceram uma noção de “tipo de estado cerebral” que preserva a similaridade de tipos de estado neuronal em espaços que difiram na sua dimensionalidade neuronal. E (tacitamente) infere que por esse motivo forneceram uma noção transdimensional de similaridade de estado-de-conteúdo também. [...] Julgamos que [Churchland] simplesmente confundiu espaços *semânticos* (cujos ocupantes típicos são estados intencionais, assim descritos), com espaços *neuronais* (cujos típicos ocupantes são estados cerebrais, assim descritos)’.¹¹⁵ F&L, portanto, crêem que Churchland não demonstrou a existência de uma correlação causal entre determinados tipos de estado neuronal e certas representações mentais.

Ainda que Churchland não tenha respondido directamente a esta objecção, os exemplos com que ilustra a possibilidade de redução de diversos aspectos da cognição ao nível neuronal através do método de Laakso e Cottrell constituem em si o argumento a favor do isomorfismo entre espaços de estado semânticos e espaços de activação neuronal que F&L consideram estar em falta na sua proposta. Desses exemplos será aqui exposto aquele que é talvez o mais ilustrativo desse facto. Considerem-se os espaços cromáticos representados na figura 11.

¹¹⁴ F&L (2002), “All at Sea in Semantic Space: Churchland on Meaning Similarity”, in *The Compositionality Papers*, p. 193

¹¹⁵ *Ibid.*, p. 194

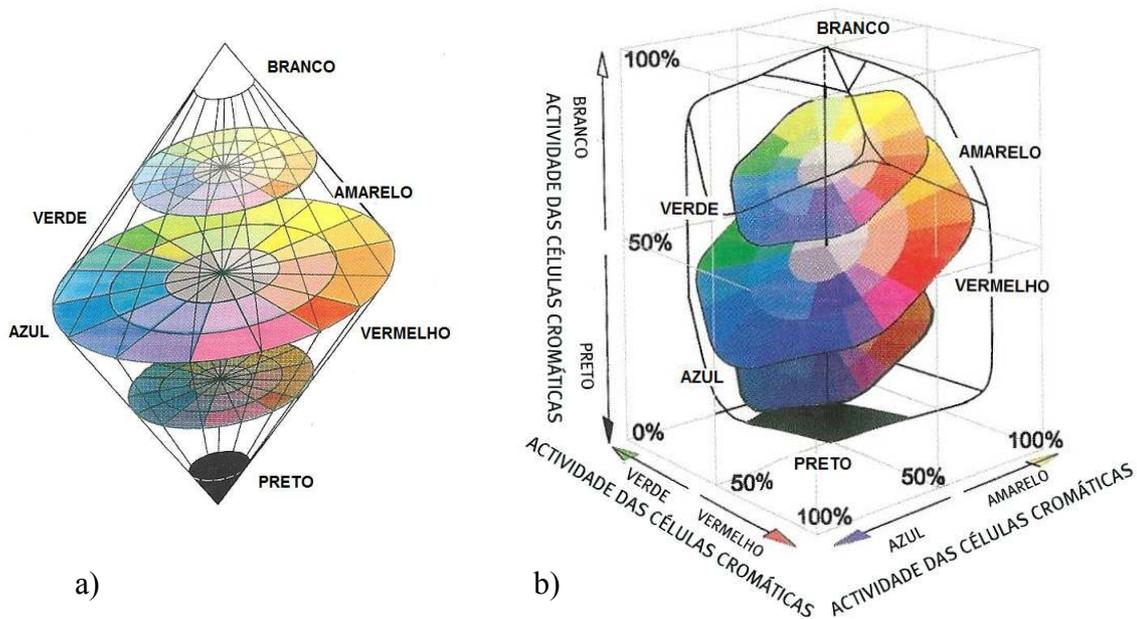


Fig. 11 a) Espaço cromático de Munsell; b) Espaço de activação de células cromáticas de Hurvich-Jameson.¹¹⁶

A figura ilustrada em 11 a) representa o que poderemos designar de *espaço fenomenológico de cores*. Foi originalmente proposto por Albert Munsell no início do século XX, e especifica a relação *subjectiva* existente entre as cores humanamente perceptíveis ao longo de três dimensões: *matiz* (eixo circular), *pureza* (eixo radial), e *luminosidade* (eixo vertical). A figura em 12 b), por seu turno, representa o espaço abstracto de activação das células responsáveis pela codificação de cores no sistema visual humano (ou seja, as cores correspondentes a cada vector de activação possível destas células).

Como Churchland observa: ‘[...] esta peculiar configuração de vectores de codificação possíveis [das células cromáticas] é estruturalmente quase idêntica à configuração peculiar, reconstruída por Munsell original e independentemente, das experiências de cor possíveis para os humanos normais’.¹¹⁷ Ora, a correspondência quase exacta entre ambos os espaços cromáticos fenomenológico e neuronal, não obstante terem sido criados “original e independentemente” com quase um século de

¹¹⁶ Churchland, Paul (2001), “*Chimerical Colors: Some Phenomenological Predictions from Cognitive Neuroscience*”, in *Neurophilosophy at Work*, a) p. 163, b) p. 169

¹¹⁷ Ibid., p. 169

intervalo entre si é, para Churchland, facilmente explicável: ‘Os paralelismos sistemáticos aqui descritos – ainda que altamente improváveis de um ponto de vista puramente *a priori* – deixam imediatamente de ser misteriosos se as experiências humanas de cor simplesmente *forem* os vectores de codificação de *output* [...]’.¹¹⁸

Esta última afirmação é de particular relevância na medida em que permite caracterizar de modo mais específico o tipo de reducionismo defendido por Churchland como radicalmente *eliminativista*. A partir desta perspectiva, o motivo pelo qual Churchland nunca propôs um argumento a favor da correspondência entre espaços de estado semânticos e neuronais torna-se evidente: Churchland não acredita que exista relação alguma de *correspondência* entre esses espaços e sim *identidade*, ou seja, as representações mentais, seja qual for o seu tipo, não são propriedades supervenientes (ontologicamente distintas) da computação neuronal, mas *são* a computação neuronal. Esta posição é obviamente controversa, mas o exemplo do isomorfismo entre os espaços de Munsell e de Hurvich-Jameson, assim como vários outros do mesmo tipo, constituem sem dúvida argumentos a seu favor. Logo, com base nesse facto, Churchland poderia afirmar em resposta à objecção inicial que o ónus de avançar um argumento a favor da tese de que espaços semânticos e espaços neuronais são ontologicamente distintos e que uma explicação da sua correlação é necessária impende sobre F&L. Até avançarem um tal argumento, a objecção falhará por incorrer numa falácia de petição de princípio.

4.2. Objecção 2: Similaridade neuronal/similaridade semântica

Uma segunda crítica que F&L dirigem a Churchland aponta para o facto de este defender que o método de Laakso e Cottrell permite determinar a semelhança entre duas redes ainda que o número de dimensões que compõem os seus respectivos espaços de estado seja diferente. Conforme referido na secção anterior, a comparação entre redes conceptuais segundo este método efectua-se pela sobreposição dos sólidos geométricos formados pelos seus respectivos pontos-protótipos na orientação dos mesmos que permita a maior congruência possível. Naturalmente, se um dos sólidos em causa pertencer a um espaço de estado constituído por um número de dimensões superior ao daquele com o qual será estabelecida a comparação, esta congruência não poderá ser perfeita. Contudo, segundo Churchland, o método que propõe continua a ser

¹¹⁸ Ibid.

perfeitamente aplicável nestes casos: ‘Se dois espaços dimensionalmente diferentes contiverem [...] um hipsólido n -dimensional com a mesma forma, [...], então ambos os espaços devem constituir ou conter um subespaço n -D, um hiperplano n -D de uma orientação ou outra, que confine cada hipsólido em questão. Uma vez que esses hipsólidos [...] têm a mesma dimensionalidade, não haverá problema em comparar os dois sólidos’.¹¹⁹ Noutros termos, o que Churchland propõe é que se os sólidos a serem comparados não forem idênticos, mas somente semelhantes – pelo facto de um ser apenas *parcialmente* congruente com o outro –, as redes conceptuais correspondentes a esses sólidos, concomitantemente, serão também *semelhantes*. Assim, se houver congruência dos sólidos nas dimensões que *ambos* partilham, serão também semanticamente semelhantes ainda que as dimensões suplementares do sólido maior não tenham correspondente no outro.

F&L contestam este último ponto por considerarem que mesmo admitindo a possibilidade de se compararem espaços *neurais* de dimensionalidade diferente por inclusão (geometricamente congruente) do sólido do espaço abstracto de activação menor num *subsólido* do espaço com mais dimensões, tal não garantiria que os espaços *semânticos* correspondentes seriam igualmente similares. O exemplo com ilustram esta objecção é o seguinte: ‘Suponha-se que tenho apenas três dimensões semânticas, duro-mole, preto-branco, pesado-leve. E suponha-se que o meu conceito de ROCHA é identificado por um vector que especifica uma região no espaço que estas dimensões definem. O mesmo no caso do leitor [...]. A diferença entre nós é que, ao passo que estas dimensões definem a *totalidade* do meu espaço semântico, definem apenas um subespaço menor do seu. Agora, resultará disto que os nossos conceitos de ROCHA serão semelhantes? Provavelmente não pensaremos que sim. Pois talvez o seu espaço possua a dimensão ANIMADO que, por hipótese, a minha não possui. E suponha-se que o leitor considera que as rochas são, de facto, animadas [...]. Suponha-se também que o seu espaço contém uma dimensão para abstracção, e que o leitor pensa que as rochas são abstractas [...]. Será o conceito de ROCHA do leitor *ainda* semelhante ao meu? Se houver *quaisquer* casos deste tipo em que a resposta seja “não”, Churchland perde’.¹²⁰

¹¹⁹ Churchland, Paul (1998), “Conceptual Similarity across Sensory and Neural Diversity: The Fodor-Lepore Challenge Answered”, in *On the contrary*, p. 89

¹²⁰ F&L, *ibid.*, p. 196

Esta crítica tem uma resposta notoriamente simples na medida em que o exemplo de F&L citado acima, que pretende ser uma redução ao absurdo da proposta de Churchland, constitui na verdade uma demonstração da flexibilidade da mesma que é um dos maiores argumentos a favor da sua plausibilidade biológica. Com efeito, Churchland reconhece em diversas ocasiões que muito provavelmente *nenhum* conceito partilhado por dois indivíduos será idêntico em todos os aspectos do seu conteúdo semântico precisamente pelo facto de que cada rede neuronal configurará um espaço de estado dimensionalmente idiossincrático. Assim, sem necessidade de acrescentar qualquer alínea à exposição original do seu método, Churchland poderia simplesmente argumentar em resposta a esta crítica de F&L que os conceitos de “rocha” dos indivíduos do exemplo em causa, não obstante não serem idênticos, seriam ainda assim o mesmo conceito pelo facto de preservarem uma certa medida de *similaridade*, demonstrada na prática pelo facto de ambos os indivíduos utilizarem esse conceito com respeito aos mesmos referentes (rochas) e por terem crenças coincidentes a respeito de um número relevante de características desses referentes. Se os indivíduos em causa possuísem conceitos de “rocha” completamente diferentes do ponto de vista semântico, tal seria efeito do facto de as vivências de ambos relativas a rochas que presidiram à configuração dos pontos-protótipo desse conceito pelas suas respectivas redes neuronais terem sido também muito diferentes. Mas nesse caso, a dissemelhança nos seus respectivos espaços de estado semânticos seria devidamente espelhada numa dissemelhança entre os seus espaços de estado neuronais.

Assumindo esta perspectiva, a objecção de F&L devém um exemplo das vantagens explicativas da proposta de Churchland relativamente às teorias da identidade, pelo facto de, ao contrário destas últimas, admitir a “matização” do conteúdo semântico que está na base da possibilidade de um mesmo conceito ser partilhado por indivíduos cujas estruturas neuronais e espaços de estado semânticos correspondentes são *por norma* dimensionalmente díspares. Em todo o caso, note-se que a resposta dada à objecção de 4.1. aplica-se com igual propriedade a esta, na medida em que a objecção de que a similaridade entre espaços de estado neuronal não preserva necessariamente a similaridade nos espaços de estado semânticos supõe a referida cisão ontológica entre ambos que o eliminativismo de Churchland nega existir.

4.3. *Objecção 3: Composicionalidade*

O último problema colocado por F&L a ser aqui considerado prende-se com a putativa incapacidade por parte dos sistemas conexionistas de assegurarem a *composicionalidade* conceptual de que qualquer teoria semântica não pode prescindir. Importa notar que a análise da discussão entre F&L e Churchland relativa a este ponto é particularmente problemática na medida em que a noção de composicionalidade tem duas interpretações distintas, sendo que aquela que Churchland tem em mente na resposta que propõe para a objecção de F&L não é aquela a que esta última se referia (ainda que ambas sejam igualmente pertinentes). Por esse motivo, no que se segue será em primeiro lugar demonstrada a solução que Churchland propõe para a vertente do problema da composicionalidade que aborda, e procurar-se-á subsequentemente responder à questão tal como foi efectivamente colocada por F&L (a que Churchland não responde), fazendo-se uso para tal do mesmo método aplicado na resposta aos problemas já abordados nesta secção deixados em aberto por Churchland, ou seja, partir de publicações suas relevantes a partir das quais seja possível extrapolar a solução que provavelmente teria dado ao problema.

As concepções alternativas de composicionalidade que estarão em causa distinguem-se entre *lexical* e *frásica*. A primeira diz respeito ao facto de os conceitos, na sua maioria, subsumirem um número indeterminado de outros conceitos que são, por assim dizer, os “átomos” a partir dos quais são constituídos.¹²¹ Por exemplo, o conceito “cão” pode ser considerado como uma “molécula semântica” composta pelos conceitos de “cauda”, “orelhas”, “patas”, etc. Nesta acepção, a noção de composicionalidade refere-se à possibilidade de concatenação e de decomposição de conceitos complexos, respectivamente a partir de e em outros mais simples. Na sua segunda acepção, por outro lado, a composicionalidade está associada à noção de *produtividade*, isto é, à possibilidade de formação de um número ilimitado de *frases* a partir da concatenação *sequencial* de um manancial limitado de conceitos numa dada linguagem, de acordo com um conjunto de regras gramaticais recursivas. Assim, numa palavra, a composicionalidade lexical aponta para o carácter composicional dos conceitos ditos “complexos” tomados individualmente, ao passo que a frásica se refere à possibilidade da criação de proposições mediante a concatenação de vários conceitos.

¹²¹ Os conceitos “simples”, ou seja, impassíveis de decomposição em conceitos mais fundamentais, serão, por exemplo, os que se referem directamente os dados dos sentidos (e.g. vermelho, amargo, áspero, etc.).

O problema que a composicionalidade apresenta ao connexionismo é o mesmo em ambas as suas acepções, e prende-se ao facto de os pontos-protótipo não serem passíveis de concatenação nem de decomposição. Um ponto-protótipo num espaço de estado semântico de uma rede neuronal correspondente ao conceito “cão”, por exemplo, não é composto por um conjunto de pontos-protótipo subordinados correspondentes aos conceitos de “patas”, “cauda”, etc., mas é em si uma unidade semântica, ou seja, o equivalente a um conceito simples. Esta impossibilidade de concatenação e de decomposição de pontos-protótipo parece colocar problemas sérios ao connexionismo, na medida em que deixa em aberto a possibilidade de um indivíduo possuir, por hipótese, o conceito de “cão”, sem que possua concomitantemente qualquer um dos referidos conceitos “atómicos” que normalmente consideraríamos indispensáveis na determinação do conteúdo semântico desse conceito (“cauda”, “patas”, “orelhas”, etc.).

Do mesmo modo, o connexionismo enfrenta dificuldades quando confrontado com a exigência da composicionalidade frásica, na medida em que a irrelevância do conteúdo semântico de cada conceito individual para a comparação de redes conceptuais segundo o método proposto por Churchland (i.e. os conceitos/pontos-protótipo podem ser considerados meramente como vértices para que a similaridade conceptual intersubjectiva possa ser determinada), apesar de ter a vantagem de tornar desnecessária a equivalência dimensional entre espaços de estado semânticos para que a comparação dos mesmos seja viável, acarreta um factor negativo: os conceitos/pontos-protótipo, *qua* vértices de sólidos geométricos, não são codificados pelas redes neuronais de acordo com a sua *categoria gramatical* (e.g. nome, predicado, quantificação, etc.), mas apenas indiferentemente como conceitos. A proposta de Churchland, assim, não obstante revelar-se promissora na explicação do modo como os conceitos adquirem o seu conteúdo semântico, deixa sem solução a questão fundamental do modo como estes se distinguem segundo a sua função sintáctica.

4.3.1. Composicionalidade lexical

Como foi dito acima, a objecção de F&L e a subsequente resposta de Churchland supõem acepções diferentes da noção de composicionalidade – em particular, os primeiros referiam-se à composicionalidade frásica ao passo que Churchland considerou exclusivamente a lexical. Uma vez que a teoria de semântica de Churchland resultará defectiva se se revelar incapaz de responder satisfatoriamente ao problema nas suas

duas vertentes, ambas serão tomadas em consideração no que se segue. Atente-se em primeiro lugar à abordagem que Churchland faz ao problema. A sua posição a respeito da noção de composicionalidade lexical é definida categoricamente: ‘Aquilo que os conceitos *não* fazem, de acordo com a nossa perspectiva, e em contraste com perspectivas históricas tais como as de Locke e de Hume, é formar hierarquias composicionais, de tal modo que os conceitos complexos seriam literalmente constituídos por uma concatenação e recursão apropriada a partir de um léxico finito de conceitos simples’.¹²²

Churchland, portanto, opõe-se ao que designa de “perspectiva conceptual molecular”, e justifica a sua posição com dois argumentos. Em primeiro lugar, sustenta que a perspectiva clássica, de acordo com a qual os conceitos complexos são compostos por outros mais simples, não se coaduna com o conhecimento actual do modo como as crianças formam as suas redes conceptuais: ‘As crianças aprendem a discriminar rostos, de outras coisas e uns dos outros, substancialmente antes de conseguirem fazer o mesmo com olhos, narizes, bocas e orelhas. E subsequentemente aprendem, por sua vez, a discriminar olhos, de outras coisas e uns dos outros, substancialmente antes de conseguirem fazer o mesmo com pupilas, pestanas, íris ou lentes [...] Em geral, os conceitos que se aprendem em primeiro lugar são os que os autores históricos teria considerado ideias altamente “complexas”, tais como “biscoito”, “cão”, “rosto”, “ave” ou “sapato”’.¹²³ Então, de acordo com Churchland, a aprendizagem de que determinados conceitos se encontram subordinados a outros é algo de *secundário* à aquisição do conceito subordinante. E, de facto, é perfeitamente normal que uma criança seja capaz de discriminar cães com um grande grau de eficiência sem que no entanto esteja ainda na posse dos conceitos correspondentes à anatomia canina (ou, por outro lado, dos conceitos de “animal” ou “mamífero” que subordinam o de “cão”).

Como explicitado na secção precedente, Churchland considera que as redes conceptuais representam o mundo como um *mapa* das suas características mais salientes. Enquanto tal, como qualquer mapa, também as redes conceptuais podem ser progressivamente *refinadas* para representarem características menos salientes do seu domínio-alvo do que as originalmente apreendidas. E, de facto, é exactamente este progressivo refinamento do espaço de estado semântico que se verifica no processo de

¹²² Churchland, Paul (2008), “*Neurosemantics: On the Mapping of Minds and the Portrayal of Worlds*”, in *Neurophilosophy at Work*, p. 147

¹²³ *Ibid.*, p. 148

treino de redes neuronais artificiais: quanto maior a frequência de exposição de uma rede a um determinado tipo de estímulo, mais estruturada se tornará a configuração dos seus pesos sinápticos e maior a quantidade de padrões esta será capaz de discriminar nos mesmos estímulos, o que se traduzirá numa progressiva subdivisão do seu espaço de estado semântico em regiões-protótipo subordinadas a outras mais genéricas e, logo, na hierarquização e alargamento do manancial conceptual que incorpora. Churchland refere como exemplo específico disso mesmo a rede de reconhecimento facial de Cottrell representada na figura 12 a), que aprendeu inicialmente a discriminar apenas entre estímulos correspondentes a “rosto” e a “não-rosto”, após o que refinou a sua capacidade discriminatória subdividindo espontaneamente a região-protótipo do seu espaço de estado semântico correspondente ao conceito de “rosto” nas subcategorias de “rosto masculino” e “rosto feminino”.

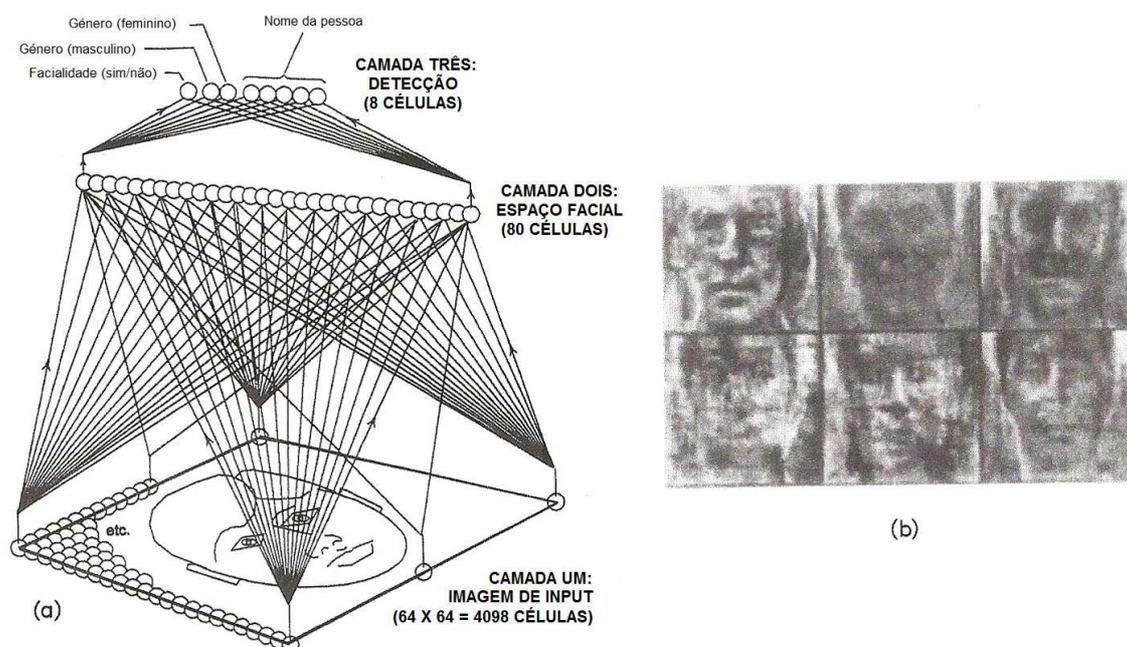


Fig. 12 a) A rede de discriminação facial de Cottrell; b) Seis possíveis padrões de activação da camada de *input* para esta rede. Cada um constitui o “estímulo preferencial” de um dos oitenta neurónios da camada intermédia, ou seja, são o padrão de activação com o qual todas as imagens de *input* recebidas da primeira camada são comparadas.¹²⁴

O segundo argumento de Churchland em resposta ao problema da composicionalidade lexical baseia-se igualmente no funcionamento da rede neuronal de

¹²⁴ Cottrell, G. (1991), “Extracting Features from Faces Using Compression Networks: Face, Identity, Emotions and Gender Recognition Using Holons”, in D. Touretzky, J. Elman, T. Sejnowski, e G. Hinton (eds.), *Connectionist Models: Proceedings of the 1990 Summer School*, San Mateo, CA: Morgan, pp. 328-337

Cottrell. Churchland pretende contrariar a noção de que as redes neuronais constroem as suas representações mediante a concatenação de unidades semânticas “simples” em qualquer acepção possível de “concatenação” que se assemelhe ao modo como Locke e Hume a conceberam. Pois não obstante a rede como um todo ter a capacidade de reconhecer rostos sem que no entanto possua os conceitos “simples” que correspondem tipicamente a rostos (olhos, nariz, boca, etc.), poder-se-ia considerar, em relação ao caso presente, que o conceito de “rostos” adquirido pela rede é formado pela concatenação de cada um dos *inputs* recebidos na camada intermédia de cada uma das 4098 células que constituem a camada sensorial, sendo que cada célula individual da camada intermédia contribuiria com a representação de uma fracção do rosto total para a formação desse conceito. Desse modo, o conceito de rosto seria afinal formado a partir de primitivos semânticos simples, correspondentes às representações das células da camada intermédia da rede, do mesmo modo que a imagem de um rosto numa televisão é composta por unidades cromáticas simples (pixéis). Logo, as redes neuronais adquiririam os seus conceitos de um modo não fundamentalmente diferente daquele que foi proposto por Locke por Hume.

O exemplo da rede de Cottrell torna evidente por que motivo esta concepção do funcionamento das redes neuronais está radicalmente errada. Considere-se a figura 12 b). Cada uma das seis imagens representadas nesta figura corresponde ao padrão de activação preferencial de *uma única* célula da camada intermédia. O que importa notar neste ponto é que cada imagem formada por cada célula individual não representa uma parcela específica do rosto apresentado à rede, mas é em si uma representação vaga e difusa da *totalidade* desse rosto. Estas representações não correspondem, portanto, a nada de semelhante a unidades semânticas “simples” no sentido clássico do termo. Por esse motivo, Churchland utiliza para designá-las o termo “*holon*”, cunhado por Janet Metcalfe, enfatizando assim o facto de não obstante serem representações atómicas, não podem constituir primitivos semânticos “simples” pelo facto de conterem já informação altamente complexa acerca do estímulo apresentado à rede *como um todo*. Assim, uma rede neuronal não identifica um determinado estímulo como sendo um rosto por concatenação das imagens parcelares recebidas na camada de *input*, mas pela semelhança (estatisticamente calculada na camada intermédia) entre essas várias imagens – cada imagem, reitera-se, já em si vagamente semelhante a um rosto –, ao “rostos-protótipo” formado pela rede no decurso do seu processo de treino (sendo que esse “rostos-protótipo”, como referido na secção II, também não constitui uma unidade

semântica “simples” e sim algo melhor descrito como a “forma platónica” de rosto para essa rede, ou, noutros termos, a sua concepção genérica de “rostidade”). Este exemplo, por extrapolação, demonstra que as redes neuronais não constroem as suas redes conceptuais a partir de nada que se assemelhe a um primitivo semântico no sentido clássico do termo.

4.3.2. Composicionalidade frásica

Com ambos os argumentos anteriores Churchland demonstra efectivamente que o problema da composicionalidade lexical não afecta de todo a sua proposta, uma vez que supõe uma noção clássica de “conceito” que não tem aplicação no contexto do conexionismo. A sua argumentação, contudo, deixa em aberto o problema da composicionalidade frásica. Em todo o caso, é possível extrapolar uma resposta possível à questão – sem sair do âmbito estrito da sua teoria – a partir de um texto alheio à sua contenda com F&L no qual discute precisamente o tópico das redes neuronais com a capacidade de reconhecerem estruturas gramaticais.¹²⁵

As redes que Churchland discute em particular são comumente designadas de “redes de Elman”, por terem sido desenvolvidas por Jeffrey Elman, especialista na aplicação de redes neuronais à teoria da linguagem. O desenvolvimento de uma exposição aturada do seu funcionamento seria impraticável no presente contexto, pelo que serão apenas descritas em traços gerais as características das mesmas necessárias à demonstração da sua proficuidade na resolução do problema em causa.

Os objectivos de Elman no desenvolvimento destas redes foram: i) determinar a possibilidade de estas aprenderem a abstrair, a partir de um conjunto de frases simples recebido no processo de treino, as categorias gramaticais dos conceitos que compunham estas últimas (pronome, verbo plural, substantivo singular, etc.); ii) e determinar se conseguiam aprender, a partir de um conjunto de frases mais complexas, a distinguir entre frases gramaticalmente correctas e incorrectas. Verificou-se que, depois de um processo de treino constituído pela apresentação de 10.000 frases de vários graus de complexidade, as redes foram bem sucedidas em ambas as tarefas. Além de terem aprendido a prever com grande eficácia o tipo de palavra que seria gramaticalmente permissível inserir-se no final de uma frase incompleta, aprenderam ainda a ‘discriminar, como gramaticais, quase exactamente o mesmo conjunto de frases geradas

¹²⁵ Churchland, Paul (1996), *The Engine of Reason, the Seat of the Soul*, pp. 137-143

como gramaticais a partir do conjunto original de regras genuinamente produtivas de tipo chomskiano'.¹²⁶ Como exemplo, Churchland refere que uma rede de Elman aprendeu a determinar correctamente a concordância sujeito-verbo em frases tão complexas quanto “rapazes que beijam rapariga que alimenta o cão perseguem gatos”.

O método computacional utilizado pelas redes de Elman para obter estes resultados difere do das redes mencionadas até aqui num ponto fundamental. As redes conexionistas normais de reconhecimento de padrões, como se explicou, recebem um determinado de estímulos durante o processo de treino e a partir desse conjunto extraem os padrões mais salientes dos mesmos. Ora, como Churchland refere, esse método falha no caso do reconhecimento da função gramatical dos conceitos em *inputs* frásicos porquanto ‘este processo deliberadamente nivela todas as variações contextuais, e assim oblitera toda a informação referente ao modo como as palavras de *input* são codificadas diferentemente, de ocasião em ocasião, em função da sua posição sequencial ou temporal variante numa frase [e] agora que estamos a considerar frases de extensão não-trivial, temos de tomar em conta exactamente essa informação’.¹²⁷ Logo, o que o processo de reconhecimento gramatical exige à partida é que as redes neuronais ‘em lugar de *nivelarem* [estatisticamente] às cegas os vectores que ocorrem em momentos variados, [que] considerem a sequência dos vectores de palavras ao longo do tempo e que se perguntem o que é que a *trajectória* resultante no espaço vectorial poderá significar’.¹²⁸ Noutros termos, ao contrário das redes anteriormente discutidas, treinadas para extrair padrões estatísticos a partir dos estímulos recebidos na fase de treino considerados na sua globalidade, as redes de Elman têm a capacidade de extrair padrões presentes na própria *sequência* da recepção desses estímulos (neste caso, palavras no contexto de frases, na medida em que, como Churchland afirma, ‘uma frase [...] é uma sequência temporal de palavras’).¹²⁹ Assim, os espaços de estado destas últimas redes, ao invés de formarem sólidos passíveis de serem comparados quanto ao seu conteúdo semântico, como Churchland propôs, formam configurações de trajetórias possíveis de vectores de activação correspondentes a tipos gramaticais de frases. O que importa relevar neste ponto é que o método de Churchland para comparação de redes conceptuais quanto ao seu conteúdo semântico aplica-se com igual propriedade às redes de Elman, na medida em que se verifica que frases gramaticalmente semelhantes

¹²⁶ Ibid., p.140

¹²⁷ Ibid.

¹²⁸ Ibid., pp. 140-141

¹²⁹ Ibid. 136

adquirem configurações de trajectórias semelhantes no espaço vectorial, e frases gramaticalmente dissemelhantes adquirem configurações dissemelhantes.

Assim, não obstante os factos de as redes de Elman terem sido modeladas no contexto da linguística e não no da filosofia da mente, e de Churchland não fazer referência às mesmas com o expresso intuito de responder ao problema da composicionalidade frásica erguido por F&L (que, reitere-se, nunca chega a abordar directamente), não é ilegítimo avançar a proposta de que este problema poderia ser solucionado por Churchland mediante a extensão do seu método de comparação de redes para além do seu âmbito original (semântico) de modo a abarcar igualmente a possibilidade de comparação de espaços de estado gramaticais tal como Elman os descreve, não já pela congruência de sólidos formados por pontos-protótipo, mas de configurações de trajectórias vectoriais.

4.4. Nota conclusiva

De tudo o que foi dito, a ilação a retirar com que se conclui a presente dissertação é a de que a teoria psicosemântica de Churchland demonstra ser sem dúvida uma alternativa forte a teorias concorrentes fundadas no paradigma simbólico/lógico-linguístico da cognição, não só pela robustez que revela possuir face às objecções erguidas pelos proponentes deste último mas acima de tudo pelo facto de o seu desenvolvimento em permanente diálogo com as neurociências a dotar de argumentos a favor da sua plausibilidade biológica que os modelos baseados na perspectiva tradicional da cognição como um processo fundamentalmente linguístico não possuem, por motivos já referidos.

Naturalmente, este resultado final reflecte-se positivamente no projecto neurofilosófico como um todo, na medida em que encontra assim na teoria semântica de Churchland o fundamento firme requerido para a prossecução do seu principal objectivo programático: a redução das ciências superiores da cognição à linguagem da neurociência. Em todo o caso, independentemente deste projecto vir ou não ser bem sucedido, a eficácia da metodologia de teorização interdisciplinar adoptada por Churchland na elaboração da sua proposta psicosemântica não pode deixar dúvidas quanto à sua fertilidade. Por esse motivo, termino citando as palavras de Churchland, com as quais exprime a sua opinião acerca daquele que deverá ser o futuro do

pensamento filosófico: ‘Para onde deverá seguir a filosofia a partir daqui? A resposta dificilmente poderia ser mais óbvia: em direcção ao cérebro’.¹³⁰

¹³⁰ Churchland, Paul (2008), “*Into the Brain: Where Philosophy Should go from Here*”, in *Neurophilosophy at Work*, p. 238

Bibliografia

Descartes, Renée (2003), *Meditações Metafísicas*, Rés-Editora, Porto

Churchland, Patricia (1986), *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, MIT Press, Nova Iorque

Churchland, Patricia (2002), *Brain-Wise*, MIT Press, Cambridge, Massachussets

Churchland, Patricia & Churchland, Paul (1998), “*Intertheoretic Reduction: A Neuroscientist’s Field Guide*”, in *On the Contrary*, MIT Press, Cambridge, Massachussets, pp. 65-80

Churchland, Patricia & Churchland, Paul (1998), “*Conceptual Similarity across Sensory and Neural Diversity*”, in *On the Contrary*, MIT Press, pp. 81-112

Churchland, Paul (1979), *Scientific Realism and the Plasticity of the Mind*, Cambridge University Press, Cambridge, Massachussets

Churchland, Paul (1988), *Matter and Consciousness* (edição revista), MIT Press, Cambridge, Massachussets

Churchland, Paul (1992), *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, MIT Press, Cambridge, Massachussets

Churchland, Paul (1996) “*Fodor and Lepore: State-Space Semantics and Meaning Holism*”, in *The Churchlands and Their Critics*, McCauley, R. N. (ed.), Blackwell Publishers Ltd., Cambridge, Massachussets, pp. 272-277

Churchland, Paul (1996), *The Engine of Reason, the Seat of the Soul*, MIT Press, Cambridge, Massachussets

Churchland, Paul (1998), “*Could a Machine Think?*”, in *On the Contrary*, MIT Press, Cambridge, Massachussets, pp. 47-64

Churchland, Paul (1998), “*Conceptual Similarity across Sensory and Neural Diversity: The Fodor-Lepore Challenge Answered*”, in *On the Contrary*, MIT Press, Cambridge, Massachussets, pp. 81-112

Churchland, Paul (2008), “*What Happens to Reliabilism Liberated from Propositional Attitudes?*”, in *Neurophilosophy at Work*, Cambridge University Press, Nova Iorque, pp. 88-112

Churchland, Paul (2008), “*Into the Brain: Where Philosophy Should go From Here*”, in *Neurophilosophy at Work*, Cambridge University Press, Nova Iorque, pp. 232-238

Churchland, Paul (2008), “*Neurosemantics: On the Mapping of Minds and the Portrayal of Worlds*”, in *Neurophilosophy at Work*, Cambridge University Press, Nova Iorque, pp. 126-160

Churchland, Paul (2008), “*On the Reality (and Diversity) of Objective Colors: How Color-Qualia Space is a Map of a Reflectance-Profile Space*”, in *Neurophilosophy at Work*, Cambridge University Press, Nova Iorque, pp. 198-231

Cottrell, G. (1991), “*Extracting Features from Faces Using Compression Networks: Face, Identity, Emotions and Gender Recognition Using Holons*”, in D. Touretzky, J. Elman, T. Sejnowski, e G. Hinton (eds.), *Connectionist Models: Proceedings of the 1990 Summer School*, San Mateo, CA: Morgan, pp. 328-337

Fodor, J. & Lepore, E. (1992), “*Paul Churchland: State Space Semantics*”, in *Holism: A Shopper’s Guide*, Blackwell Publishers Ltd., Cambridge, Massachussets, pp. 187-208

Fodor, J. & Lepore, E. (1996), “*Reply to Churchland*”, in *The Churchlands and Their Critics*, McCauley, R. N. (ed.), Blackwell Publishers Ltd., Cambridge, Massachussets, pp. 159-162

Fodor, J. & Lepore, E. (2002), “*All at Sea in Semantic Space: Churchland on Meaning Similarity*”, in *The Compositionality Papers*, Clarendon Press, Oxford

Hebb, D. O. (1949), *The Organization of Behaviour*, John Wiley & Sons, Nova Iorque

Prinz, Jesse J. (2006), “*Empiricism and State Space Semantics*”, in *Paul Churchland*, Keeley, B. (ed.), Cambridge University Press, Nova Iorque, pp. 88-112

Laakso, A., & Cottrell, G. (1998), “*How can I know what you think?: Assessing representational similarity in neural systems*”, Proceedings of the 20th Annual Cognitive Science Conference

Laakso, A., & Cottrell, G. (2000), “*Content and cluster analysis: Assessing representational similarity in neural systems*”, in *Philosophical psychology*, 13, pp. 47-76

Land, Edwin (1977), “*The Retinex Theory of Color Vision*”, Scientific American, Dezembro, pp. 108-128

McCulloch, W.S., Pitts, W. (1943), “*A logical calculus of the ideas immanent in nervous activity*”, Bulletin of Mathematical Biophysics, 5

Medler, David (1998), “*A Brief History of Connectionism*”, Neural Computing Surveys, 1 (2)

Newell, A. & Simon, H. A. (1976), “*Computer Science as empirical enquiry: Symbols and Search*”, Communications of the ACM, 19:113-126

Pitts, W., McCulloch, W.S. (1947), “*How we know universals: the perception of auditory and visual forms*”, Bulletin of Mathematical Biophysics, 9

Quine, W. V. O. (1951), “*Two Dogmas of Empiricism*”, *The Philosophical Review* 60: 20-43

Quine, W. V. O. (1969), “*Epistemology Naturalized*”, in *Ontological Relativity & other Essays*, Columbia University Press, Nova Iorque, pp. 69-90

Quine, W. V. O. (1969), “*Natural Kinds*”, in *Ontological Relativity & other Essays*, Columbia University Press, Nova Iorque, pp. 114-138

Rosenblatt, F. (1958), “*The Perceptron: A probabilistic model for information storage and organization in the brain*”, *Psychological Review*, 65:386-408

Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986), “*Learning representations by back-propagating errors*”, *Nature*, 323:533-536

Selfridge, O. G. (1959), “*Pandemonium: A paradigm for learning*”, *Proceedings of the Symposium on Mechanisation of Thought Processes*, pp. 511-529

Widrow, B., Hoff, M.E. (1960), “*Adaptive switching circuits*”, 1960 IRE WESCON, *Convention Record*, pp. 96-104

Wiener, N. (1948), *Cybernetics*, John Wiley & Sons, Nova Iorque