

Capítulo 8

E o significado?

Cláudia Freitas

Publicado em: 26/09/2023

Semântica lida com o sentido do que é comunicado por meio da linguagem (em oposição ao que é comunicado por imagens ou sons não verbais, por exemplo). Assim, a **semântica estuda o significado de palavras e frases**. Mas a simplicidade relativa ao que é semântica acaba aí. Nos estudos linguísticos, a semântica é conhecida como “um domínio de investigação de limites movediços” e para o qual não há jargões bem estabelecidos (Ilari; Geraldi, 1985, p. 6). A questão “o que é o significado de uma palavra?” (e também o de uma frase) é um dos problemas nucleares da investigação semântica, e sua resposta irá depender da perspectiva teórica adotada. Essa característica é diferente de outras áreas do conhecimento, como a zoologia, por exemplo, em que não há controvérsia sobre o que é um animal.

No PLN, diferentes maneiras de conceber o significado se manifestam em diferentes abordagens para o tratamento do sentido, como veremos nos Capítulos Semântica com técnicas simbólicas e Semântica distribucional.

Mas... sabemos o que é significado e os dicionários – objetos que contêm o **significado** das palavras – não só existem, como são úteis. Então, por que tanta dificuldade? Por que “limites movediços”?

Para ilustrar essa ideia, vamos fazer uma analogia entre a observação – e descrição – do significado de uma palavra e a observação de uma onda, como narrada em “Palomar na praia”, de um livro de Ítalo Calvino. Trata-se de um capítulo curto, e alguns trechos são transcritos abaixo. A história gira em torno de alguém – Palomar – tentando conhecer algo – uma onda – de forma maximamente objetiva. Palomar, aliás, além de nome do protagonista, é o nome de um observatório astronômico que, durante muito tempo, ostentou o maior telescópio do mundo¹. Vamos à história, apresentada no Quadro 8.1.

Quadro 8.1: Trechos da história “Palomar na praia” de Ítalo Calvino.

O senhor Palomar está de pé na areia e observa uma onda. Não que esteja absorto na contemplação das ondas. Não está absorto, porque sabe bem o que faz: quer observar uma onda e a observa. (...). Em suma, não são “as ondas” que ele pretende observar, mas uma simples onda e pronto: no intuito de evitar as sensações vagas, ele predetermina para cada um de seus atos um objetivo limitado e preciso.

O senhor Palomar vê uma onda apontar na distância, crescer, aproximar-se, mudar de forma e de cor, revolver-se sobre si mesma, quebrar-se, desfazer-se. A essa altura poderia convencer-se de ter levado a cabo a operação a que se havia proposto e ir-se embora. Contudo, isolar uma

¹Devo à Helena Franco Martins a apresentação deste texto como alegoria tanto para as tentativas de apreensão do significado como para a crise relativa ao conhecimento/ciência.



onda da que se lhe segue de imediato e que parece às vezes suplantá-la ou acrescentar-se a ela e mesmo arrastá-la é algo muito difícil, assim como separá-la da onda que a precede e que parece empurrá-la em direção à praia, quando não dá até mesmo a impressão de voltar-se contra ela como se quisesse fechá-la. (...).

Em suma, não se pode observar uma onda sem levar em conta os aspectos complexos que concorrem para formá-la e aqueles também complexos a que essa dá ensejo. Tais aspectos variam continuamente, decorrendo daí que cada onda é diferente de outra onda; mas da mesma maneira é verdade que cada onda é igual a outra onda, mesmo quando não imediatamente contígua ou sucessiva; enfim, são formas e sequências que se repetem, ainda que distribuídas de modo irregular no espaço e no tempo. Como o que o senhor Palomar pretende fazer neste momento é simplesmente ver uma onda, ou seja, colher todos os seus componentes simultâneos sem descurar de nenhum, seu olhar se irá deter sobre o movimento da água que bate na praia a fim de poder registrar os aspectos que a princípio não havia captado (...).

(...) Foi uma dessas línguas baixas de areia que o senhor Palomar escolheu como ponto de observação, porque as ondas nelas batem obliquamente de uma parte e de outra, e ao cavalgarem por cima da superfície semi-submersa vão encontrar-se com as que chegam da outra parte. (...).

O senhor Palomar está procurando agora limitar seu campo de observação; se tem presente um quadrado de, digamos, dez metros de praia por dez metros de mar, pode levantar um inventário de todos os movimentos de ondas que ali se repetem com frequência variada dentro de um dado intervalo de tempo. A dificuldade está em fixar os limites desse quadrado, porque, por exemplo, se ele considera como o lado mais distante de si a linha em relevo de uma onda que avança, essa linha ao aproximar-se dele irá, erguendo-se, ocultar de sua vista tudo o que está atrás; e eis que o espaço tomado para exame se destaca e ao mesmo tempo se comprime. (...)

Contudo, o senhor Palomar não perde o ânimo e a cada momento acredita haver conseguido observar tudo o que poderia ver de seu ponto de observação, mas sempre ocorre alguma coisa que não tinha levado em conta. Prestar atenção em um aspecto faz com que este salte para o primeiro plano, invadindo o quadro, como em certos desenhos diante dos quais basta fecharmos os olhos e ao reabri-los a perspectiva já mudou. (...).

O vento estaria mudando? É pena que a imagem que o senhor Palomar havia conseguido organizar com tanta minúcia agora se desfigure, se fragmente e se perca. Só conseguindo manter presentes todos os aspectos juntos, ele poderia iniciar a segunda fase da operação: estender esse conhecimento a todo o universo.

Bastaria não perder a paciência, coisa que não tarda a acontecer. O senhor Palomar afasta-se ao longo da praia, com os nervos tensos como havia chegado e ainda mais inseguro de tudo. CALVINO, Ítalo. *Palomar*. São Paulo: Companhia das Letras, 1994. p.7-11.

O que vemos, por trás da simplória tarefa de observação de **uma única onda**, é a dificuldade de Palomar diante de um objeto que se transforma continuamente durante a própria atividade de observação. Ainda que Palomar defina, de modo preciso, seu objetivo e seu objeto (“observar uma simples onda e pronto”) e busque uma abrangência descritiva (“Colher todos os seus componentes simultâneos sem descurar de nenhum”), é difícil, na observação, isolar o objeto de suas “adjacências”, reduzir as diferentes instâncias do objeto a uma essência comum (“sempre ocorre alguma coisa que não tinha levado em conta”), controlar a subjetividade, suspender as pressões externas (“bastaria não perder a paciência”), encontrar o ponto de vista superior ou ideal (“Foi uma dessas línguas baixas de areia que o sr. Palomar escolheu como ponto de observação”). Enfim, definitivamente, Palomar não é bem sucedido em sua empreitada, por mais simples que esta parecesse inicialmente.

De volta à semântica, podemos nos imaginar como Palomar na tentativa de observar o significado de uma única palavra. Podemos escolher a palavra “quente”, e teremos

uma “sopa quente”, um “dia quente” e uma “cerveja quente”. Estamos tratando da mesma temperatura, do mesmo significado, o que há em comum em todos eles, e que os define? Podemos escolher outra palavra, “medo”, em “medo de altura”, “medo de perder o emprego”, “medo de barata”, “medo do mar”, “medo de sofrer”. Estamos falando exatamente do mesmo “medo”? Qual o significado exato, preciso, de “medo”? Se estamos diante de uma mesma palavra, e de uma palavra que sabemos usar, não seria esperado que soubéssemos definir, de forma clara e precisa, seu significado? Qual o significado (ou significados) de “tomar”, tomando como exemplos combinações como “tomar um susto”, “tomar um porre”, “tomar cuidado”, “tomar um suco”, “tomar remédios”, “tomar uma decisão”, “tomar conta”, “tomar um tombo”, “tomar ciência”, “tomar porrada”, “tomar dois banhos”, “tomar um susto” etc.

É exatamente este tipo de dificuldade que justifica a existência, nos estudos linguísticos, de duas grandes perspectivas que irão problematizar o que é o significado. São perspectivas concorrentes, e de um modo bastante simplificado podemos chamá-las de **representacionistas ou essencialistas**, por um lado, e de **pragmáticas** (ou, **pragmáticas radicais**), por outro (Martins, 2000, 2004). No PLN, estas visões se manifestam em diferentes maneiras de lidar com o significado: usando técnicas simbólicas (veja Semântica com técnicas simbólicas) ou usando representações distribuídas (veja Semântica distribucional), respectivamente.

A perspectiva **representacionista/essencialista** é a visão hegemônica, estando presente em boa parte dos estudos linguísticos e no senso comum – e, até recentemente, em boa parte do PLN também.

Nesta visão, palavras seriam como “substitutos” de **entidades extralinguísticas**, entidades externas à linguagem (entidades mentais, reais ou virtuais). As palavras, nessa perspectiva, importam pouco, importando mesmo as ideias (as entidades extralinguísticas) que elas **representam**. Significado e palavra são, assim, entidades distintas, ainda que relacionadas (falamos “do significado das palavras”, por exemplo), e a relação entre elas é hierárquica, com a entidade significado se sobrepondo à entidade palavra (ou à palavra e seus sinônimos), que apenas fornece matéria/forma para “hospedar” o significado.

Ainda de acordo com esta visão, apesar da multiplicidade de usos e contextos que podem existir associados a uma mesma palavra (por exemplo “tomar”, ou “quente”, ou ainda “liberdade”, “violência”, “aprender”, “significado”, “compreensão”) a comunicação é possível porque esses diferentes usos estão **associados** a uma **essência comum** (a entidade extralinguística), e por isso reconhecemos a palavra como sendo a mesma em diferentes situações. A associação entre a palavra (ou a palavra e seus sinônimos) e sua essência/ideia/conteúdo/conceito/significado, por sua vez, é guiada por regras. **Aprender uma língua**, aqui, é aprender a estabelecer a conexão entre a palavra e a entidade extralinguística que ela representa (e diferentes línguas irão variar quanto às palavras usadas para representar estes conceitos/ideias). Este conceito/ideia/significado, que algo é separado da palavra, é um “objeto” extralinguístico (do mundo mental, real, virtual) estável e com contornos bem definidos – mas que, por sua vez, também será descrito por meio de palavras.

Podemos agrupar sob esta visão – apresentada aqui de maneira muito simplificada – uma série de correntes teóricas que, de alguma maneira, compartilham a ideia de que a **estabilidade do significado** (e a compreensão) é o **resultado da representação de algo que lhe é exterior**.

Ainda segundo esta visão, os significados das palavras são, de certo modo, o que o dicionário diz. O fato de dicionários representarem os significados de maneira objetiva,

estável e discreta (vemos isso na maneira pela qual as acepções estão claramente separadas e numeradas), faz parecer que os significados das palavras se organizam “naturalmente” assim². No PLN, reconhecemos esta maneira de lidar com o significado em recursos como *wordnets*, por exemplo, que são bases de dados lexicais que contêm “nomes, verbos, adjetivos e advérbios agrupados em conjuntos de sinônimos cognitivos, **cada um representando um conceito distinto**”³ (grifo meu). Uma apresentação do que se pode fazer partindo desse ponto de vista, e de por que ele continua tendo espaço no PLN, está no Semântica com técnicas simbólicas.

Já do ponto de vista pragmático (ou, mais precisamente, pragmático **radical**)⁴, – e fazendo igualmente uma apresentação bastante simplificada – o significado de uma palavra é decorrência de situações concretas (e não o correspondente a uma entidade extralinguística), e situações concretas são variáveis. Nesta visão, os vários usos de uma palavra não se organizam em torno de um núcleo semântico comum (a entidade extralinguística), garantidor da estabilidade do que elas significam. A estabilidade do significado será sempre **provisória**, e o significado dependerá do uso, do contexto, do tempo, do espaço, de quem fala ... A **comunicação se dá no risco** (isto é, pode dar certo ou não, podemos nos entender ou não), e os mal-entendidos existem, estão aí – não são um desvio ou uma falha, são parte do jogo. O que determina se **compreendemos o significado** de um enunciado linguístico é o fato de **a manifestação dessa compreensão** (um comportamento) **ser considerada adequada** no contexto em que é produzida. Por exemplo, a um enunciado como “Está quente aqui”, seriam manifestações legítimas ações como abrir a janela ou respostas como “Não acho” ou “Por que não tira o casaco?”, entre outras. Mas dificilmente aceitaríamos como manifestações de compreensão do enunciado “Está quente aqui” dar uma cambalhota ou uma resposta como “Prefiro melão”⁵. **Aprender uma língua**, aqui, é aprender a tomar parte nas atividades humanas, um aprendizado que nunca se completa.

Assim como uma onda, os limites do significado de uma palavra não têm – aliás, podem não ter, pois não se trata de uma exigência – a precisão ou os limites definidos, necessários à formalização que sempre se buscou fazer. Segundo esta visão, o significado é flexível e maleável, não havendo uma “essência”, algo que perpassa todos os usos, e sobre o qual seja possível se sustentar, se estabilizar. No PLN, esta visão se alinha às representações distribuídas (veja Semântica distribucional)⁶.

Sabemos que uma mesma palavra pode aparecer em contextos diferentes – desde contextos completamente distintos, como “banco” e “manga”, até contextos ligeiramente diferentes, como os exemplos de “quente”, “medo” ou “tomar”, que já vimos⁷. Nesse caso, e considerando os modelos de representações distribuídas mais complexos e dinâmicos, cada forma “quente” ou “medo” será representada de uma maneira – e por isso nesses casos falamos de vetores contextuais (*contextual word embeddings*). Nos vetores estáticos, que

²Por trás de dicionários estão lexicógrafos e decisões editoriais.

³<http://wordnet.princeton.edu/>

⁴Dentre as linhas de investigação pragmáticas há as que poderiam ser também enquadradas em um paradigma representacionista. Isto porque algumas correntes da pragmática recomendam a análise das propriedades da prática da comunicação como maneira de fornecer uma explicação do que são as línguas e os significados. Por isso a especificação indicando a “radicalidade” da visão que será apresentada.

⁵Mas mesmo estas poderiam ser aceitas se assim fosse previamente estipulado.

⁶Mas não se alinha à busca do algoritmo capaz de fornecer a representação distribuída “correta”, ou “verdadeira”.

⁷É importante notar que, diferentemente do que supõe o senso comum, os casos de “banco” ou “manga”, apesar de fartamente citados como exemplos de ambiguidade, estão longe de ser prototípicos. Pelo contrário, são raros os casos em que dois sentidos se apresentam tão claramente distintos. O mais comum são casos como “quente” ou “medo”.



irão representar de uma única maneira as várias formas “quente” ou “medo”, o alinhamento à visão não-representacionista/não essencialista se mantém, uma vez que não há uma fonte (ou entidade) externa que determina o significado da palavra. Vetores produzidos a partir de conjuntos de dados diferentes irão levar a representações diferentes⁸.

Durante muito tempo, a semântica computacional esteve ancorada em visões essencialistas-representacionistas (ou simbólicas), como ilustram os capítulos “*Semantics*” de dois compêndios da área de PLN: Jurafsky; Martin (2023) e Mitkov (2003). No entanto, trabalhos de PLN que dialogam claramente com perspectivas não-essentialistas também não são novidade, como Kilgarrieff (1997); Kilgarrieff (2003) e Brewster; Wilks (2004), por exemplo.

Entre as técnicas simbólicas e as representações distribuídas existem ainda os *datasets* (ou *corpora*) com **anotação semântica** (veja Conjunto de dados, *dataset* e *corpus*), uma terceira maneira de lidar com o significado no PLN. Se, por um lado, tais *datasets* se alinham às abordagens probabilísticas, uma vez que podem ser usados como fonte para o aprendizado de máquina (para o **aprendizado do significado**), por outro lado, a atividade de anotação de significado se alinha às abordagens representacionistas. Neste tipo de anotação (também chamada de anotação de *word senses*), cada palavra (ou segmento de texto) é anotada com informação relativa ao significado de acordo com o contexto específico em que aparece no *corpus*. A informação relativa ao significado, por sua vez, vem de **fontes externas** (como dicionários, *wordnets*, *verbnet*s e *framenet*s) e a tarefa de anotação pode ser descrita como um **trabalho** de desambiguação, pois consistiria em selecionar, dentre os vários sentidos possíveis de uma palavra, aquele usado no contexto da frase. O que a anotação faz, deste modo, é criar uma **representação estável** entre a palavra e o seu significado, no contexto em que está sendo usada. Cada ocorrência de uma palavra poderá estar associada a um significado diferente (e aqui vemos uma aproximação com abordagens pragmáticas), desde que este significado esteja presente no inventário de significados usado na anotação (aqui vemos uma aproximação com abordagens representacionistas). Para as pessoas responsáveis pela anotação, a principal dificuldade está na escolha do sentido adequado conforme o contexto, uma vez que os sentidos frequentemente se sobrepõem, como as ondas observadas por Palomar.

Por exemplo, tomando a palavra “trabalho” destacada no parágrafo anterior, a tarefa consiste em escolher, dentre opções listadas no quadro abaixo, retiradas do dicionário Caldas-Aulete online⁹, aquela adequada ao contexto (se a anotação usasse o inventário de uma *wordnet* como fonte, o inventário de significados poderia ser diferente¹⁰).

Quadro 8.2: Acepções da palavra *trabalho* conforme dicionário

1. Emprego da força física ou intelectual para realizar alguma coisa
2. Aplicação dessas forças como ocupação profissional: *Seu trabalho é de gari.*
3. Local onde isso se realiza: *Mora longe do trabalho.*
4. Esmero, cuidado que se emprega na confecção ou elaboração de uma obra
5. A confecção, elaboração ou composição de uma obra
6. Obra realizada: *Essa cômoda é um belo trabalho de marcenaria.*

⁸Mas mesmo representações distribuídas podem ser associadas a visões representacionistas, quando se assume que tais representações são úteis apenas enquanto não encontramos a forma (ou a representação) correta de uma palavra.

⁹<https://www.aulete.com.br/trabalho>

¹⁰Aqui é possível consultar a OpenWordNet-PT para os significados de “trabalho”: https://www.openwordnet-pt.org/search?search_field=all&term=trabalho



7. Grande esforço; TRABALHÃO; TRABALHEIRA
8. Exercício para treino: *A professora passou muito trabalho para casa.*
9. Ação contínua de uma força da natureza e seu efeito: *O trabalho do vento resulta na erosão eólica.*
10. Med. Fenômeno orgânico que se opera no interior dos tecidos (trabalho inflamatório; trabalho de cicatrização)
11. Resultado do funcionamento de uma máquina, um aparelho etc.: *o trabalho de uma pá mecânica.*
12. Obrigação ou responsabilidade; DEVER; ENCARGO: *Seu trabalho é protegê-lo do assédio da imprensa.*
13. Econ. Conjunto das atividades humanas empregado na produção de bens: *O capital e o trabalho são os pilares da economia.*
14. Tarefa a ser realizada: *Contratou-o para um trabalho temporário.*

Fonte: (Freitas, 2022)

Parece que as acepções 1, 2 e 14 são aceitáveis no contexto da frase, o que já é um problema se precisamos escolher apenas um sentido, e por isso não é exagero dizer que as pessoas responsáveis pela anotação se sentem como Palomar na tentativa de isolar uma onda.

Corpora anotados com este tipo de informação são escassos, e um dos motivos é justamente a dificuldade de isolar o significado/conteúdo/essência das palavras enquanto estão sendo efetivamente usadas.

O estudo de Baker et al. (2017) tentou entender por que, com este tipo de anotação, era tão difícil conseguir uma boa concordância entre anotadores (veja Conjunto de dados, *dataset* e *corpus*), isto é, era tão difícil que as pessoas concordassem quanto à escolha do significado utilizado. Afinal, a tarefa é simples: associar cada palavra ao seu significado, e, se sabemos a nossa língua, sabemos o significado das palavras que usamos. No estudo, diferentes pessoas deveriam anotar as mesmas palavras, nas mesmas frases, considerando o mesmo inventário de sentidos. Os resultados indicaram uma variação bem maior que o previsto¹¹. Vamos lembrar que, na **anotação**, os significados precisam ser vistos como **unidades discretas e de conteúdo estável** – uma necessidade de ordem prática que se alinha harmoniosamente com visões representacionistas, mas que não encontra respaldo em visões pragmáticas.

O fato de representações distribuídas terem levado a resultados positivos no PLN não deve ser visto como argumento contrário às técnicas simbólicas. São maneiras diferentes de lidar com o sentido das palavras. Como tirar o melhor proveito destas diferentes visões e abordagens, no PLN, é uma das questões que se coloca. O que temos visto é a limitação de cada uma delas, tomada individualmente. Se consideramos o significado como uma entidade estável, como lidar com as mudanças, que inclusive podem ser capturadas pelos dicionários (dicionários, recentemente, mudaram a definição da palavra “família”¹²)? Por outro lado, se consideramos a instabilidade e a dependência dos dados, como evitar vieses indesejados, como a associação entre os sentidos, por exemplo, de “paraguai” e “de baixa qualidade”, quando dizemos “uísque paraguai”?

¹¹A seção “Anotações Semânticas” de Freitas (2022) traz um levantamento dos principais estudos sobre anotação semântica e seus desafios, bem como uma apresentação linguística da alternativa oferecida pelas representações distribuídas para o tratamento do significado.

¹²Dicionários mudam definição de família <https://www.metropoles.com/vida-e-estilo/comportamento/dicionario-houaiss-muda-significado-da-palavra-familia>



Os próximos capítulos aprofundam cada uma dessas maneiras de trabalhar com o significado no PLN.

Referências

BAKER, C.; FELLBAUM, C.; PASSONNEAU, R. Semantic Annotation of MASC. Em: **Handbook of Linguistic Annotation**. [s.l.] Springer Netherlands, 2017. p. 699–717.

BREWSTER, C.; WILKS, Y. **Ontologies, taxonomies, thesauri: learning from texts**. (M. Deegan, Ed.) Proceedings of Use of Computational Linguistics in the Extraction of Keyword Information from Digital Library Content Workshop. **Anais...**2004. Disponível em: <http://www.cbrewster.com/papers/KeyWord_FMO.pdf>

FREITAS, C. **Linguística Computacional**. [s.l.] Parábola Editorial, 2022.

ILARI, R.; GERALDI, J. W. **Semântica**. [s.l.] Ética, 1985.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 3rd. ed. USA: Prentice Hall PTR, 2023.

KILGARRIFF, A. I Don't Believe in Word Senses. **Computers and the Humanities**, 1997.

KILGARRIFF, A. **Thesauruses for Natural Language Processing**. Proceedings of Natural Language Processing and Knowledge Engineering. **Anais...**2003. Disponível em: <<https://www.kilgarriff.co.uk/Publications/2003-K-Beijing-thes4NLP.pdf>>

MARTINS, H. Sobre a estabilidade do significado em Wittgenstein. **Veredas**, v. 4, n. 2, p. 19–42, 2000.

MARTINS, H. Três Caminhos na Filosofia da Linguagem. Em: **Introdução à Linguística. Volume III**. [s.l.] Editora Cortez, 2004.

MITKOV, R. **The Oxford handbook of Computational Linguistics**. [s.l.] Oxford University Press, 2003.

