

**Instruções para a validação de relações semânticas entre palavras usando o VARRA -  
Validação, Avaliação e Revisão de Relações semânticas no AC/DC**  
Versão 1.1, 18 de Dezembro de 2009

Cláudia Freitas

## 1. VISÃO GERAL

O VARRA é um sistema que, por meio da ocorrência de palavras em textos, pretende auxiliar a validação de relações semânticas entre essas palavras. Cada relação é representada como uma tripla *palavra1\_relacao\_palavra2* (por exemplo “*mentira\_sinônimo\_ilusão*”).

O VARRA foi desenvolvido no âmbito do projeto AC/DC (<http://www.linguateca.pt/ACDC/>) (que oferece uma interface comum para acesso e disponibilização de corpos em português), de forma a obter julgamentos de falantes nativos mais completos em quanto às relações semânticas em questão, buscando validá-las a partir do uso das palavras em contextos autênticos, representados por frases dos corpos do projeto AC/DC.

É importante mencionar que o VARRA não está diretamente vinculado a qualquer ontologia lexical, léxico, tesouro ou dicionário. O sistema pode – e deve – ser utilizado para verificar relações semânticas advindas de quaisquer recursos, embora tenha sido inicialmente desenvolvido para validar as relações semânticas do PAPEL (<http://www.linguateca.pt/PAPEL/>), que foram automaticamente calculadas a partir de um dicionário de português de Portugal.

Com esse serviço, buscamos construir uma base confiável de julgamentos sobre uma dada relação, assim como criar uma tarefa mais parecida com a interpretação humana (em oposição à validação de relações entre palavras fora de contexto). Assim, ao invés de perguntarmos se “*mentira* é sinônimo de *ilusão*”, perguntamos se um dado contexto/frase ilustra a relação de sinonímia entre *mentira* e *ilusão*. Nos trechos abaixo, por exemplo, as respostas seriam SIM e NÃO, respectivamente.

*Mudança maior, porém, vem do novo presidente do Supremo Tribunal Federal, ministro Sepúlveda Pertence, que afirmou: ` Desde que se superou a **mentira** de que um juiz, particularmente um juiz constitucional, é um puro técnico capaz de extrair uma norma supostamente de um único sentido válido de um fato, desde que essa **ilusão** foi desfeita, a verdade é que o juiz é um homem, enquanto cidadão, com crenças, convicções, tendências conscientes e inconscientes . (Chave, AC/DC)*

*Há muitas acções que respeitadas à luz da liberdade podem não o ser quando desencadeiam a desordem, a mistificação, a **mentira**, a **ilusão**, a feitiçaria, o roubo, disse o prelado, acrescentando no entanto que como*

*bispo da Igreja Católica, a sua função era a de impedir e lutar contra toda e qualquer caça às bruxas . (Chave, AC/DC)*

A fim de termos respostas mais informativas do que simplesmente SIM / NÃO, para os casos em que o texto NÃO ilustra as relações prevemos as seguintes possibilidades (todos os exemplos se referem à relação de sinonímia entre *mentira* e *ilusão*):

**2:** O texto não valida a relação, embora seja, de alguma maneira, compatível com ela.

*A primeira vez que vi foi chocante, surpreendente, pelo jogo que faz entre verdade e **mentira**, pela discussão sobre a **ilusão** e seu poder, sobre a mistificação . (Chave, AC/DC)*

*Uma fábula sobre as relações do real e da **ilusão**, sobre a **mentira**, o tempo que volta para trás, a beleza da paternidade e o enigma da fraternidade quebrada . (Chave, AC/DC)*

*Gaspar estava a entrar na história da **ilusão**, a palavra que num dicionário de português aparece como qualquer erro proveniente de uma falsa aparência, erro de percepção, mas também com sentidos díspares e meio contraditórios: **mentira**, fantasia, traição, quimera . (Chave, AC/DC)*

**3:** O texto não valida a relação, e é completamente não relacionado

*Esta **ilusão** de ótica é tão interessante que até parece de **mentira**. (Frase extraída da internet)*

**4:** O texto não valida a relação; pelo contrário, invalida-a

*Não era uma **mentira**, era **ilusão**, o cinema possui esse poder de criar **ilusão**. (Chave, AC/DC)*

Ainda para os casos em que NÃO seja possível validar a relação pela frase exemplo, é possível simplesmente assumir que não se pode obter qualquer conclusão a partir da frase exemplo:

**5:** Não sei mesmo

*Distingue, nas promessas que lhe fazem, o que é **mentira** ou **ilusão** . (Chave, AC/DC)*

Na interface do VARRA, cada uma possíveis relações entre o texto e a tripla corresponde a um número:

- **1:** Sim
- Não
  - **2:** É compatível, mas não exatamente
  - **3:** O texto é completamente não relacionado
  - **4:** Pelo contrário, invalida-a
  - **5:** Não sei mesmo

## 2. COMO VALIDAR AS RELAÇÕES

A tarefa de validação consiste em, após a leitura da frase que ilustra a relação em questão, anotar, na coluna ao lado direito da frase, o número correspondente ao tipo de relação existente.

Existe ainda um campo, *opcional*, para que se escreva algum comentário a respeito da relação ou frase exemplo. Para a frase abaixo, relacionada à sinonímia entre *mentira* e *ilusão*,

*Distingue, nas promessas que lhe fazem, o que é **mentira** ou **ilusão** .*

o campo COMENTÁRIO poderia conter algo como

“não é possível saber se a conjunção **ou** expressa alternância (exclusão) ou adição.”

## 3. COMO UTILIZAR OS ARQUIVOS

Os arquivos que contêm as frases e as relações a serem validadas são arquivos html. No entanto, para que possam ser editados – isto é, para que cada relação seja julgada quanto a uma determinada frase (cf. Secção 2) é preciso antes que o arquivo seja salvo como um documento do tipo Word. Assim, antes de iniciar o processo de validação propriamente siga os passos seguintes:

Para abrir o arquivo recebido:

- a) clique com o botão direito do mouse;
- b) selecione “Abrir com”
- c) selecione “Word”

Com isso, é possível trabalhar normalmente como se trabalha com um arquivo Word.

Para SALVAR o arquivo:

- a) Em “Arquivo”, selecione “Salvar como”
- b) Em “Nome do Arquivo:”, acrescente, ao nome que está lá, `_seunome`
- c) Em “Salvar como tipo:”, selecione “Documento do Word”

## 4. SOBRE AS RELACOES SEMÂNTICAS:

De uma maneira bastante simplificada, podemos definir as relações semânticas envolvidas na validação da seguinte maneira:

A relação de SINONÍMIA envolve palavras fonologicamente distintas, mas de significados idênticos ou muito próximos.

Menino sinônimo de garoto

Na delegacia, o lavrador disse que não percebeu que o **garoto** havia entrado embaixo do trator e que, quando viu o **menino**, já era tarde . (Chave, AC/DC)

Alimento sinônimo de comida

O **alimento** baixa a mortalidade infantil, mas não podemos sustentar o programa só com a distribuição de **comida**, disse Peliano. (Chave, AC/DC)

A relação de HIPONÍMIA é uma relação de inclusão: o nome incluído é o hipônimo e, de maneira inversa, o nome mais abrangente é o hiperônimo:

Cachorro, escorpião e cobra são hipônimos de animal (e animal é hiperônimo de cachorro, gato e baleia)

Quase qualquer **animal** pode virar um prato: **escorpião** frito, sopa de **cachorro** ou assado de **cobra**. (Chave, AC/DC)

Flauta é hipônimo de instrumento

Eu, como é óbvio, como é que posso ser fã de música eletrônica, tocando um **instrumento** como a **flauta** de bambú ? (Chave, AC/DC)

Ódio é hipônimo de sentimento

Todo mundo sabe que o **ódio**, quando explode, é um **sentimento** purificador . (Chave, AC/DC)

Sarampo é hipônimo de doença

O **sarampo** é uma **doença** contagiosa que ocorre na infância, provocada por um vírus . (Chave, AC/DC)

A relação de PARTE\_DE (também chamada relação de MERONÍMIA) envolve uma parte e o todo que a compõe:

Memória é parte\_de computador

Todo **computador** sai de fábrica com **memória** RAM . (Chave, AC/DC)

Leite é parte\_de iogurte

Depois, você vê como o **leite** se transforma em **iogurte** e manteiga. (Chave, AC/DC)

Bico é parte\_de galinha

Minha **galinha** teria **bico** e penas como todas, com a diferença de que traria uma novidade: a descoberta científica sobre seus dentes. (Chave, AC/DC)

Para uma descrição detalhada das relações semânticas, sugerimos a leitura de John Lyons, *Semântica*. Lisboa, Presença/Martins Fontes, 1980. 2v.

A seguir, apresentamos 3 arquivos, validados por alguém chamado João. O [primeiro](#) arquivo exemplifica a relação SINÔNIMO\_DE; o [segundo](#), a relação PARTE\_DE e o [terceiro](#), a relação HIPERÔNIMO\_DE.