

**Instruções para a validação de relações semânticas entre palavras usando o VARRA -
Validação, Avaliação e Revisão de Relações semânticas no AC/DC**
Versão 2, 29 de Novembro de 2010

Cláudia Freitas

- I. Visão Geral
- II. Alternativas de validação do VARRA
- III. Instruções aos validadores
 - III. 1. Como validar as relações
 - III. 1.2. O campo Comentários
 - III. 2. Como guardar (salvar) os resultados

I. VISÃO GERAL

O VARRA é um sistema que pretende auxiliar a validação de relações semânticas entre pares de palavras considerando sua ocorrência em contextos de uso – em textos. Cada relação é representada como uma tripla *palavra1 relação palavra2* (“*menino SINONIMO_DE garoto*”).

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 precisos 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, por exemplo, 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 ações 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)*

No entanto, a fim de termos respostas mais informativas do que simplesmente SIM / NÃO, para os casos em que o texto NÃO ilustra as relações prevemos 4 possibilidades de resposta. A seção a seguir exemplifica e explica todas as alternativas de validação do VARRA.

II. ALTERNATIVAS DE VALIDAÇÃO DO VARRA

O VARRA prevê 5 diferentes alternativas para dar conta das relações entre os pares de palavras e os textos que ilustram as relações. Estas alternativas correspondem a 5 possíveis respostas à pergunta “*Os textos dos exemplos ilustram a relação entre as duas palavras apresentada na primeira coluna?*”

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

1. SIM.

Essa deve ser a alternativa escolhida quando o texto ilustrar a relação-alvo entre os pares de palavras.

Exemplos:

1a - Relação a ser validada: feijão PARTE_DE feijoada

Texto: *Sábado é o dia de Tsholent, um tipo de **feijoada** judaica feita com **feijão** branco, batatas e carne de boi, no Cecília.*

1b - Relação a ser validada: fruta HIPERONIMO_DE abacaxi

Texto: *Bebidas típicas são à base de rum e **frutas** tropicais, como o **abacaxi** e o **coco**.*

2. NÃO. É compatível com a relação, mas não a exemplifica.

Essa deve ser a alternativa escolhida quando, conhecendo a relação-alvo e o par de palavras envolvido, a frase se encaixa, isto é, é compatível com a relação, embora não a ilustre.

Exemplos:

2a - Relação a ser validada: grana SINONIMO_DE dinheiro

Texto: *Um cidadão cheio da **grana**, desesperado dos conselhos de seus assessores, ousou me consultar a respeito de um problema que eu nunca tive: em caso de reviravolta social, o que ele deveria fazer para perder **dinheiro** rápida e honestamente.*

Explicação: Nesse exemplo, os termos em negrito são sinônimos, mas a frase não ilustra, necessariamente, a relação de sinonímia entre **grana** e **dinheiro**, embora seja compatível com ela. Isto é, apenas pela frase, não podemos ter certeza da relação de sinonímia envolvida. Poderíamos, por exemplo, substituir um dos termos por outros não sinônimos, sem perda de coerência, como a seguir: *“Um cidadão cheio de **preocupações**, desesperado dos conselhos de seus assessores, ousou me consultar a respeito de um problema que eu nunca tive: em caso de reviravolta social, o que ele deveria fazer para perder **dinheiro** rápida e honestamente”*

2b- Relação a ser validada: raio PARTE_DE tempestade

Texto: *Malcolm começa o jogo na lixeira de Kyrandia, onde um **raio**, numa noite de **tempestade** -- como convém numa estória de fantasia -- o acorda do seu sono pétreo.*

Explicação: Embora a frase do exemplo se encaixe na relação – porque conhecemos o significado de **raio** e **tempestade** – não chega a exemplificá-la. Poderíamos ter, por exemplo, *“Malcolm começa o jogo na lixeira de Kyrandia, onde um **pássaro**, numa noite de **tempestade** -- como convém numa estória de fantasia -- o acorda do seu sono pétreo.”*, em que não há relação entre pássaro e tempestade.

2c - Relação a ser validada: feijão PARTE_DE feijoada

Texto: *Depois de passar um dia tentando preparar uma **feijoada**, com um pedaço de aipo na boca e toda suja de **feijão**, a atriz se rende e diz: Não sei se a feijoada vai dar certo, mas a cerveja eu garanto.*

Explicação: Assim como nos exemplos anteriores, podemos estabelecer a relação devido ao nosso conhecimento prévio do significado de **feijoada** e **feijão**. No entanto, se a frase fosse *“Depois de passar um dia tentando preparar uma feijoada, com um pedaço de aipo na boca e toda suja de **pão**, a atriz se rende e diz: Não sei se a feijoada vai dar certo, mas a cerveja eu garanto.”*

IMPORTANTE:

- a) Ainda com relação às respostas do tipo **2**, frequentemente usamos termos hiperônimos para dar coesão ao texto, em uma relação de anáfora – e, nesses

casos, eles acabam por funcionar também como sinônimos. Os exemplos **2d** e **2e** ilustram esse fenômeno.

2d. Relação a ser validada: fruta HIPERONIMO_DE abacaxi

Texto: *A região de Bauru (345 km a noroeste de São Paulo) é a maior produtora de **abacaxi** de sobremesa do Estado, com 600 hectares da **fruta** plantados.*

2e Relação a ser validada: aparelho HIPERONIMO_DE telefone

Texto: *Por não ter sido bem identificado ao **telefone**, a pessoa da TV que o atendeu não acreditou que fosse uma autoridade da Censura e desligou o **aparelho**.*

Ou seja, nesses casos, a frase poderia ser usada para, **ERRADAMENTE**, validar a relação de *sinonímia* entre *abacaxi – fruta* e entre *telefone – aparelho*, respectivamente, uma relação que claramente vai contra o conhecimento do falante. Para essas situações, que imaginamos frequentes na relação HIPERONIMO_DE, a alternativa marcada deve ser 2 – sabemos que a frase exemplifica a relação, embora não a prove – e, nesse caso, o campo **Comentário** (cf. Seção III.1.2) deve ser preenchido com "*também compatível com sinonímia*" (ou, simplesmente, **TCS**).

IMPORTANTE:

- b) É possível que, em alguns casos, decidir entre uma resposta do tipo 1 e uma do tipo 2 seja uma questão de interpretação pessoal.

3. NÃO. O texto é completamente não relacionado.

Essa deve ser a alternativa escolhida quando, a partir do texto, não é possível concluir nem que a relação é válida nem que a relação é inválida. Isso pode acontecer porque alguma das palavras está sendo usada num sentido que não o indicado pelo par.

3a.Relção validada: fruta HIPERONIMO_DE manga

Texto: *Durante o banquete, Nasrudin ia jogando as **frutas** e a comida pela **manga** da túnica.*

Explicação: No exemplo, a palavra **manga** está sendo usada com o sentido de parte de uma roupa, e por isso não é possível validar a relação.

3b.Relção a ser validada: roda PARTE_DE carro

Texto: O novo **carro** médio da Fiat já **roda** no país

Explicação: No exemplo, a palavra **roda** refere-se ao verbo “rodar”, e não à parte do carro chamada **roda**.

A alternativa **3** também deve ser usada quando, mesmo compreendendo a frase, as palavras e a relação, a frase em si não permite (isto é, não é relevante) a validação de que se está à procura. Nos exemplos abaixo, não é possível estabelecer qualquer relação entre **feijão** e **feijoada** (3c); **fruta** e **abacaxi** (3d), **dedo** e **mão** (3e), **aparelho** e **telefone** (3f).

3c - Relação a ser validada: feijão PARTE_DE feijoada

Texto: *Mesmo assim, aí vai uma lista: vários bacalhaus (cozido, cru, assado e arroz, sopas (de alheiras, de tomate, de grão, de batata, de alho, de **feijão**, etc.), arroz de pato, caldeirada (de peixe e de cabrito), moamba de galinha, cabrito assado nas brasas, perdiz (várias receitas) e **feijoada** de lebre.*

3d. Relação a ser validada: fruta HIPERONIMO_DE abacaxi

Texto: *Ensina a fazer fofinho de chocolate (rocambole), **frutas** com gelatina, friturinha de maçã, coroa de **abacaxi** e cajuzinho*

3e. Relação a ser validada: dedo PARTE_DE mão

Texto: *A mão autoritária e o **dedo** caríssimo do decorador-estrela viraram cafona.*

3f. Relação a ser validada: aparelho HIPERONIMO_DE telefone

Texto: *As informações são trocadas pelo modem (**aparelho** que permite enviar e receber informações pelo **telefone** usando um computador) com outras escolas participantes do convênio do Grupo com a Escola do Futuro.*

IMPORTANTE: A alternativa **3** também engloba exemplos em que não se consegue validar a relação porque pelo menos uma das palavras envolvidas faz parte de expressões maiores, isto é, pelo menos uma das palavras do par integra o que chamamos expressões multivocabulares (EMV). Por exemplo, para a relação adega PARTE_DE casa, pensamos que “casa lotérica”, “casa de banho” (“banheiro”, em Portugal) ou “casas agrícolas” não devem ser usadas **em contexto** para decidir sobre “casa”. Nesses casos, o campo **Comentário** (cf. Seção III.1.2) deve ser preenchido por “expressão multivocabular” (ou, simplesmente, **EMV**).

Exemplos:

3g. Relação a ser validada: roda PARTE_DE carro

Texto: *A lei obriga a ter no **carro** chaves de **roda** e de fenda, macaco, triângulo e extintor de incêndio.*

Explicação: chave de roda é EMV

3h. Relação a ser validada: adega PARTE_DE casa

Texto: *Na terça-feira, em Samora Correia, responsáveis da RTR reuniram-se com proprietários de **casas** agrícolas, **adegas** e representantes de diversas câmaras municipais, para deliberar a distribuição dos 40 mil contos que vão ser gastos no projecto.*

Explicação: casas agrícolas é EMV

4. NÃO. O texto não valida a relação; pelo contrário, invalida-a

A alternativa **4** deve ser usada quando não apenas é impossível validar a relação a partir do texto, mas quando o exemplo contraria, ou invalida, a relação.

4a. Relação a ser validada: mentira SINONIMO_DE ilusão

Texto: *Não era uma **mentira**, era **ilusão**, o cinema possui esse poder de criar ilusão.*

5: Não sei

A alternativa 5 deve ser usada quando o validador não consegue perceber a relação expressa no texto.

5a. Relação a ser validada: mentira SINONIMO_DE ilusão

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

A alternativa **5** se aplica, também, aos casos em que há dúvida entre as opções, e no **Comentário** (cf. Seção III.1.2), é possível explicar o motivo da dúvida.

Por fim, também é possível escolher a alternativa **5** porque o texto é incompreensível.

Lembramos que a alternativa **5** (“Não sei”) difere da **3** (“O texto é completamente não relacionado”) pois, ao escolher a **3**, o validador sabe porque não é possível validar a relação pelo texto – porque, no texto, não há relação entre as palavras, ou porque as palavras estão sendo usadas em um sentido que não permite o estabelecimento da relação em questão - e, ao escolher a alternativa **5**, o validador não sabe bem o que fazer.

III. INSTRUÇÕES AOS VALIDADORES

III.1. COMO VALIDAR AS RELAÇÕES

A tarefa de validação é feita com a leitura e análise das relações, em contexto. Cada validador recebe um arquivo que pode ser aberto em qualquer programa editor de textos, como o Word, que contém as relações, os contextos e o campo para preenchimento.

No entanto, antes de começar o processo de validação propriamente, isto é, antes de receber o documento com as relações, cada validador receberá uma lista com as relações que serão validadas SEM CONTEXTO. A ideia é que a lista seja preenchida com base na intuição, para que depois possamos também comparar as intuições que temos com relação ao comportamento das palavras com os valores (ou significados) que elas efetivamente assumem em contexto (figura 1).

Após o preenchimento dessa lista (que não deve levar mais que 10 minutos), os validadores receberão o dossiê completo para validação, com os termos, as relações e os contextos (figura 2).

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 existente¹.

Antes de ver os contextos, acha que as relações abaixo são

(a) Correta
(b) Incorreta
(c) às vezes correta outras vezes incorreta
(d) Não sabe.

Por favor não mude esta resposta mesmo que tenha mudado de opinião depois da validação.

Relação a validar	Julgamento sem contexto
abdômen PARTE_DE corpo	
adega PARTE_DE casa	
agulha PARTE_DE bússola	
aluno PARTE_DE escola	
anel PARTE_DE véu	
avião PARTE_DE panelha	
aço PARTE_DE espada	
batalhão PARTE_DE regimento	
berma PARTE_DE plataforma	
bico PARTE_DE arara	
bico PARTE_DE codorniz	
bandolim PARTE_DE choro	
bico PARTE_DE colheiro	
bico PARTE_DE pelicano	
bico PARTE_DE pena	
bigode PARTE_DE barba	
braço PARTE_DE alavanca	
braço PARTE_DE membro	
broca PARTE_DE berbequim	
buffer PARTE_DE memória	
cabelo PARTE_DE medalhão	

Figura 1 - lista de relações sem contexto

III.1.2 O campo Comentário

Em cada dossiê existe ainda um campo chamado *Comentário*, de preenchimento **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**.*

Uma resposta possível é a 5 (*Não sei*), e o campo COMENTÁRIO poderia conter algo como

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

Além disso, é possível que um dado exemplo ilustre mais que uma relação. Se for este o caso, escolha a alternativa adequada (e imaginamos que seja 1 ou 2) e, no campo Comentário, indique qual a OUTRA relação que também está exemplificada.

Por fim, como já mencionado nas seções anteriores, há dois casos em que **enfatizamos a relevância do preenchimento** do campo Comentário:

- a) Quando, na validação da relação de HIPERONIMO_DE, a relação entre as palavras também for indicativa de sinonímia. Nesse caso, deve-se marcar a

¹ Pessoalmente, acho que facilita a leitura (e o preenchimento do dossiê) ativar, no Word, o modo de visualização que é “Layout de Web”. Para tanto, no menu superior do Word, vá em “Exibição” e selecione a opção “Layout de Web”. (C.F.)

relação como **2** (pois a frase exemplifica a relação, mas não a prova), e o campo Comentário pode ser preenchido com **TCS** ("*também compatível com sinonímia*"). (conferir explicação referente à alternativa **2**.)

NOME:

As colunas abaixo apresentam uma relação semântica entre dois termos; o código usado na procura por esses termos no corpo "Colecção CHAVE v. 4.3"; os exemplos de ocorrência desses termos na mesma sentença encontrados; um espaço para suas respostas e seus comentários.

Leia os exemplos e complete as colunas **Resposta** e **Comentário**.

Os textos dos exemplos ilustram a relação entre as duas palavras apresentada na primeira coluna?

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

Para cada linha, escolha uma das possibilidades 1 a 5, e comente se achar necessário.

Relação	Procura	Exemplo	Resposta (1-5)	Comentário
abdómen PARTE_DE corpo	MU meet abdómen corpo s	<i>docid= PUBLICO-19940316-062:</i> Segundo informações dos Bombeiros de Avintes, que procederam à remoção dos corpos, juntamente com os Sapadores de Gaia, um terceiro operário, que se encontrava dentro da máquina escavadora a controlar as manobras, desmaiou na altura em que os bombeiros retiravam o corpo de Adriano Pereira, que, apesar de não ter sido completamente soterrado, tinha sofrido grandes ferimentos na zona do abdómen .		
abdómen PARTE_DE corpo	MU meet abdómen corpo s	<i>docid= PUBLICO-19940610-138:</i> Vinte facadas no pescoço e abdómen , após uma luta corpo a corpo num terreno baldio, acabam com a vida de Carlos Moreno Fernandes, de 52 anos, casado, empregado de limpeza.		
abdómen PARTE_DE corpo	MU meet abdómen corpo s	<i>docid= PUBLICO-19950130-058:</i> Duas das vítimas sofreram apenas cortes superficiais, a terceira, identificada com o apelido Portas, deu entrada no Hospital de Santo António, apresentando perfurações por todo o corpo e uma de particular gravidade no		

Figura 2- dossiê para validação

b) Quando pelo menos uma das palavras do par fizer parte de uma expressão multivocabular, como "chave" em "chave de fenda". Nesse caso, deve-se validar a relação com a alternativa **3** e o campo Comentário pode ser preenchido com **EMV**. (conferir explicação referente à alternativa **3**.)

IMPORTANTE: Ao escrever no campo Comentário, por favor, **NÃO UTILIZE "Enter"** para introduzir um novo parágrafo ou quebra de linha. Deixe seus comentários como um único parágrafo (do contrário, teremos problemas com o programa que processa os dados dos dossiês).

A figura 3 ilustra um dossiê preenchido.

III. 2. COMO GUARDAR (salvar) OS RESULTADOS

Após o preenchimento dos dossiês, é importante salvar os dados e enviá-los para a equipe do VARRA, para que os dados possam ser contabilizados.

Todos os dossiês estão no formato RTF, o que permite serem utilizados pela maioria dos programas processadores de texto. Terminado o processo de validação, é importante salvar o arquivo, MANTENDO o mesmo formato RTF.

Quando terminar o preenchimento de um dossiê, envie-o para let1910@gmail.com com o campo ASSUNTO preenchido com “dossie do VARRA”.

Leia os exemplos e complete as colunas **Resposta** e **Comentário**.

Os textos dos exemplos ilustram a relação entre as duas palavras apresentada na primeira coluna?

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

Para cada linha, escolha uma das possibilidades 1 a 5, e comente se achar necessário.

Relação	Procura	Exemplo	Resposta (1-5)	Comentário
mentira SINONIMO_DE ilusão	MU ment mentira ilusão s	<i>docid= FSP940826-029:</i> E fácil, portanto, aplicar um detector de mentiras nas campanhas, considerando o previsível coquetel de ilusões que será oferecido para gerar empregos .	2	
mentira SINONIMO_DE ilusão	MU ment mentira ilusão s	<i>docid= FSP951221-004:</i> Poesia romena, música afegã, dança belga, tanto faz: o importante é propagar as ilusões do exótico, as mentiras da autenticidade .	3	
mentira SINONIMO_DE ilusão	MU ment mentira ilusão s	<i>docid= PUBLICO-19940512-132:</i> Não era uma mentira , era ilusão , o cinema possui esse poder de criar ilusão . "	4	
mentira SINONIMO_DE ilusão	MU ment mentira ilusão s	<i>docid= PUBLICO-19950423-084:</i> Distingue, nas promessas que lhe fazem, o que é mentira ou ilusão .	5	Não é possível saber se "ou" expressa alternância ou adição

Figura 3- Exemplo de um dossiê preenchido