



## Geração de língua natural para descrição de objetos em contextos visuais

Ivandr  Paraboni (USP / EACH)

**Resumo:** Este documento apresenta uma proposta de projeto de pesquisa na  rea de Gera o de L ngua Natural (GLN), enfocando o problema de descri o de objetos em contextos visuais. Em especial, ser o enfocados tr s desafios computacionais em aberto nesta  rea, e que s o foco de pesquisas atualmente em andamento nesta institui o: (a) o uso de descri es relacionais, (b) a gera o de descri es com diferentes graus de especifica o, e (c) a quest o da varia o humana na produ o destas descri es. Para este fim, o projeto prev  a coleta e anota o de um recurso lingu stico-computacional independente de l ngua e de larga escala - um corpus de descri es de objetos em um dom nio visual - e o uso deste recurso para proposta de um modelo computacional que estende trabalho pr vio relacionado  s quest es (a-c). O corpus constru do ser  disponibilizado para reuso pela comunidade cient fica em geral, e deve assim servir de refer ncia para a pesquisa na  rea de GLN aos moldes de recursos de alto impacto como os corpora TUNA, GRE3D7 e similares. O modelo proposto deve avan ar o estado da arte no problema computacional da descri o de objetos em dom nios visuais, com aplica es pr ticas em  reas como a de sistemas baseados em mundos virtuais interativos e outros.

**Palavras-chave:** Gera o de L ngua Natural, Sele o de Conte do, Contextos Visuais.

# 1 Introdução

## 1.1 O problema computacional da geração de descrições de objetos visuais em língua natural

Sistemas de geração de língua natural (GLN) - que produzem descrições textuais a partir de uma entrada de dados geralmente não linguística - são empregados quando o uso de texto predefinido não é suficiente, ou seja, quando é necessária uma maior variação linguística nos documentos gerados e/ou maior proximidade em relação ao desempenho humano. Aplicações de GLN incluem sistemas de diálogo humano-computador em língua natural, geração de relatórios a partir de bases de dados, sumarização de documentos WEB e muitas outras.

Com o avanço em áreas da ciência da computação como processamento gráfico e interação humano-computador, problemas de geração de língua natural até então puramente textuais encontraram domínios mais realistas - e efetivamente novas aplicações - em diversas frentes. Este é o caso, por exemplo, de sistemas de geração de instruções de navegação em língua natural em contextos visuais como *GIVE* (*Generating Instructions in Virtual Environments*) (Byron et al., 2007) e, mais recentemente, *Gruve* (*Giving Route instructions in Uncertain Virtual Environments*) (Janarthanam et al., 2012). Um exemplo deste tipo de aplicação é ilustrado na Figura 1, na qual o sistema gera uma instrução de navegação (na parte inferior da tela) que inclui a descrição de um objeto do mundo virtual (o prédio mais alto, ao fundo).



**Figura 1** – Geração de descrições no ambiente Gruve (Janarthanam et al., 2012).

Neste documento apresentamos uma proposta de pesquisa que trata do problema computacional de *seleção de conteúdo* em aplicações deste tipo, ou seja, da tarefa de descrever objetos em contextos visuais com as quais o usuário deve interagir. Além de compartilhar vários dos desafios típicos da geração de língua a partir de texto (Novais e Paraboni, 2012), a seleção de conteúdo em contextos visuais apresenta ainda uma série de questões particulares, algumas das quais documentadas em trabalho prévio (Paraboni et al., 2007; Paraboni e Deemter, 2013), e cujo aprofundamento é o foco da presente proposta de pesquisa. De forma mais específica, este projeto procura avançar três aspectos do problema computacional de descrição de objetos visuais que são atualmente tema de pesquisas nesta instituição, a saber:

- (a) O uso de descrições relacionais.
- (b) A geração de descrições com diferentes graus de especificação.
- (c) A variação humana na geração destas descrições.

Descrições relacionais - que envolvem relações semânticas entre o objeto-alvo e um ponto de referência como ‘o computador *ao lado da porta*’ - são abundantes em situações em que o uso de propriedades atômicas do próprio alvo (e.g., cor, tipo, tamanho etc.) não são suficientes. O uso de descrições deste tipo, com as envolvendo relações *espaciais* entre objetos, é particularmente importante em domínios visuais complexos como os encontrados em aplicações baseadas em mundos virtuais interativos (Byron et al., 2007; Janarthanam et al., 2012). Não por acaso, a decisão de quando e como usar propriedades relacionais ao descrever um objeto nestes domínios permanece um dos grandes desafios em aberto na área (Viethen e Dale, 2011).

O grau de especificação de uma expressão de referência é provavelmente o tema mais recorrente na pesquisa sobre seleção de conteúdo para GLN desde as primeiras formulações algorítmicas para o problema (Dale e Reiter, 1995). Descrições produzidas por participantes humanos do discurso normalmente exibem grande variação de volume de informação representada, cobrindo todo o espectro de referências subespecificadas (i.e., com algum grau de ambiguidade em um ou mais de seus termos) até descrições altamente superespecificadas (i.e., com algum grau de redundância lógica, seja de forma espúria ou intencional). Considerando-se que uma descrição gerada de forma automática deva ser psicologicamente plausível, isso é, aproximar-se tanto quanto possível da expressão que um participante humano produzira na mesma situação, coloca-se assim o desafio de como produzir um modelo computacional para geração de descrições com o grau ‘adequado’ de sub ou superespecificação (Krahmer e Deemter, 2012).

Finalmente, observa-se que não existe uma estratégia de produção de língua dita ‘universal’, ou nem mesmo uma estratégia única para cada locutor. Pessoas diferentes descrevem um mesmo objeto de formas diferentes, e a mesma pessoa pode descrever um objeto de formas diferentes

em momentos distintos (i.e., mesmo em situações idênticas). Desenvolver modelos computacionais de geração de descrições que sejam capazes de se adaptar à estratégia de referência de um indivíduo ou grupo de indivíduos permanece também uma questão de pesquisa em aberto (Viethen et al., 2013). Além disso, as questões anteriores (i.e., o uso de descrições relacionais e o grau de sub/superespecificação da descrição) são, ao menos em parte, influenciadas pela variação humana, e o estudo destas questões sob a ótica da variação humana é assim uma condição para seu melhor entendimento.

Assim como em diversas outras linhas de pesquisa do processamento de línguas naturais (PLN), a investigação de questões como acima (a-c) envolve a observação empírica do uso da língua para treinamento e teste de artefatos computacionais. Em outras palavras, também no caso da seleção de conteúdo para GLN faz-se necessário o uso de corpora representativos do fenômeno investigado.

Por outro lado, e diferentemente de outras áreas do PLN, o uso de corpora de propósito geral (e.g., uma coleção de artigos jornalísticos como o corpus NILC (Nunes et al., 1996)) em GLN é limitado pelo fato de que o conhecimento a partir do qual o texto foi produzido normalmente não se encontrar disponível. Em outras palavras, o texto que encontramos em um corpus de propósito geral é apenas o produto final de um processo de produção humana da língua, e que normalmente traz pouca informação sobre o processo em si. Além disso, aspectos relevantes do processo, mesmo que presentes, tendem a ocorrer misturados a diversos outros fenômenos, e permanecem assim fora do controle do pesquisador.

Na pesquisa em GLN e áreas afins como psicolinguística, ciências cognitivas etc., a forma usual de reproduzir condições de produção de língua natural é a realização de experimentos controlados com uso de participantes humanos. Experimentos deste tipo fornecem um certo tipo de estímulo textual ou visual aos participantes - que podem assumir o papel de falante, ouvinte ou ambos - e registram suas reações, geralmente manifestas na forma oral, escrita, ou ainda por ações de interação com o ambiente computacional. Em outras palavras, para pesquisa e validação de vários tipos de sistemas e artefatos de GLN, faz-se necessário examinar não apenas o texto resultante do processo, mas também modelar as condições contextuais nas quais este texto foi produzido. Dependendo do objetivo da pesquisa, isso pode envolver, por exemplo, a definição do conteúdo que estava em discussão no momento em que a sentença foi produzida, do objetivo da comunicação, do conhecimento prévio do falante, do seu estado de atenção etc.

Experimentos de GLN que resultaram em conjuntos de dados atualmente considerados referência na área incluem, por exemplo, o corpus TUNA (Gatt et al., 2007), GRE3D7 (Viethen e Dale, 2011), GIVE-2 (Gargett et al., 2010) e outros. Vários destes recursos e técnicas associadas têm sido empregadas regularmente na pesquisa realizada pelo proponente deste projeto. Por exemplo, experimentos controlados para validação de modelos computacionais de GER foram re-

alizados em (Paraboni e Deemter, 2006; Paraboni et al., 2006; Lucena e Paraboni, 2009; Araujo et al., 2010; Paraboni et al., 2007; Paraboni e Deemter, 2013), e o reuso de corpora de GLN já existente (Lucena et al., 2010; Pereira et al., 2012) ou coletado a partir da WEB (Cuevas e Paraboni, 2008; Cuevas et al., 2008; Aziz et al., 2008; Novais et al., 2011; Novais e Paraboni, 2012) tem sido amplamente utilizado para fins de avaliação de artefatos deste tipo.

De modo geral, entretanto, cada um destes recursos é normalmente construído com um propósito altamente específico<sup>1</sup>. Os corpora existentes tendem assim a limitar o escopo das conclusões possíveis sobre outros tipos de fenômeno além daqueles originalmente contemplados em seus experimentos de origem, e já exaustivamente discutidos em seus respectivos projetos de pesquisa. Por estes motivos, e apesar da inegável contribuição de recursos deste tipo para a pesquisa na área, novas questões de pesquisa tendem a exigir recursos com características ainda não oferecidas.

Do ponto de vista da presente pesquisa, as questões (a-c) assim compartilham - além do interesse comum no problema computacional na geração de descrições em língua natural - uma mesma necessidade: o suporte de um conjunto de dados de referência de alta qualidade, possivelmente desenvolvido em modo colaborativo (i.e., envolvendo situações de comunicação entre dois ou mais participantes humanos), e que atenda três requisitos: (a) envolver situações complexas de referência em um domínio visual, estimulando assim o uso de descrições relacionais; (b) contemplar situações de referência que exijam diferentes graus de especificação; e (c) envolver um número considerável de participantes de modo que suas diferentes estratégias de referência sejam observáveis.

Conforme será discutido na Seção 3.5, é razoável supor que recursos de GLN existentes não atendem de forma satisfatória todos estes requisitos. Diante deste cenário, apresentamos neste documento uma proposta de projeto de pesquisa em geração de língua natural enfocando a seleção de conteúdo em contextos visuais. O projeto prevê a construção de um novo recurso linguístico-computacional de larga escala privilegiando as questões (a-c) acima, e o uso deste recurso para extensão do trabalho em (Paraboni et al., 2007; Paraboni e Deemter, 2013) na forma de um novo modelo computacional abordando estas questões.

## 1.2 Objetivos

O objetivo da pesquisa é construir um corpus produzido em modo colaborativo contendo descrições de objetos em contextos visuais, e utilizá-lo na construção de um modelo computacional de seleção de conteúdo para geração de descrições deste tipo.

---

<sup>1</sup>Por exemplo, o corpus TUNA (Gatt et al., 2007) enfoca o caso de descrições não-relacionais, ou seja, contendo apenas propriedades atômicas como tipo, cor etc.

De forma mais específica, o corpus a ser construído será anotado de modo independente de língua, e deve oferecer suporte à proposta de um novo modelo computacional baseado na proposta em (Paraboni e Deemter, 2013), contemplando a geração de descrições relacionais com diferentes graus de especificação, e levando em conta a questão da variação humana inerente a este processo.

### 1.3 Justificativas

Considerando-se que a pesquisa em GLN é ainda relativamente incipiente se comparada à grande área de PLN, torna-se assim fundamental não apenas reutilizar recursos desenvolvidos por terceiros, mas também contribuir com iniciativas de pesquisa básica para disponibilização de novos corpora de referência para a área. Assim, acreditamos que a disponibilização de um novo corpus de descrições de objetos visuais de maior cobertura do fenômeno, e envolvendo um número elevado de participantes, seja um requisito para o avanço do estado da arte, e uma forma importante de fomento à pesquisa na área.

Além da necessidade de corpora de GLN, as questões teóricas que motivam a presente proposta são igualmente atuais. A geração de descrições de objetos em língua natural é um componente essencial de sistemas de GLN baseadas em mundos virtuais interativos, com aplicações diretas no desenvolvimento de jogos, software educacional etc., e o amadurecimento das tecnologias aqui envolvidas requer soluções algorítmicas de maior cobertura para diversos problemas em aberto. Estes problemas incluem, de forma obviamente não exaustiva, as três questões de pesquisa aqui destacadas, as quais são amplamente documentadas na literatura da área, e atualmente foco de estudos nesta instituição.

## 2 *Resultados esperados*

O projeto prevê a disponibilização de dois produtos principais, a saber:

- (1) Um corpus anotado de modo independente de língua, contendo descrições produzidas em modo colaborativo, e privilegiando as questões do uso de descrições relacionais, a geração de descrições envolvendo diferentes graus de especificação, e a variação humana.
- (2) Um novo modelo computacional de GER contemplando as questões propostas.

O recurso linguístico-computacional construído será disponibilizado para reuso pela comunidade científica em geral, podendo servir de referência para a pesquisa na área de GLN aos moldes de iniciativas de alto impacto como os corpora TUNA (Gatt et al., 2007), GIVE-2 (Gargett et al., 2010), GRE3D7 (Viethen e Dale, 2011) e outros. Este potencial de reutilização torna-se mais evidente se considerarmos que o recurso a ser desenvolvido é independente de língua, ou seja, embora o corpus seja produzido a partir de um experimento envolvendo falantes do Português, o produto final consiste de um conjunto de anotações semânticas potencialmente aplicáveis à geração de qualquer língua-alvo, do mesmo modo que corpora produzido por falantes do inglês (Gatt et al., 2007; Viethen e Dale, 2011) e alemão (Gargett et al., 2010) têm sido utilizados em nossa própria pesquisa envolvendo a geração de descrições em Português (Lucena et al., 2010; Silva e Paraboni, 2013b).

O modelo computacional resultante deste estudo deve avançar o estado da arte no problema computacional da geração de descrições de objetos em domínios visuais, ampliando o conhecimento atual sobre como sistemas de GLN devam ser construídos e avaliados. Este conhecimento deverá permitir o desenvolvimento de algoritmos de GER linguisticamente mais motivados, capazes de gerar descrições relacionais, com grau adequado de especificação e levando em conta a variação humana na escolha da estratégia de referência.

De forma mais específica, o projeto também auxiliará na consolidação da linha de pesquisa em GLN na USP-EACH, dando continuidade e/ou complementando estudos como em (Paraboni et al., 2007; Lucena et al., 2010; Novais e Paraboni, 2012; Paraboni e Deemter, 2013). Além disso, à luz do recente programa de pós-graduação em Sistemas de Informação (PPgSI) da mesma, espera-se engajar pela primeira vez estudantes de mestrado nesta linha de pesquisa, bem como estreitar os laços de colaboração com a instituição parceira no exterior com a qual compartilhamos interesses neste tema.

## 3 *Desafios científicos e tecnológicos*

Nesta seção apresentamos uma visão geral do problema computacional da geração de descrições de objetos em língua natural e de alguns de seus desafios em aberto. Na Seção 3.1 é exemplificado o funcionamento de um algoritmo de seleção de conteúdo típico. Nas seções seguintes, são retomadas as três questões de maior interesse na presente pesquisa, a saber: o uso de descrições relacionais (Seção 3.2), a geração de descrições com diferentes graus de especificação (Seção 3.3) e a questão da variação humana na geração destas descrições (Seção 3.4). Finalmente, a Seção 3.5 apresentam um levantamento dos principais recursos linguístico-computacionais que mais se aproximam das necessidades deste projeto, bem como suas limitações.

### 3.1 A seleção de conteúdo para GLN

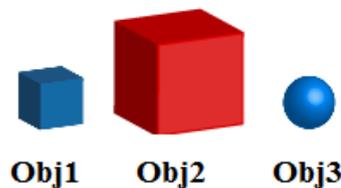
Objetos de um determinado contexto podem ser referenciados com uso de descrições definidas, indefinidas e diversas outras formas, como em ‘o prédio da esquina’ ou ‘uma rua à esquerda, depois do supermercado’. Em sistemas de geração de língua natural, denominamos *seleção de conteúdo* a tarefa computacional de decidir quais propriedades semânticas de um objeto-alvo devem ser realizadas em uma descrição deste tipo (Krahmer e Deemter, 2012).

Um dos algoritmos de seleção de conteúdo mais conhecidos na área, e que ajudou a definir o próprio problema computacional de geração de expressões de referência (GER), é o algoritmo Incremental apresentado em (Dale e Reiter, 1995). Este algoritmo recebe como entrada um contexto  $C$  formado por um grupo de objetos denominados distraidores, o objeto-alvo ou referente  $r$  que se deseja descrever, e suas propriedades semânticas na forma de pares (*atributo - valor*), como em (*cor - azul*).

A seleção de conteúdo consiste em produzir um conjunto  $L$  de propriedades de  $r$  tal que  $L$  seja capaz de distinguir  $r$  de todos os outros objetos em  $C$ . As propriedades são incluídas em  $L$  incrementalmente (de onde provem o nome do algoritmo) segundo uma ordem  $P$  predefinida, desde que contribuam para a desambiguação do referente (i.e., excluindo pelo menos um objeto do contexto  $C$ ). O algoritmo termina quando um conjunto único (i.e., livre de ambiguidade) de propriedades é produzido (caso em que  $L$  poderia ser realizada, por exemplo, como uma descrição definida), ou até que todas as propriedades possíveis em  $P$  tenham sido consideradas (caso em que  $L$  permaneceria ambígua e poderia ser realizada, por exemplo, como uma descrição indefinida).

Considere o exemplo da Figura 2 ilustrando um contexto composto por três objetos: dois cubos (um pequeno e outro grande, tendo o cubo grande a cor vermelha e o cubo pequeno a cor azul), e uma esfera de tamanho pequeno e cor azul. A semântica deste contexto poder ser representada como segue.

- Obj1:  $\{(tipo - cubo), (tamanho - pequeno), (cor - azul)\}$
- Obj2:  $\{(tipo - cubo), (tamanho - grande), (cor - vermelho)\}$
- Obj3:  $\{(tipo - esfera), (tamanho - pequeno), (cor - azul)\}$



**Figura 2** – Um contexto visual formado por dois cubos e uma esfera.

Seja o objeto-alvo  $r = Obj1$ , o contexto  $C = \{Obj2, Obj3\}$ , e a ordem preferencial de propriedades a serem consideradas  $P = \{tipo, cor, tamanho\}$ . O algoritmo inicia com uma lista  $L$  vazia e percorre  $P$  na ordem estabelecida, inserindo em  $L$  cada propriedade que exclua pelo menos um objeto em  $C^1$ . Neste exemplo, o algoritmo escolhe inicialmente a propriedade  $(tipo, cubo)$  por excluir  $Obj3$ , que é *tipo* esfera. A seguir (na ordem em  $P$ ), a propriedade  $(cor, azul)$  exclui  $Obj2$ , que é de *cor* vermelha. Como o contexto  $C$  não possui mais elementos, o algoritmo retorna a expressão  $L$  atual, que poderia ser realizada, por exemplo, como ‘o cubo azul’.

A abordagem Incremental não pode entretanto ser considerada uma solução algorítmica completa para o problema de GER. A ordem preferencial  $P$  - cuja definição é deixada em aberto em (Dale e Reiter, 1995) justamente por ser dependente de domínio - tem grande impacto sobre o tipo de expressão produzida pelo algoritmo. Por exemplo, se fosse considerada a ordem  $P = \{tipo, tamanho, cor\}$  para o domínio acima, a mesma referência a  $Obj1$  seria ‘o cubo pequeno’.

Por este motivo, o algoritmo em (Dale e Reiter, 1995) é mais adequadamente entendido como um *framework* geral de GER do que como uma solução algorítmica completa. Além disso, o algoritmo básico não trata de diversos outros critérios que passaram a ser considerados na pesquisa da área em anos recentes, como a questão da naturalidade da expressão (ou *human-likeness* em (Gatt e Belz, 2007)). Uma visão mais ampla da área e seus principais desafios é apresentada em (Krahmer e Deemter, 2012).

<sup>1</sup>Em obediência à máxima de brevidade de Grice (Grice, 1975), algoritmos deste tipo tendem a favorecer a seleção de atributos discriminatórios.

## 3.2 O uso de descrições relacionais

Em sua proposta original, o algoritmo Incremental manipula apenas propriedades atômicas do próprio objeto-alvo, como tipo, cor etc. Com certas adaptações, entretanto, pode também ser aplicado ao caso de propriedades relacionais (Paraboni et al., 2007). Esta adaptação é de especial importância, por exemplo, para o uso de *relações espaciais* como em ‘o cubo à esquerda a esfera’. Neste caso, o objeto-alvo  $r$  (o cubo vermelho) teria, além de suas propriedades atômicas (e.g., *(tipo - cubo)*, *(cor - vermelho)* etc.) uma propriedade relacional do tipo *(esquerda - o)*, no qual  $o$  é um objeto (esfera) usado como ponto de referência para descrição do objeto-alvo  $r$ .

Relações espaciais desempenham um papel central na descrição de objetos em contextos visuais, e são comuns em aplicações baseadas em mundos virtuais do tipo GIVE (Byron et al., 2007) e outros. Entretanto, o conhecimento sobre quando ou como produzir descrições deste tipo é ainda escasso. Alguns estudos como (Dale e Haddock, 1991) sugerem o uso de relações apenas como último recurso, ou seja, quando não é possível distinguir o objeto-alvo usando apenas suas propriedades atômicas. Esta observação entretanto contrasta estudos como (Viethen e Dale, 2011), que demonstram que o uso de propriedades relacionais é comum mesmo quando estas propriedades não são estritamente necessárias para desambiguação.

Alguns sistemas participantes da série de desafios *GIVE Challenge* (Byron et al., 2009; Koller et al., 2010; Striegnitz et al., 2011) implementam certos recursos de manipulação de relações deste tipo, ainda que de forma pouco documentada. Entretanto, como estes sistemas foram avaliados apenas de forma extrínseca (i.e., medindo-se o desempenho global de usuários na tarefa de navegação) não é possível distinguir o eventual impacto da seleção de conteúdo das outras funcionalidades de cada sistema, as quais incluem, por exemplo, diversas melhorias não relacionadas a esta tarefa.

De interesse mais imediato para a presente proposta, o estudo em (Paraboni e Deemter, 2013) estende o modelo de seleção de conteúdo proposto em (Paraboni et al., 2007) para domínios espaciais do tipo *GIVE* (Byron et al., 2007), resultando em um modelo de produção de relações espaciais significativamente mais informado. Além disso, em pesquisa de mestrado recentemente concluída (sob orientação do responsável por esta proposta) os resultados de estudos como (Viethen e Dale, 2011; Kelleher e Costello, 2009) foram também estendidos para este domínio. Resultados preliminares desta investigação encontram-se submetidos para publicação (Silva e Paraboni, 2013b, 2013a), mas carecem ainda de complementação e validação empírica de grande escala, baseada em exemplos de produção humana de língua.

### 3.3 A geração de descrições sub e superespecificadas

Descrições de objetos em contextos visuais podem variar drasticamente quanto ao volume de informação que representam. Em um extremo, observamos em (Teixeira et al., 2013) o caso de descrições relacionais subespecificadas, que exigem a interpretação global da expressão para que cada um de seus componentes individuais possa ser desambiguado. Por exemplo, suponha um domínio visual simples composto de um pequeno grupo de pessoas e algumas cadeiras, no qual apenas uma pessoa está sentada. Assim como em (Dale e Haddock, 1991), neste caso seria aceitável descrever está pessoa como ‘o homem sentado na cadeira’, muito embora tanto ‘homem’ como ‘cadeira’ sejam, se consideradas de forma isolada, descrições subespecificadas.

No outro extremo, observamos o caso da geração de descrições altamente superespecificadas como característica intrínseca do domínio (Viethen e Dale, 2011) ou por razões de ordem epistêmica (Paraboni e Deemter, 2013), dentre muitas outras. Em um contexto visualmente complexo (e.g., um auditório repleto de pessoas e muitas cadeiras) pode ser necessário acrescentar um certo grau de redundância lógica (i.e., informação que não seria rigorosamente necessária para desambiguação) de modo a facilitar a identificação do objeto-alvo. Por exemplo, pode ser necessário produzir uma descrição superespecificada como ‘o homem sentado na segunda cadeira da terceira fileira’. Entretanto, o acréscimo deste tipo de informação ‘redundante’ não é tratado de forma sistemática pelos algoritmos do gênero (Paraboni et al., 2007).

A questão de quando sub ou superespecificar uma descrição está no centro do problema computacional de seleção de conteúdo, e por este motivo tem sido tema de pesquisa do responsável por esta proposta (Paraboni et al., 2007; Paraboni e Deemter, 2013). O leque de novas oportunidades de investigação permanece no entanto amplo. Uma possibilidade particularmente interessante, e que permanece a ser explorada, é a possível integração destes estudos com a teoria de esforço cognitivo em (Kelleher e Costello, 2009). Esta teoria foi recentemente utilizada na já citada pesquisa em (Silva e Paraboni, 2013b, 2013a), tendo apresentado resultados iniciais promissores também submetidos para publicação (Paraboni et al., 2013; Teixeira et al., 2013).

### 3.4 A variação humana na geração de descrições

Conforme observado em (Viethen e Dale, 2011), não existe uma estratégia de produção de língua dita ‘universal’, ou nem mesmo uma estratégia única para cada falante. Pessoas diferentes descrevem um mesmo objeto de formas diferentes, e a mesma pessoa pode descrever um objeto de formas diferentes em momentos distintos (i.e., mesmo em situações idênticas).

Diversos estudos na área de psicolinguística realizaram experimentos com corpora de GER. A variação humana no corpus de diálogos Coconut (Eugenio et al., 2000) é investigada em (Gupta e Stent, 2005). O corpus TUNA (Gatt et al., 2007) figura em uma série de experimentos

derivados da série de desafios TUNA (Gatt e Belz, 2007), como em (Bohnet, 2007, 2008, 2009; Fabbrizio et al., 2008). Os corpora GRE3D e GRE3D7 são aplicados em (Viethen e Dale, 2011; Viethen et al., 2013). De modo geral, entretanto, desenvolver modelos computacionais de geração de descrições que sejam capazes de se adaptar - preferencialmente de forma automática - à estratégia de referência de um indivíduo ou grupo de indivíduos permanece ainda em aberto. Em especial, as abordagens existentes são fortemente baseadas no domínio em questão, e não há iniciativas de tratamento desta questão de forma mais generalizada.

A questão da variação humana na geração de descrições é tema de uma pesquisa de mestrado atualmente em andamento nesta instituição. Neste estudo, é apresentada uma abordagem de GER que explora a variação humana nos corpora GRE3D7 (Viethen e Dale, 2011) e Stars (Paraboni et al., 2013). Resultados preliminares - também em fase de publicação (Ferreira e Paraboni, 2013) - são promissores e, no caso específico da geração de descrições a partir do corpus GRE3D7, superiores aos apresentados em (Viethen e Dale, 2011).

### 3.5 Corpora para GLN

O desenvolvimento de modelos computacionais capazes de contemplar questões como as abordadas nas seções anteriores requer exemplos de uso prático da língua, tipicamente organizados na forma de um corpus para GLN. Diferentemente de corpora de propósito geral (e.g., do tipo normalmente utilizados em outros tipos de pesquisa em PLN), um corpus para GLN é normalmente o produto de um experimento controlado combinando fatores que motivaram a produção da língua e o resultado (i.e., o texto) obtido.

Experimentos para coleta de corpora para GLN fornecem estímulo textual ou visual aos participantes e registram suas reações, manifestas na forma oral, escrita, ou por ações de interação com o ambiente computacional. De acordo com o papel assumido pelo participante humano no experimento, podemos dividir estes estudos em experimentos orientados ao falante, experimentos orientados ao ouvinte, e em modo colaborativo. Nesta seção apresentamos um breve levantamento dos principais recursos produzidos destas três formas que se encontram publicamente disponíveis para pesquisa na área.

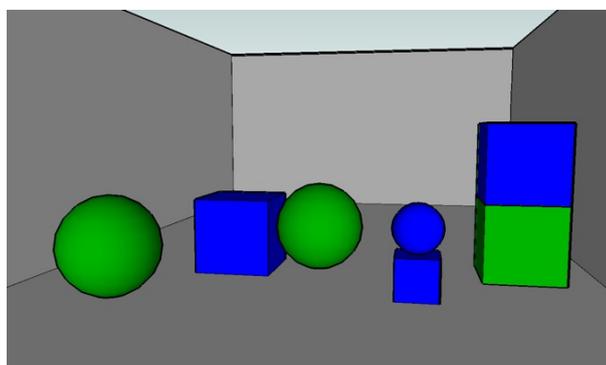
Experimentos orientados ao falante são utilizados tanto para ganhar entendimento sobre um determinado aspecto da produção de língua (e.g., coletando-se dados de treinamento) como para validação de artefatos ou sistemas previamente existentes (e.g., coletando-se dados de teste). Um dos primeiros exemplos de experimento de GLN orientado ao falante com construção de um conjunto de dados de uso público é o caso do corpus *Drawer* de expressões de referência (Viethen e Dale, 2006). Este corpus é baseado em uma única cena representando um armário com 16 gavetas de cores variadas, contendo 140 descrições geradas por 20 participantes. O objetivo do estudo foi o de examinar a questão da variação humana de estratégias de produção de língua,

a qual é de interesse para a presente proposta. O pequeno número de instâncias coletadas, entretanto, naturalmente limita o escopo de seu eventual reuso.

Um exemplo mais representativo de experimento orientado ao falante é o projeto de construção do corpus *TUNA* (Gatt et al., 2007), que contém expressões coletadas propositalmente para o estudo de fenômenos e algoritmos de geração de expressões de referência. O corpus *TUNA* descreve situações de referência em dois domínios distintos: peças de Móvel (*Furniture*), e fotos de Pessoas (*People*). O corpus *TUNA* contém 2280 expressões (780 singulares e 1500 plurais) e seus respectivos contextos. As descrições foram geradas a partir de experimentos controlados realizados com 50 participantes que usavam o Inglês como idioma nativo, ou possuíam fluência no mesmo. O propósito da descrição era apenas a identificação do objeto indicado.

O corpus *TUNA* de expressões de referência foi utilizado para treinamento e teste de algoritmos de GER em uma grande variedade de projetos, incluindo três competições da área (Gatt e Belz, 2007; Gatt et al., 2008, 2009). Este corpus foi também aproveitado em uma série de pesquisas nesta instituição (Lucena e Paraboni, 2008; Lucena et al., 2010; Pereira et al., 2012), e para participações em duas das edições do desafio *TUNA* nos anos de 2008 e 2009. Apesar da sua ampla reutilização, o domínio *TUNA* é no entanto insuficiente para os propósitos desta pesquisa. Em especial, o corpus *TUNA* carece de descrições contendo propriedades relacionais, o que na prática inviabiliza seu aproveitamento no estudo de situações de referência mais complexas e/ou domínios visuais mais realistas.

Um exemplo recente de experimento orientado ao falante é a construção dos corpora *GRE3D* e *GRE3D7* (Dale e Viethen, 2009; Viethen e Dale, 2011), que trata da questão do uso de relações espaciais em um contexto visual tridimensional simplificado. Nestes experimentos, 294 participantes foram instruídos a descrever objetos geométricos do tipo esfera e cubo. O procedimento resultou no corpus *GRE3D* (Dale e Viethen, 2009), posteriormente ampliado para o corpus *GRE3D7* (Viethen e Dale, 2011), contendo 4480 descrições destes tipos de objetos. Um exemplo deste tipo de domínio é ilustrado na figura 3.



**Figura 3** – Exemplo de estímulo visual utilizado na construção do corpus *GRE3D*, adaptado de (Dale e Viethen, 2009).

O corpus *GRE3D7* reúne diversas características desejáveis para o presente estudo, e foi inclusive utilizado em uma iniciativa de pesquisa nesta instituição (Silva e Paraboni, 2013b, 2013a). Em especial, o corpus *GRE3D7* constitui uma coleção de proporções significativa - é possivelmente o maior corpus do gênero disponível para fins de pesquisa - e contendo relações espaciais de interesse para descrição de objetos em contextos visuais. Apesar destas vantagens, entretanto, o domínio *GRE3D7* é ainda excessivamente simplificado para aproveitamento no presente estudo. O número reduzido de atributos atômicos e relacionais possíveis representa uma limitação considerável para estudos de maior cobertura além do já apresentado em (Viethen e Dale, 2011). Em especial, não é possível investigar a questão do grau de especificação destas expressões com base apenas nestes dados.

Experimentos orientados ao falante são também úteis para a construção de corpora de diálogos, como no caso do corpus *iMap* (Guhe, 2009) de instruções de rota em mapas bidimensionais. O corpus *iMap* consiste de 256 diálogos construídos com base em mapas. Em cada diálogo, uma dupla de participantes revezava-se nos papéis de instrutor e receptor das instruções. Ambos participantes tinham acesso ao mapa, mas apenas o mapa do instrutor exibia o traçado do caminho a ser seguido. A tarefa do instrutor consistia assim em descrever este caminho de modo que o receptor pudesse desenhá-lo em seu próprio mapa.

O conjunto de dados formado pelo corpus *iMap* (i.e., diálogos, imagens etc.) constitui um recurso potencialmente valioso para estudos em GLN. Entretanto, como estes dados não estão disponíveis publicamente por razões de confidencialidade (os experimentos incluíam pesquisas na área de psicologia), não há uso prático além do estudo apresentado em (Guhe, 2009).

Finalmente, o corpus *GIVE-2* (Gargett et al., 2010) de instruções em mundos virtuais foi construído por meio de experimentos envolvendo 36 pares de participantes de língua inglesa e alemã alternando-se nas tarefas de instrutor e jogador. Este corpus foi produzido em condições semelhantes às do modelo instrutor-receptor empregado no corpus *iMap*, e contém todas as instruções fornecidas pelo instrutor, e as respectivas decisões tomadas pelo jogador (e.g., movimentos, ações de pressionar botões etc.). O conjunto de dados multimodal resultante pode ser visualizado na forma de animação com uso da ferramenta *Replay* em (Gargett et al., 2010).

Mesmo constituindo um domínio mais próximo de uma aplicação computacional real, entretanto, o corpus *GIVE-2* também apresenta algumas dificuldades de reuso na presente proposta. A representação em formato de animação acompanhada de diálogos torna as tarefas de anotação e processamento significativamente complexas, e provavelmente não justificáveis face ao número reduzido de descrições de interesse (cerca de 992 descrições de objetos do tipo botão, que é o único elemento manipulável em mundos *GIVE*).

## 4 *Cronograma de atividades*

O projeto proposto é dividido em duas grandes etapas, correspondendo cada uma a aproximadamente um ano de duração: uma etapa de construção e preparação do corpus propriamente dito, e uma etapa investigativa para tratamento das questões levantadas na seção anterior, incluindo a proposta e validação de um modelo computacional de GER baseado no corpus produzido. As principais atividades contempladas são relacionadas a seguir.

Na primeira etapa - construção do corpus - são contempladas três atividades principais: (1) a elaboração de um experimentos online em modo colaborativo para coleta de descrições de objetos sob condições controladas, (2) a execução do experimento propriamente dito e (3) a anotação do corpus por um grupo de avaliadores para posterior disponibilização à comunidade científica. Para fins de investigação das questões de interesse para a presente pesquisa, o corpus coletado deverá privilegiar a produção de descrições relacionais em situações de referência que exijam graus variados de especificação, e devem ser produzido por um número considerável de participantes em modo colaborativo. Aos moldes do corpus GIVE-2 (Gargett et al., 2010), espera-se contar com pelo menos 30 duplas de participantes revezando-se no papel de falante e ouvinte.

No momento da elaboração da presente proposta, dois protótipos de experimento deste tipo encontram-se em desenvolvimento: o primeiro, realizado em colaboração com a Universidade de Córdoba (Argentina), objetiva a construção de um corpus bilíngue (Português-Espanhol, com possibilidade de futura expansão para o Inglês) de descrições de objetos em mapas geográficos; o segundo, mais voltado para questões de sub e superespecificação de expressões de referência discutidas na seção anterior, é atualmente tema de um grupo de pesquisa nesta instituição, e inclui um auxílio FAPESP na modalidade IC<sup>1</sup>.

O produto final desta etapa é o corpus produzido a partir do experimento online colaborativo, anotado com a representação semântica das descrições produzidas em grau satisfatório de concordância entre avaliadores. O corpus deve ser representado em formato XML segundo práticas consagradas da área (Gatt et al., 2007) e disponibilizado para reuso.

Uma vez construído e anotado, o corpus será então empregado na segunda etapa do projeto, que trata do estudo e desenvolvimento de um modelo computacional contemplando as questões enunciadas na seção anterior, e que são uma extensão de temas de pesquisa já em andamento nesta instituição.

---

<sup>1</sup>Processo FAPESP nro. 2013/10641-3.

A primeira destas questões - o uso de descrições relacionais - deverá beneficiar-se diretamente das descrições relacionais disponibilizadas pelo corpus. Mais especificamente, espera-se obter evidência de quando e como estas descrições devam ser geradas, e de como estas informações podem ser incorporadas a um algoritmo de GER de uso prático. Estudos preliminares nesta área têm sido conduzidos nesta instituição como em (Paraboni et al., 2007; Paraboni e Deemter, 2013; Silva e Paraboni, 2013b), e incluem submissões recentes como (Silva e Paraboni, 2013a). Uma possibilidade a ser investigada neste caso é o uso da hierarquia de esforço cognitivo proposta em (Kelleher e Costello, 2009) para guiar as decisões de um algoritmo de GER sobre qual propriedade relacional utilizar.

A segunda questão de pesquisa - o estudo da geração de expressões sub e superespecificadas - deverá explorar a variação no grau de especificação das descrições do corpus. Dentre outras possibilidades, consideramos por exemplo a aprendizagem de um modelo de subespecificação que poderia ser incorporado ao proposto em (Paraboni e Deemter, 2013), e que já trata do fenômeno simétrico a este, ou seja, a superespecificação. Juntos, espera-se que estes dois estudos constituam um primeiro passo em direção a uma teoria de ampla cobertura do fenômeno de referência em contextos visuais sob condições realistas, possivelmente também atrelada à hierarquia de esforço cognitivo proposta em (Kelleher e Costello, 2009). Um primeiro passo nesta direção é detalhado nas recentes submissões em (Paraboni et al., 2013; Teixeira et al., 2013).

Finalmente, a terceira questão - o estudo da variação humana de estratégias de referência - deverá explorar as diferenças entre descrições produzidas por participantes distintos em uma mesma situação de referência. Em especial, contemplamos a possibilidade de modelar diversos aspectos da estratégia humana de referência - incluídas aqui as duas questões anteriores, mas não limitado a estas - com base nos exemplos do corpus. Uma recente submissão em (Ferreira e Paraboni, 2013) apresenta um estudo mais detalhado sobre este assunto, e explora a variação humana em combinação com o grau de especificação de cada descrição (cf. questão anterior).

Estratégias individuais podem também ser agrupadas com uso de técnicas de aprendizagem supervisionadas (Viethen e Dale, 2011) ou não-supervisionadas (Viethen et al., 2013), o que levanta uma série de questões de pesquisa relacionadas à modelagem de comportamento de grupos versus indivíduos. De forma incipiente, um estudo deste tipo foi apresentado em (Pereira et al., 2012), no qual foi utilizada a técnica de indução de árvores de decisão para aprendizagem de estratégias de referência mais frequentes no corpus *TUNA* (Gatt et al., 2007).

Com base nos resultados do estudo destas três questões, será proposto um modelo computacional de geração de descrições de objetos em contextos visuais (5) a ser implementado na forma de um algoritmo prático de GER (6). A solução proposta será avaliada (7) de forma intrínseca com base nos dados do corpus produzido utilizando-se métricas consagradas da área como Dice (Dice, 1945), MASI (Passonneau, 2006) e outras.



## ***5 Avaliação e disseminação***

A anotação do corpus - embora relativamente simples no caso de domínios fechados como utilizados em experimentos deste tipo - será realizada por um grupo de juízes, e avaliado com uso de medidas de concordância tais como Kappa (Landis e Koch, 1977).

O modelo computacional desenvolvido será avaliado de forma intrínseca com base em métricas consagradas da área, tais como Dice (Dice, 1945) e MASI (Passonneau, 2006). Métricas deste tipo têm sido amplamente utilizadas na avaliação da tarefa computacional de seleção de conteúdo e algoritmos de GER como em (Gatt e Belz, 2007; Gatt et al., 2008, 2009; Lucena et al., 2010; Paraboni et al., 2007; Paraboni e Deemter, 2013).

Todos os resultados da pesquisa e os produtos desenvolvidos ao longo do projeto serão divulgados em eventos e publicações científicas da área de PLN/GLN.

## 6 *Recursos*

A instituição de destino (USP/EACH) oferecerá o espaço físico e recursos computacionais básicos (inclusive provenientes de apoio anterior desta mesma agência de fomento) que se façam necessários para o início deste estudo. Os itens solicitados são assim destinados à modernização dos recursos computacionais existentes e participação em reuniões e eventos científicos de divulgação da pesquisa realizada. Estes itens são relacionados na documentação anexa à proposta no sistema SAGE.

Além da já citada colaboração com a Universidade de Córdoba, o grupo local conta com apoio de estudantes de graduação e pós-graduação (incluindo duas pesquisas em nível de mestrado diretamente ligadas ao tema proposto) para apoio às tarefas de coleta e anotação de dados, validação de hipóteses e outros.

A pesquisa proposta conta também com uma considerável infraestrutura de software de GLN desenvolvida ao longo dos anos nesta instituição. Estes recursos incluem ambientes de desenvolvimento e teste de modelos do tipo proposto, ferramentas de avaliação, algoritmos de GER já implementados (que podem ser estendidos e/ou reutilizados como sistemas de *baseline*), e protótipos integrados aos ambientes interativos 3D do tipo GIVE (Byron et al., 2007) e GRUVE (Janarthanam et al., 2012).

## Referências

- ARAUJO, R. P. A. de; OLIVEIRA, R. L. de; AIS, E. M. de N.; TADEU, T. D.; PEREIRA, D. B.; PARABONI, I. SINotas: the evaluation of a NLG application. In: *7th International Conference on Language Resources and Evaluation (LREC-2010)*. 2010. p. 2388–2391.
- AZIZ, W. F.; PARDO, T. A. S.; PARABONI, I. An experiment in spanish-portuguese statistical machine translation. *LNAI*, Springer-Verlag, v. 5249, p. 248–257, 2008.
- BOHNET, B. IS-FBN, IS-FBS, IS-IAC: The adaptation of two classic algorithms for the generation of referring expressions in order to produce expressions like humans do. In: *MT Summit XI, UCNLG+ MT*. 2007. p. 84–86.
- BOHNET, B. The fingerprint of human referring expressions and their surface realization with graph transducers. In: *5th International Natural Language Generation Conference*. 2008. p. 207–210.
- BOHNET, B. Generation of referring expression with an individual imprint. In: *12th European Workshop on Natural Language Generation*. 2009. p. 185–186.
- BYRON, D.; KOLLER, A.; OBERLANDER, J.; STOIA, L.; STRIEGNITZ, K. Generating instructions in virtual environments (GIVE): A challenge and evaluation testbed for NLG. In: *Workshop on Shared Tasks and Comparative Evaluation in Natural Language Generation*. 2007.
- BYRON, D.; KOLLER, A.; STRIEGNITZ, K.; CASSELL, J.; DALE, R.; MOORE, J.; OBERLANDER, J. Report on the first NLG challenge on generating instructions in virtual environments (GIVE). In: *12th European Workshop on Natural Language Generation (ENLG)*. Athens: , 2009.
- CUEVAS, R. R. M.; HONDA, W. Y.; LUCENA, D. J. de; PARABONI, I.; OLIVEIRA, P. R. Portuguese pronoun resolution: resources and evaluation. *LNCS*, Springer-Verlag, v. 4919, p. 344–350, 2008.
- CUEVAS, R. R. M.; PARABONI, I. A machine learning approach to portuguese pronoun resolution. *LNAI*, Springer-Verlag, v. 5290, p. 262–271, 2008.
- DALE, R.; HADDOCK, N. J. Content determination in the generation of referring expressions. *Computational Intelligence*, v. 7, p. 252–265, 1991.
- DALE, R.; REITER, E. Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, v. 19, 1995.
- DALE, R.; VIETHEN, J. Referring expression generation through attribute-based heuristics. In: *12th European Workshop on Natural Language Generation*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2009. (ENLG '09), p. 58–65.
- DICE, L. R. Measures of the amount of ecologic association between species. *Ecology*, v. 26, n. 3, p. 297–302, 1945.
- EUGENIO, B. D.; JORDAN, P. W.; THOMASON, R. H.; MOORE, J. D. The agreement process: An empirical investigation of human-human computer-mediated collaborative dialogues. *International Journal of Human-Computer Studies*, v. 53, n. 6, p. 1017–1076, 2000.
- FABBRIZIO, G. D.; STENT, A.; BANGALORE, S. Trainable speaker-based referring expression generation. In: *12th Conference on Computational Natural Language Learning*. 2008. p. 151–158.
- FERREIRA, T. C.; PARABONI, I. Speaker variation in referring expression generation (submetido). In: . 2013.
- GARGETT, A.; GAROUFI, K.; KOLLER, A.; STRIEGNITZ, K. The GIVE-2 corpus of giving instructions in virtual environments. In: *Proceedings of LREC-2010*. 2010.
- GATT, A.; BELZ, A. The attribute selection for GRE challenge: Overview and evaluation results. In: *UCNLG+MT: Language Generation and Machine Translation*. 2007.
- GATT, A.; BELZ, A.; KOW, E. The TUNA challenge 2008: Overview and evaluation results. In: *Fifth International Natural Language Generation Conference (INLG-2008)*. 2008. p. 198–206.
- GATT, A.; BELZ, A.; KOW, E. The TUNA challenge 2009: Overview and evaluation results. In: *12nd European Workshop on Natural Language Generation*. 2009. p. 174–182.
- GATT, A.; SLUIS, I. van der; DEEMTER, K. van. Evaluating algorithms for the generation of referring expressions using a balanced corpus. In: *Proceedings of ENLG-07*. 2007.
- GRICE, H. P. Logic and conversation. In: COLE, P.; MORGAN, J. L. (Ed.). *Syntax and semantics*. : New York: Academic Press, 1975. v. 3.
- GUHE, M. Generating referring expressions with a cognitive model. In: *Production of Referring Expressions: Bridging the gap between computational and empirical approaches to reference*. 2009.

- GUPTA, S.; STENT, A. Automatic evaluation of referring expression generation using corpora. In: *1ST WORKSHOP ON USING CORPORA IN NLG*. 2005.
- JANARTHANAM, S.; LEMON, O.; LIU, X. A web-based evaluation framework for spatial instruction-giving systems. In: *Proceedings of ACL-2012*. 2012. p. 49–54.
- KELLEHER, J. D.; COSTELLO, F. J. Applying computational models of spatial prepositions to visually situated dialog. *Computational Linguistics*, v. 35, n. 2, p. 271–306, 2009.
- KOLLER, A.; STRIEGNITZ, K.; GARGETT, A.; BYRON, D.; CASSELL, J.; DALE, R.; MOORE, J.; OBERLANDER, J. Report on the second NLG challenge on generating instructions in virtual environments (GIVE-2). In: *6th International Natural Language Generation Conference (INLG)*. Dublin: , 2010.
- KRAHMER, E.; DEEMTER, K. van. Computational generation of referring expressions: A survey. *Computational Linguistics*, v. 38, n. 1, p. 173–218, 2012.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, p. 159–174, 1977.
- LUCENA, D. J. de; PARABONI, I. Combining frequent and discriminating attributes in the generation of definite descriptions. *LNAI*, Springer-Verlag, v. 5290, p. 252–261, 2008.
- LUCENA, D. J. de; PARABONI, I. The design of an experiment in anaphora resolution for referring expressions generation. In: *Recent Advances in Natural Language Processing (RANLP-2009)*. 2009. p. 14–16.
- LUCENA, D. J. de; PEREIRA, D. B.; PARABONI, I. From semantic properties to surface text: The generation of domain object descriptions. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, Asociacion Espanhola para la Inteligencia Artificial, v. 14, n. 45, p. 48–58, 2010.
- NOVAIS, E. M. de; PARABONI, I. Portuguese text generation using factored language models. *Journal of the Brazilian Computer Society*, Springer-Verlag, p. 1–12, 2012.
- NOVAIS, E. M. de; PARABONI, I.; FERREIRA, D. T. Highly-inflected language generation using factored language models. *Computational Linguistics and Intelligent Text Processing*, Springer-Verlag, LNCS 6608, p. 429–438, 2011.
- NUNES, M. das G. V.; VIEIRA, F. M. C.; ZAVAGLIA, C.; SOSSOLOTE, C. R. C.; HERNANDEZ, J. A construção de um léxico para o português do Brasil: lições aprendidas e perspectivas. In: *II PROPOR*. 1996. p. 61–70.
- PARABONI, I.; DEEMTER, K. van. Referring via document parts. *Computational Linguistics and Intelligent Text Processing*, Springer, LNCS 3878, p. 299–310, 2006.
- PARABONI, I.; DEEMTER, K. van. Reference and the facilitation of search in spatial domains. *Language and Cognitive Processes*, DOI 10.1080/01690965.2013.805796, 2013.
- PARABONI, I.; DEEMTER, K. van; MASTHOFF, J. Generating referring expressions: Making referents easy to identify. *Computational Linguistics*, MIT Press, v. 33, n. 2, p. 229–254, 2007.
- PARABONI, I.; MASTHOFF, J.; DEEMTER, K. van. Overspecified reference in hierarchical domains: measuring the benefits for readers. In: *Proceedings of INLG-2006*. 2006. p. 55–62.
- PARABONI, I.; YAMASAKI, A. K.; SILVA, A. S. R. da; TEIXEIRA, C. V. M. Underspecified relational descriptions: a corpus study (submetido). In: . 2013.
- PASSONNEAU, R. Measuring agreement on set-valued items (MASI) for semantic and pragmatic annotation. In: *International Conference on Language Resources and Evaluation (LREC)*. 2006.
- PEREIRA, H. V. L.; NOVAIS, E. M. de; MARIOTTI, A. C.; PARABONI, I. Corpus-based referring expressions generation. In: *8th International Conference on Language Resources and Evaluation (LREC-2012)*. 2012. p. 4004–4009.
- SILVA, D. dos S.; PARABONI, I. Generating spatial references in interactive 3D worlds (submetido). In: . 2013.
- SILVA, D. dos S.; PARABONI, I. Geração de expressões de referência usando relações espaciais. In: *Proceedings of STIL-2013*. 2013. p. 88–97.
- STRIEGNITZ, K.; DENIS, A.; GARGETT, A.; GAROUFI, K.; KOLLER, A.; THEUNE, M. Report on the second second challenge on generating instructions in virtual environments (GIVE-2.5). In: *Generation Challenges Session at the 13th European Workshop on Natural Language Generation (ENLG)*. 2011. p. 270–279.
- TEIXEIRA, C. V. M.; PARABONI, I.; SILVA, A. S. R. da; YAMASAKI, A. K. Generating relational descriptions involving mutual disambiguation (submetido). In: . 2013.
- VIETHEN, J.; DALE, R. Algorithms for generating referring expressions: Do they do what people do? In: *Proceedings of INLG-2006*. 2006. p. 63–70.
- VIETHEN, J.; DALE, R. GRE3D7: A corpus of distinguishing descriptions for objects in visual scenes. In: *UCNLG+Eval: Language Generation and Evaluation Workshop*. 2011. p. 12–22.
- VIETHEN, J.; MITCHELL, M.; KRAHMER, E. Graphs and spatial relations in the generation of referring expressions. In: *14th European Workshop on Natural Language Generation*. : Association for Computational Linguistics, 2013. p. 72–81.