

Geração de Expressões de Referência em Ambientes Virtuais Interativos*

Referring Expression Generation in Interactive Virtual Worlds

Diego dos Santos Silva

School of Arts, Sciences and Humanities (EACH)
University of São Paulo (USP)
diego.silva@usp.br

Ivandré Paraboni

School of Arts, Sciences and Humanities (EACH)
University of São Paulo (USP)
ivandre@usp.br

Resumo

Sistemas de geração de instruções em mundos virtuais interativos 3D possuem uma ampla gama de aplicações em áreas como educação, desenvolvimento de jogos etc. Neste artigo discutimos o problema computacional da geração de expressões de referência em ambientes deste tipo, enfocando a questão do uso de relações espaciais para descrever objetos do domínio.

Palavras chave

Geração de Língua Natural, Expressões de Referência

Abstract

Instruction-giving systems for virtual interactive 3D worlds have a wide range of applications in education, games and others. This paper discusses the computational task of referring expression generation for systems of this kind, focusing on the use of spatial relations to describe domain objects.

Keywords

Natural Language Generation, Referring Expressions

1 Introdução

A geração de expressões de referência (GER) é um dos componentes fundamentais de aplicações de geração de língua natural (GLN) a partir de dados de entrada não linguísticos. Algoritmos de GER tratam da tarefa de seleção do conteúdo semântico a ser realizado, por exemplo, na forma de descrições definidas¹ como a seguir:

a *O velho*

b *O homem de óculos, à esquerda*

*Este trabalho conta com apoio FAPESP.

¹Em contraste à questão da *interpretação* (e.g., anafórica) de expressões existentes, (Paraboni, 1997; Cuevas e Paraboni, 2008).

c *O segundo homem, de casaco preto e ao lado do rapaz que está fumando*

A escolha de um determinado conjunto de propriedades semânticas para compor uma expressão de referência como nos exemplos acima traz uma série de consequências para o leitor (ou ouvinte), tanto no que diz respeito à sua interpretação linguística como à sua resolução (aqui entendida como sendo a identificação do objeto-alvo da referência). Por exemplo, uma descrição excessivamente breve como (a) pode, em determinado contexto (e.g., uma multidão) dificultar a identificação do referente. Por outro lado, uma descrição muito extensa como (c) pode apresentar maior dificuldade de interpretação.

De forma mais ampla, pode-se dizer que a tarefa computacional de GER consiste em produzir descrições que sejam psicologicamente plausíveis, ou seja, o mais próximo possível das descrições que seriam produzidos por sujeitos humanos em condições semelhantes. Isso inclui, por exemplo, a necessidade de evitar a produção de descrições ambíguas, contendo falsas implicações lógicas, excessivamente breves ou extensas, ou que façam uso de propriedades incomuns para aquele tipo de contexto, dentre vários outros objetivos geralmente conflitantes.

GER é uma ativa linha de pesquisa em GLN, tendo sido inclusive objeto de uma série de competições (ou *shared tasks*) recentes (Belz e Gatt, 2007; Gatt, Belz e Kow, 2008; Gatt, Belz e Kow, 2009). Abordagens existentes, entretanto, tendem a considerar principalmente domínios simplificados e/ou bidimensionais. O problema de referência em domínios físicos mais realistas (e.g., com grande complexidade estrutural, tridimensionalidade etc.) permanece pouco explorado na pesquisa da área, possivelmente em virtude da própria dificuldade em criar bons modelos computacionais deste tipo.

Mais recentemente, entretanto, este cenário começou a mudar com iniciativas como o projeto GIVE (*Generating Instructions in Virtual Environments*) (Koller et al., 2009). GIVE é uma plataforma para desenvolvimento e teste de sistemas de GLN em mundos virtuais interativos, na qual o sistema encarrega-se de todo gerenciamento do ambiente gráfico e de interatividade, permitindo ao desenvolvedor concentrar-se apenas na tarefa de GLN e produzir rapidamente uma aplicação de teste. A Figura 1 ilustra este sistema.

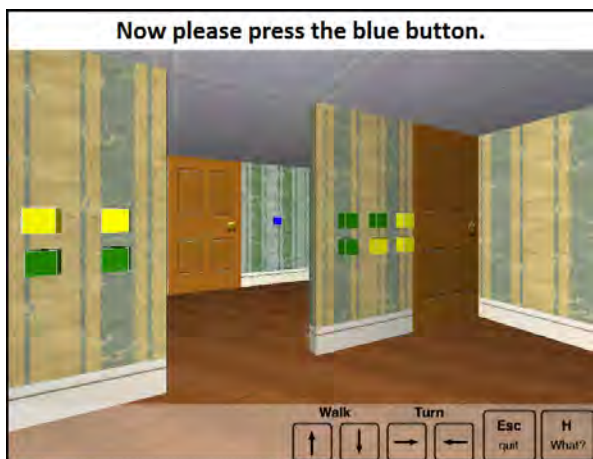


Figura 1: Exemplo de contexto GIVE.

O ambiente GIVE é composto de objetos manipuláveis do tipo botão, usados para abrir e fechar portas e outras funções de interação com o mundo virtual, e objetos maiores como peças de mobiliário e afins. Estes objetos podem, a critério do desenvolvedor de GLN, ser empregados como ponto de referência nas instruções fornecidas ao usuário, como em ‘Aperte o botão ao lado da porta’. O uso de pontos de referência não é, entretanto, um recurso nativo do sistema, e sua implementação requer a extração e manipulação de relações espaciais (e.g., ‘ao lado de’) por parte do algoritmo de GER.

Este artigo discute o desenvolvimento de um algoritmo de GER que faz uso de relações espaciais em ambientes do tipo GIVE. Assim como em (Tenbrink, 2005; Baltaretu, Krahmer e Maes, 2013), assume-se que o propósito único da informação espacial seja a desambiguação do referente para fins de identificação, e não a localização de um objeto previamente identificado (Barclay e Galton, 2008). Será tratada especificamente a questão da seleção de conteúdo semântico, deixando de lado o problema da realização textual destas expressões (Pereira e Paraboni, 2007; Pereira e Paraboni, 2008; de Novais e Paraboni, 2012).

2 Trabalhos Relacionados

2.1 O problema computacional de GER

Um dos algoritmos mais conhecidos na área, e que ajudou a definir o próprio problema computacional de 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, o objeto-alvo ou referente r que se deseja descrever, e suas propriedades semânticas na forma de pares (*atributo-valor*), como em (*cor-azul*).

O objetivo do algoritmo é 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) seguindo 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 obtido (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 a seguir, ilustrando um contexto composto por três objetos: dois cachorros (um preto e um branco, sendo o preto de tamanho pequeno e o branco de tamanho grande), e um gato preto e pequeno.

- Obj1:
(*tipo-cachorro*), (*cor-preto*), (*tamanho-pequeno*)
- Obj2:
(*tipo-cachorro*), (*cor-branco*), (*tamanho-grande*)
- Obj3:
(*tipo-gato*), (*cor-preto*), (*tamanho-pequeno*)

Seja:

- o objeto-alvo $r = Obj1$;
- o contexto $C = \{Obj2, Obj3\}$;

e a ordem preferencial

$$P = \langle \textit{tipo}, \textit{cor}, \textit{tamanho} \rangle$$

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 . Neste exemplo, o algoritmo escolhe inicialmente a propriedade (*tipo-cachorro*) por excluir *Obj3*, que é *tipo* gato. A seguir (na ordem em P), a propriedade (*cor-preto*) exclui

Obj2, que é de *cor* branca. Como o contexto *C* não possui mais elementos, o algoritmo retorna a expressão *L*, que poderia ser realizada, por exemplo, como ‘o cachorro preto’.

A ordem preferencial de seleção de atributos da lista *P* tem grande impacto sobre o tipo de expressão produzida pelo algoritmo. Por exemplo, se fosse considerada a ordem

$$P = \langle \text{tipo}, \text{tamanho}, \text{cor} \rangle$$

para o domínio acima, a mesma referência a *Obj1* seria ‘o cachorro pequeno’. Em obediência à máxima de brevidade de Grice (Grice, 1975), algoritmos de GER tendem a favorecer a seleção de atributos discriminatórios. Em anos recentes, no entanto, passaram a ser consideradas também questões como a naturalidade da expressão (ou *humanlikeness* em (Belz e Gatt, 2007)), dentre muitos outros objetivos desejáveis. Uma visão geral da área de GER e seus principais desafios é apresentada em (Krahmer e van Deemter, 2012).

2.2 Geração de descrições relacionais

Exemplos de algoritmos de GER que produzem descrições relacionais incluem (Dale e Haddock, 1991; Paraboni e van Deemter, 1999; Krahmer e Theune, 2002). De modo geral, entretanto, estes estudos não levam em conta as peculiaridades do fenômeno de referência em mundos virtuais.

Em sua proposta original, o algoritmo Incremental manipula apenas propriedades atômicas, mas com adaptações a serem discutidas na seção 4.2 pode também ser aplicado ao caso de propriedades relacionais. Esta modificação é de especial importância para o uso de relações espaciais de que trata este trabalho, como em (*acima,o*), no qual *o* é um objeto usado como ponto de referência para a descrição do objeto-alvo.

Alguns sistemas participantes da série de competições *GIVE Challenge* (Byron et al., 2009; Koller et al., 2010; Striegnitz et al., 2011) implementam certos recursos de manipulação de relações espaciais, ainda que de forma pouco documentada (Braunias et al., 2010; Schutte e Dethlefs, 2010; Garoufi e Koller, 2011; Akkersdijk et al., 2011). Entretanto, como estes sistemas foram avaliados apenas de forma extrínseca (i.e., medindo-se o desempenho global de usuários GIVE na tarefa de navegação) não é possível distinguir o eventual impacto do uso de relações espaciais das outras funcionalidades de cada sistema, as quais incluem, por exemplo, um grande número de melhorias não relacionadas à tarefa de GER.

2.3 Corpora de GER

Uma questão recorrente na pesquisa em GER é como obter evidência empírica sobre o fenômeno de referência. Uma opção natural, e amplamente empregada na grande área de Processamento de Língua Natural (PLN), é a avaliação baseada em corpus. Por exemplo, em aplicações de tradução automática (Aziz, Pardo e Paraboni, 2008), grandes coleções de textos paralelos (i.e., na língua-alvo e na língua-destino) são empregadas para treinamento de modelos estatísticos deste tipo.

O uso de corpora textuais para pesquisa em GER é entretanto limitado pelo fato de que o conhecimento de entrada que produziu o texto presente no corpus normalmente não está disponível. Ou seja, o conteúdo de um corpus textual típico é meramente o produto final de um processo de geração de língua natural realizado por humanos, e que não contém informações explícitas sobre o processo em si. Para diversos tipos de pesquisa em GER e áreas afins, faz-se necessário assim examinar não apenas o texto resultante do processo, mas também modelar as condições contextuais nas quais o texto foi produzida.

A forma usual de reproduzir estas condições na pesquisa em GER é a realização de experimentos controlados com uso de participantes humanos para coleta de corpus de GER, aqui entendido como sendo um conjunto de expressões de referência e os respectivos contextos (e.g., cenas) nos quais cada expressão foi produzida. Alguns exemplos de recursos deste tipo incluem TUNA (Gatt, van der Sluis e van Deemter, 2007), GRE3D3 (Dale e Viethen, 2009), GRE3D7 (Viethen e Dale, 2011) e Stars (Teixeira et al., 2014). De modo geral, entretanto, estes corpora tendem a representar apenas situações estáticas de referência (i.e., sem interatividade) e/ou envolvendo contextos visuais bidimensionais.

Uma exceção de especial interesse para esta pesquisa é o corpus GIVE-2 (Gargett et al., 2010) de instruções em mundos virtuais como ‘dobrar à esquerda’, ‘pressionar o segundo botão, ao lado da porta’ etc. GIVE-2 foi construído por meio de experimentos envolvendo 36 pares de participantes de língua inglesa e alemã alternando-se nas tarefas de instrutor e jogador. O corpus contém todas as instruções fornecidas pelo instrutor e as respectivas decisões tomadas pelo jogador (e.g., movimentos, ações de pressionar botões etc.) em três mundos de exemplo. Este conjunto de dados multimodal pode ser visualizado na forma de animação com uso da ferramenta *Replay* apresentada em (Gargett et al., 2010).

3 Extração e Preparação de Dados

O presente trabalho faz uso de informações contextuais extraídas dos três mundos virtuais que compõem o corpus GIVE-2 (Gargett et al., 2010), e das expressões de referência produzidas nestes contextos. A extração e preparação destes dados é discutida individualmente nas seções a seguir.

3.1 Extração de relações espaciais

Objetos em um ambiente GIVE (Koller et al., 2009) possuem apenas uma propriedade atômica básica representando seu *tipo* (botões, portas, cadeiras etc.) e, no caso dos botões, uma propriedade *cor*. O primeiro passo deste trabalho foi assim a implementação de um conjunto de métodos básicos para computar relações espaciais de diversos tipos a partir de um mundo virtual.

As relações espaciais computadas para um dado objeto-alvo r e ponto de referência o são: *acima*, *abaixo*, *esquerda*, *direita*, *frente* e *atrás*. Para extração destas relações, foram utilizadas as funções propostas em (Kelleher e Costello, 2009), baseadas na posição angular de um objeto em relação ao outro no plano cartesiano.

O algoritmo de extração de relações espaciais utiliza uma constante de distância máxima k única para cada tipo de relação. A posição física das entidades no ambiente GIVE é definida pela coordenada de seu ponto central, e assim entidades maiores como sofás, portas etc. possuem um ponto central mais distante das bordas. Uma entidade está próxima de outra entidade se a diferença entre os valores para os eixos x , y e z é no máximo k . Se esta condição for verdadeira, considera-se que há uma relação espacial válida do ponto de vista semântico.

Tendo em vista o propósito de gerar expressões de referência livres de ambiguidade, a questão da transitividade destas relações foi aqui desconsiderada. Assim, relações como *esquerda-o* devem ser entendidas como ‘imediatamente à esquerda’ do objeto o , e não contemplando objetos mais distantes que também possam estar à esquerda de o .

3.2 Preparação dos dados do corpus

As expressões de referência aqui consideradas são as que descrevem objetos do tipo botão no mundo GIVE. Botões são frequentemente referenciados nas instruções de navegação por serem os únicos elementos manipuláveis neste ambiente.

Foram extraídas do corpus todas instruções contendo a palavra ‘button’ e formas equivalen-

Relação	Ocorrências	%
próximo	217	21,87%
esquerda	121	12,19%
direita	92	9,27%
acima	17	1,71%
canto	14	1,41%
frente	6	0,60%

Tabela 1: Relações espaciais envolvendo objetos do *tipo* ‘botão’ em GIVE-2 (Gargett et al., 2010).

tes como ‘this’ e ‘box’ inferidas pelo uso de verbos como ‘press’ ou ‘click’. Como no entanto este procedimento não foi exaustivo, é possível que uma pequena parcela de descrições menos comuns tenha sido excluída da presente análise.

No total, foram identificadas 992 descrições de interesse. No caso da porção em alemão do corpus, as instruções foram previamente traduzidas para o inglês com uso da ferramenta *Google Translate*² de modo a facilitar sua interpretação.

Apenas dois tipos de propriedades atômicas foram observadas nas expressões de referência coletadas: *tipo* e *cor*. Das expressões coletadas, 467 (47,07%) utilizam algum tipo de relação espacial, sendo 248 (25%) do tipo topológica (e.g., ‘O botão *perto* da planta’), e 219 (22,07%) do tipo projetiva (e.g., ‘O botão *à esquerda* da planta’) (Kelleher e Costello, 2009). A Tabela 1 sumariza os tipos de propriedades relacionais identificados.

No corpus GIVE-2 há três tipos de expressões que não foram consideradas neste trabalho: as que incluem relações com o jogador (e.g., ‘o botão *à sua frente*’), propriedades comparativas (e.g., ‘o botão *mais distante* da lâmpada’) e envolvendo negações (e.g., ‘o botão *que não está perto* da lâmpada’). Foram identificadas 68 descrições destes tipos, correspondendo a (6,8%) das expressões consideradas. Nestes casos é assumido um ônus para a solução proposta, que nem sempre será capaz de produzir descrições idênticas às observadas no corpus.

Um breve exame das expressões coletadas é suficiente para constatar que, em uma mesma situação de referência, pessoas diferentes podem usar ou não uma relação espacial, variação esta que pode ser tomada por um indicador da complexidade da tarefa de geração considerada. A Tabela 2 ilustra esta variação, representada pelo número de casos em que uma relação espacial foi ou não utilizada para cada referência nos três mundos (ou domínios) do corpus GIVE-2.

Pelos dados da Tabela 2 é possível observar,

²<http://translate.google.com.br/>

mundo 1			mundo 2			mundo 3		
r	sim	não	r	sim	não	r	sim	não
b19	5	32	b18	32	6	b1	29	4
b18	26	0	b11	1	6	b12	1	38
b3	2	36	b12	26	10	b5	74	2
b20	30	0	b2	0	47	b4	39	2
b11	5	30	b5	19	17	b10	32	0
b5	0	72	b10	5	75	b15	39	33
b4	0	1	b6	6	39	b36	3	33
b9	9	29	b9	26	0			
b6	30	2	b14	31	6			

Tabela 2: Uso de relações espaciais no corpus GIVE-2 (Gargett et al., 2010).

por exemplo, casos em que todos participantes usaram a mesma estratégia de referência, como na descrição do alvo *b20* no mundo 1. Por outro lado, há também casos em que a divisão é de quase (50%), como no caso do alvo *b15* no mundo 3. Casos deste último tipo ilustram algumas das dificuldades de modelagem computacional do problema de geração de relações espaciais.

A tarefa de preparação de dados produziu um conjunto de descrições anotadas com informações sobre seus atributos atômicos e relacionais, bem como suas informações contextuais (i.e., o objeto-alvo e demais objetos do contexto, seus atributos atômicos e relacionais). Este conjunto foi então dividido em um conjunto de treinamento (794 instâncias) e teste (198 instâncias) selecionadas aleatoriamente. O uso destes conjuntos é descrito nas seções a seguir.

4 A Abordagem Proposta

Nesta seção apresentamos uma proposta de solução para o problema computacional de GER que faz uso de relações espaciais para descrever objetos do domínio GIVE (Koller et al., 2009). A proposta é dividida em duas etapas: a tarefa de seleção de relações espaciais adequadas a partir do contexto de entrada, e o algoritmo de GER propriamente dito.

4.1 Seleção de pontos de referência

A extração de relações espaciais a partir do contexto visual (cf. seção 3.1) produz uma ampla gama de possibilidades de referência. Entretanto, o fato de que dois objetos mantêm uma relação espacial entre si não necessariamente significa que esta relação seja uma forma típica ou aceitável de referência a estes objetos. Por exemplo, uma expressão como “a caixa que contém um relógio” pode não ser apropriada em um contexto com várias caixas, e no qual não seja possível re-

conhecer a relação de forma imediata (e.g., porque um objeto oculta o outro). Em outras palavras, uma relação espacial pode ser perfeitamente válida do ponto de vista semântico, mas de uso limitado para fins de referência por questões pragmáticas variadas.

Uma forma de decidir o que constitui ou não uma relação espacial válida para fins de referência, e assim filtrar casos que não deveriam ser considerados na produção destas expressões, é pela observação de exemplos de uso real da língua. A primeira etapa da solução proposta trata assim da tarefa de determinar quais objetos mantêm relações válidas - para fins de referência - com outros objetos do mesmo contexto.

Como forma de manter um certo grau de independência do domínio, a implementação deste módulo faz uso de uma abordagem de aprendizado de máquina semelhante à adotada em (Viethen, 2010), porém descartando-se características que não se aplicam ao domínio GIVE³, e acrescentando-se outras que capturam aspectos específicos da situação a ser tratada.

Dado um objeto-alvo *r* e um candidato a ponto de referência *o*, utilizamos um classificador binário *use_relation* para determinar se *r* pode ser referenciado via *o* através de uma relação espacial. As características de aprendizagem consideradas foram extraídas da porção de treinamento do corpus, e são sumarizadas na Tabela 3.

As características *distractors*, *landmarks*, *ambiguity* e *distance* são auto explicativos. As demais, que são baseadas em observações feitas em (Viethen, 2010) para os corpora GRE3D3 e GRE3D7, são detalhadas a seguir.

A definição de *unique_relation* é motivada pela observação feita em (Viethen, 2010) de que a relação espacial entre o alvo e o ponto de re-

³Por exemplo, no ambiente GIVE todos objetos de um mesmo tipo possuem o mesmo tamanho, o que torna pouco útil a definição de características de aprendizagem baseadas neste tipo de informação.

Característica	Descrição
<i>distractors</i>	quantidade de objetos do mesmo tipo que o alvo
<i>landmarks</i>	quantidade de objetos que mantêm relações espaciais com o alvo
<i>ambiguity</i>	quantidade de objetos iguais ao alvo e na mesma sala
<i>distance</i>	distância entre o alvo e o ponto de referência
<i>unique_relation</i>	indica se a relação entre o alvo e o ponto de referência é única no contexto
<i>most_salient_landmark</i>	indica se o ponto de referência é mais saliente que o alvo
<i>equal_landmarks</i>	indica se o ponto de referência é o do mesmo tipo e cor que o alvo

Tabela 3: Características de aprendizagem para a classe binária *use_relation*.

ferência tem mais chance de ser incluída em uma descrição quando for única, já que assim diminui-se a complexidade de identificação do alvo. Um exemplo de aplicação deste princípio em um ambiente GIVE seria um contexto em que deseja-se descrever um botão, e este é o único objeto do tipo botão que se encontra à esquerda de uma cadeira. Neste caso, o uso da relação espacial seria altamente recomendado. O mesmo não ocorreria, entretanto, em um contexto em que houvesse outros botões à esquerda de alguma cadeira.

A definição de *most_salient_landmark* é baseada na observação feita em (Viethen, 2010) de que uma relação espacial entre o alvo e o ponto de referência é mais utilizada quando o ponto de referência é mais saliente que o alvo. Um exemplo de aplicação deste princípio em um ambiente GIVE seria um contexto em que deseja-se referenciar um botão (que é um objeto pequeno) próximo a uma porta (que é um objeto grande), resultando em uma expressão do tipo ‘o botão ao lado da porta’. Para fins de implementação desta regra, a saliência das entidades é definida de acordo com seu tamanho físico. Outros fatores, como a distância entre o alvo e o ponto de referência, serão considerados na próxima seção, quando será discutido um ranking de relações espaciais. Deve-se ressaltar que as entidades consideradas próximas são as entidades que se encontram na mesma sala do respectivo mundo virtual.

Finalmente, a definição de *equal_landmarks* é também baseada em uma observação feita em (Viethen, 2010), segundo a qual uma relação espacial é preferível quando há similaridade visual entre o alvo e o ponto de referência. Um exemplo de aplicação deste princípio em um ambiente GIVE seria um contexto em que o alvo compartilha a mesma cor e mesmo tamanho que o ponto de referência, ou envolvendo objetos que não possuem variações de cor (e.g., sofás, cadeiras, portas, etc.), nos quais a similaridade visual se reduz à coincidência de tipos. Observa-se entretanto que no ambiente GIVE não existe variação de tamanho entre objetos do mesmo tipo, o que reduz a aplicabilidade da observação feita

em (Viethen, 2010), a qual tinha um sentido original mais amplo já que naquele domínio havia variação de tamanho e cor.

As instâncias de aprendizagem para a classe *use_relation* foram geradas da seguinte forma. Para cada objeto-alvo r do corpus, foi computada uma lista de n objetos que seriam candidatos a pontos de referência em uma possível descrição de r naquele contexto, utilizando-se o método de extração de relações espaciais do domínio descrito na seção 3.1.

A seguir, para cada par alvo-candidato $(r, o_{i=1..n})$, foi gerada uma instância de aprendizagem, totalizando assim n instâncias para cada objeto-alvo. O conjunto de n instâncias de cada objeto-alvo r foi rotulado da seguinte forma: se a descrição de r no corpus não usa uma relação espacial, então todas n instâncias são rotuladas como negativas (*use_relation* = *não*), isso é, nenhum dos candidatos pode ser recomendado como ponto de referência para r . Por outro lado, se a descrição do objeto r no corpus usou uma relação com um ponto de referência o , então a instância que representa o par (r, o) é rotulada como positiva (*use_relation* = *sim*), e todas outras $n-1$ instâncias como negativas.

Como resultado deste procedimento, foram geradas 3246 instâncias de aprendizagem compostas pelas características acima, sendo 335 positivas e 2911 negativas. A classificação propriamente dita foi realizada com o algoritmo de indução de árvores de decisão *J48* disponibilizado pelo pacote WEKA (Witten, Frank e Hall, 2011), utilizando-se *10-fold cross-validation* e demais parâmetros *default* do algoritmo (confiança $C=0,25$ e mínimo de $M=2$ instâncias por folha).

A Tabela 4 apresenta a matriz de confusão com base nos dados de treinamento. A Tabela 5 apresenta os resultados do treinamento de acordo com as medidas de precisão (P), cobertura (C) e medida F_1 . Cabe reiterar que o resultado da aplicação deste classificador aos dados de teste será tratado na seção 5.

	sim	não
usa relação	213	122
não usa relação	97	2814

Tabela 4: Matriz de confusão para *use_relation* sobre dados de treinamento.

use_relation	P	C	F_1
sim	0,68	0,63	0,66
não	0,95	0,96	0,96
média	0,93	0,93	0,93

Tabela 5: Resultados para *use_relation* sobre dados de treinamento.

Estes resultados podem ser considerados satisfatórios na medida em que o classificador objetiva modelar apenas o comportamento *médio* dos 72 participantes do experimento que deu origem ao corpus GIVE-2, os quais frequentemente adotam estratégias de referência conflitantes. Conforme discutido na secção 3.2, por exemplo, um mesmo botão pode ser descrito por participantes distintos como ‘o botão azul’ ou ‘o botão ao lado da porta’, dentre muitas outras possibilidades observadas no corpus.

O classificador foi incorporado ao ambiente GIVE e constitui o primeiro módulo a ser invocado na produção de uma descrição de um objeto-alvo *r*. Um exemplo completo é descrito a seguir, com base no contexto da Figura 2.

Neste exemplo, considera-se o objetivo de descrever o alvo *b2* (um botão amarelo) em um contexto contendo também uma cadeira *c1* e dois outros botões *b1* (vermelho) e *b3* (amarelo). Os rótulos associados a cada objeto, bem como a seta que aponta para o alvo *b2* são meramente ilustrativos, e não fazem parte da imagem real.

Dado o objetivo de descrever *b2*, o módulo classificador recebe como entrada as propriedades relacionais extraídas conforme discutido na secção 3.1, conforme a Tabela 6.



Figura 2: Exemplo de contexto visual.

Id	Propriedade	Distância
<i>r1</i>	<i>direita-b1</i>	0,30
<i>r2</i>	<i>atrás-c1</i>	1,25
<i>r3</i>	<i>esquerda-b3</i>	1,80

Tabela 6: Propriedades relacionais do alvo *b2*.

A seguir, é construído um conjunto de instâncias de teste *i1..i3* para verificar a possibilidade de uso de cada uma das relações *r1..r3* computadas. As instâncias assim obtidas são relacionadas na Tabela 7.

As instâncias *i1..i3* são submetidas à árvore de decisão, e o resultado da classificação determina se cada uma das relações correspondentes (*r1..r3*) deve ou não ser recomendada para uso pelo algoritmo de GER. Neste exemplo, apenas as relações *r1* e *r2* atendem aos critérios do classificador: a relação *r3*, a distância entre *b2* e *b3* foi considerada excessiva para fins de referência usando esta relação (ou, para ser mais exato, é superior à distância observada no corpus nos casos em que esta relação foi utilizada). As duas relações recomendadas serão fornecidas como entrada ao próximo módulo do sistema, a seguir.

4.2 Geração de expressões de referência usando relações espaciais

Utilizando-se o procedimento descrito na secção anterior é possível determinar, para uma determinada situação de referência a um objeto-alvo *r*, quais relações espaciais seriam adequadas para descrever *r* naquele contexto. A etapa seguinte consiste então em gerar a descrição propriamente dita, a qual pode ou não incluir uma das relações espaciais sugeridas.

O algoritmo proposto para este fim é uma versão modificada do algoritmo Incremental (Dale e Reiter, 1995) para manipular propriedades relacionais, e também integrado ao sistema GIVE (Koller et al., 2009).

Uma propriedade é incluída na expressão resultante desde que elimine ao menos uma en-

	#	<i>i1</i>	<i>i2</i>	<i>i3</i>
<i>distractors</i>	3	3	3	3
<i>landmarks</i>	3	3	3	3
<i>ambiguity</i>	2	2	2	2
<i>distance</i>	0,30	1,25	1,80	
<i>unique_relation</i>	true	false	true	
<i>most_salient_landmark</i>	false	true	false	
<i>equal_landmarks</i>	true	false	true	

Tabela 7: Instâncias de teste para o exemplo da Figura 2.

r	Propriedades selecionadas	Exemplo de realização possível
b1	$Ref-b1 = [(tipo-botão), (cor-vermelho)]$	o botão vermelho
b2	$Ref-b2 = [(tipo-botão), (cor-amarelo), (direita-b1)]$ $Ref-b1 = [(tipo-botão), (cor-vermelho)]$	o botão amarelo, à direita do botão vermelho
b3	$Ref-b3 = [(tipo-botão), (cor-amarelo)]$	o botão amarelo (ambíguo)
c1	$Ref-c1 = [(tipo-cadeira)]$	a cadeira

Tabela 8: Exemplos de descrições para os objetos da Figura 2.

tidade do contexto. Além disso, propriedades atômicas e relacionais são consideradas para inclusão em ordem de frequência, conforme observado na porção de treinamento do corpus GIVE-2, levando à definição da seguinte lista de preferência a ser usada como parâmetro P do algoritmo. Esta lista é única para todos tipos de objetos, ou seja, tanto para o alvo (que é sempre um objeto do tipo botão) como para os diversos tipos de objetos usados como pontos de referência (outros botões, cadeiras, mesas etc.):

$$P = \langle \text{tipo, cor, esq, dir, acima, frente, abaixo, atrás} \rangle$$

Observa-se também a ausência de relações espaciais de proximidade (e.g., *próximo, ao lado* etc.) na definição de P . No presente trabalho optamos por suprimir a geração de relações deste tipo em favor das formas mais específicas (e.g., *esquerda, direita, acima* etc.), as quais apresentam maior poder discriminatório e, conseqüentemente, podem levar à construção de expressões mais breves. Por exemplo, no caso da descrição de $b2$ no contexto da Figura 2, a expressão ‘o objeto à direita do botão vermelho’ seria preferível à forma ambígua ‘o objeto próximo ao botão vermelho’, que pode ser interpretada tanto como em referência a $b2$ como à cadeira $c1$.

Assim como em (Paraboni, 2000), no caso de expressões envolvendo um objeto e um ponto de referência, o algoritmo descreve cada objeto de forma independente. Isso pode, em alguns casos, acarretar superespecificação. Por exemplo, em um contexto com duas mesas e dois livros, em que apenas um dos livros está sobre uma das mesas, o presente algoritmo faria uso de propriedades adicionais, como em ‘o livro vermelho, sobre a mesa da esquerda’⁴. Esta medida foi adotada para evitar possíveis problemas de identificação de objetos em domínios espaciais complexos, como os discutidos em (Paraboni e van Deemter, 2013).

Na Tabela 8 são apresentados exemplos de

saídas do algoritmo para cada um dos botões do contexto da Figura 2, considerando-se as propriedades relacionais da Tabela 6 e a ordem de preferência P acima, e observando-se que cada descrição é composta de uma série de listas de propriedades (cláusulas Ref-) que podem referenciar outros objetos.

A referência a $b3$ neste exemplo permanece ambígua, pois neste caso o contexto oferecido como entrada para o algoritmo não contém propriedades em número suficiente para que o objeto possa ser identificado de forma única. A modelagem de uma relação *extremidade-direita*, por exemplo, permitira a geração de descrições como ‘o botão amarelo na extrema direita’.

5 Avaliação

5.1 Procedimento

Na avaliação do trabalho desenvolvido, considerou-se inicialmente a possibilidade de avaliação extrínseca, baseada no uso real do algoritmo de GER proposto em uma aplicação do tipo GIVE. O procedimento neste caso seria semelhante ao adotado em (Paraboni e van Deemter, 2013), medindo-se o desempenho de um usuário na tarefa de navegação ao reagir a uma série de instruções contendo expressões de referência produzidas pelo algoritmo proposto, ou por algum sistema de *baseline* usado como termo de comparação.

Uma análise dos diferentes tipos de expressões produzidas, e das limitações do ambiente GIVE, nos leva entretanto à conclusão de que este tipo de avaliação não seria factível. No ambiente GIVE, as únicas métricas de avaliação possíveis são aquelas baseadas na distância percorrida pelo usuário até aproximar-se do objeto-alvo, ou baseadas no tempo transcorrido entre a produção da descrição e a seleção do objeto referenciado. Estas métricas, entretanto, não nos permitiriam capturar de forma significativa a possível diferença entre alternativas como ‘pressione o botão vermelho’ e ‘pressione o botão ao lado do botão amarelo’, que são saídas típicas de um algoritmo

⁴Diferentemente de (Dale e Haddock, 1991), por exemplo, que neste caso permitiria mútua desambiguação como em ‘o livro sobre a mesa’.

do tipo proposto. Em outras palavras, diferenças como esta exigiriam recursos de medição mais sofisticados, como técnicas de *eye-tracking* (Koller et al., 2012) ou uso de imagens cerebrais (Engelhardt, Demiral e Ferreira, 2011), as quais estão fora do escopo deste trabalho.

Além da inadequação para o tipo de avaliação aqui exigido, medidas de tempo e distância em ambientes GIVE são também altamente sujeitas a ruído. Por exemplo, o usuário pode ter dificuldades de manipulação da interface, ou pode interromper a navegação para considerar uma decisão com mais cautela, dentre muitas outras possibilidades que tornam as métricas de tempo e distância ainda menos confiáveis (embora naturalmente não inviabilize a avaliação de outros aspectos de um sistema deste tipo, como o caso da geração de instruções de navegação abordada na série de desafios GIVE (Byron et al., 2009; Koller et al., 2010; Striegnitz et al., 2011)).

Em razão destas dificuldades, diversos aspectos do trabalho proposto foram assim avaliados de forma intrínseca com base no conjunto de teste extraído do corpus GIVE-2 (Gargett et al., 2010) descrito na seção 3.2. Para este fim, três aspectos de interesse foram considerados: a política de seleção da relação espacial, a ordem de preferência para seleção de propriedades (o parâmetro P do algoritmo de GER), e o tratamento de relações espaciais redundantes.

Com relação à política de seleção da relação espacial, consideramos duas alternativas: a proposta original - que seleciona a relação espacial *mais frequente* no corpus de treinamento - e uma alternativa na qual a propriedade espacial é selecionada de forma *aleatória*.

Com relação à ordem de preferência P utilizada pelo algoritmo, consideramos também duas alternativas: a proposta original - que faz ordenação *por frequência* conforme observado no conjunto de treinamento - e uma estratégia *gulosa* na qual propriedades de maior poder discriminatório (ou seja, aquelas que diferenciam o objeto-alvo do maior número possível de objetos do contexto) têm preferência.

Finalmente, com relação ao tratamento de relações espaciais redundantes, consideramos duas alternativas: a proposta original - que *não* inclui propriedades redundantes (i.e., aquelas que não contribuem para a desambiguação do referente, como em (Dale e Reiter, 1995)) - e uma estratégia na qual a relação espacial é selecionada mesmo que seja *redundante*.

A avaliação destes três aspectos da solução leva ao enunciado de $(2 \times 2 \times 2)$ 8 algoritmos distintos, dos quais a proposta original corres-

ponde à alternativa que seleciona a relação espacial mais frequente, ordena a lista de preferências P também por frequência, e inclui uma propriedade relacional na expressão apenas se esta for discriminatória.

Por simplicidade, todas alternativas avaliadas consideram como contexto de referência o conjunto de objetos na mesma sala onde se encontra o objeto-alvo. Entretanto, cabe observar que, no caso do corpus GIVE-2, o contexto utilizado pelos participantes do experimento foi, em alguns casos, formado apenas pelos objetos visíveis naquele instante. Por exemplo, quando o jogador já estava muito próximo do objeto referenciado pelo instrutor, em alguns casos este optou por desconsiderar os objetos mais distantes e produzir uma descrição breve como ‘o botão’, mesmo havendo outros botões na mesma sala. Uma vez que estes casos não foram contemplados na presente avaliação, assume-se assim um ônus para todos os algoritmos avaliados.

Cada uma das 198 situações de referência constantes no corpus de teste - aqui denominado conjunto *Referência* - foi fornecida como entrada para cada um dos 8 algoritmos, resultando assim em 8 conjuntos de expressões denominados *Sistema 1..8*. A avaliação propriamente dita consistiu em comparar cada conjunto *Sistema 1..8* com o conjunto *Referência*.

Para a comparação de cada um dos 1584 pares (8×198) *Sistema-Referência*, utilizamos duas métricas amplamente utilizadas (e.g., (Belz e Gatt, 2007; de Lucena, Paraboni e Pereira, 2010)): o coeficiente *Dice* (Dice, 1945), que mede o grau de similaridade entre os dois conjuntos de propriedades, assumindo um valor entre 0 (totalmente distintos) e 1 (idênticos); e *MASI* (Pasonneau, 2006), que possui correlação com *Dice*, porém atribuindo maior peso no caso de uma expressão ser subconjunto da outra.

Além de calcular coeficientes Dice e MASI, será considerada para fins ilustrativos a medida de exatidão, que é definida como o número de coincidência total entre cada par *Sistema-Referência* (número de vezes que o coeficiente Dice é igual a 1). Considerando-se entretanto que diversos aspectos da solução de GER estão fora do escopo deste trabalho (por exemplo, o tratamento de propriedades não espaciais que também fazem parte da expressão), a coincidência total em relação às descrições do corpus não seria um objetivo realista. Espera-se assim que os valores de exatidão sejam baixos para todos os algoritmos em questão, e que as métricas de menor granularidade Dice e MASI possam refletir de forma mais precisa o mérito de cada algoritmo.

5.2 Resultados

Antes de discutir a avaliação do sistema proposto e suas alternativas, será discutida a avaliação do classificador responsável pela seleção de possíveis pontos de referência (cf. seção 3.1) aplicado aos dados de teste. Este módulo é fixo para todas as versões do sistema consideradas nesta avaliação, e portanto seu desempenho não se reflete na análise a ser realizada.

A matriz de confusão obtida com base nos dados de teste é ilustrada na Tabela 9. O resultado obtido pelas etapas de extração e classificação de pontos de referência é sumarizado na Tabela 10.

Os 198 itens de teste apresentaram um total de 86 expressões contendo relações espaciais (43,43%). Destas, 36 itens (18,18%) faziam referência ao próprio ouvinte receptor da expressão, como em ‘pressione o botão ao seu lado’. Conforme discutido na seção 3.2, estes casos não foram cobertos pela presente proposta e acarretam assim uma margem de erro a todas as implementações avaliadas. Os 8 algoritmos e seus resultados são sumarizados na Tabela 11. O algoritmo originalmente proposto é o primeiro (#1).

As seguintes comparações entre algoritmos foram realizadas utilizando-se o teste de *Wilcoxon* sobre coeficientes *Dice*. Primeiramente, observa-se que o algoritmo proposto apresenta os melhores resultados para todas as três métricas de avaliação. Entretanto, seus resultados são idênticos ao da alternativa #5, que difere apenas na estratégia de seleção. Em outras palavras, não houve diferença significativa entre a seleção da propriedade mais frequente e a simples seleção aleatória. Este resultado se explica pelo fato de que nos dados de teste a maioria dos objetos-alvo só possui um ponto de referência possível.

Em segundo lugar, a ordenação por frequência é significativamente superior à ordenação gulosa para todos os pares de algoritmos avaliados (i.e., comparando-se #1 com #3, #2 com #4 etc.) ($W=16110$, $Z=11,6$, $p < 0,001$). Este resultado contrasta propostas de algoritmos para geração de descrições breves ou mínimas como em (Gardent, 2002). Por outro lado, a estratégia de não inserir redundância explícita (isso é, evitando a inclusão de atributos desnecessários para desambiguação) é significativamente superior ao

	sim	não
usa relação	13	4
não usa relação	9	172

Tabela 9: Matriz de confusão para *use_relation* sobre dados de teste.

seu uso para todos os pares de algoritmos (i.e., comparando-se #1 com #2, #3 com #4 etc.). A menor diferença observada, mas ainda altamente significativa, foi entre os algoritmos #7 e #8 ($W=986$, $Z=5,75$, $p < 0,001$).

Quanto à exatidão dos algoritmos propostos, observamos ainda que os algoritmos #1 e #5 relevaram-se superiores a todos os demais, sendo os únicos a gerar descrições completamente idênticas às do corpus de teste.

6 Discussão

Este artigo descreveu o desenvolvimento e avaliação de um algoritmo de GER que faz uso de relações espaciais em ambientes do tipo GIVE (Koller et al., 2009). Apesar da relativa simplicidade da proposta, este estudo contribuiu para o entendimento de três aspectos do problema: a política de seleção de propriedades espaciais, a ordenação das propriedades consideradas pelo algoritmo de GER, e o tratamento de propriedades redundantes.

Mesmo não tendo sido observada uma diferença significativa entre as alternativas de seleção de propriedade espacial, as estratégias de ordenação baseada em frequência e de inclusão de propriedades espaciais discriminatórias revelaram-se superiores às demais. Em outras palavras, não parece haver uma preferência geral por propriedades discriminatórias, mas as propriedades espaciais parecem ser empregadas predominantemente desta forma.

Como trabalho futuro planejamos expandir a presente análise para incluir aspectos adicionais da solução, tais como a prioridade da relação espacial considerada pelo algoritmo, e uma política de seleção de relações espaciais mais sofisticada. Além disso, esperamos testar a substituição do presente modelo de classificação baseado em árvores de decisão por outros, como máquinas de vetor de suporte recentemente aplicadas ao problema de GER em (Ferreira e Paraboni, 2014).

use_relation	P	C	F_1
sim	0,59	0,77	0,67
não	0,98	0,95	0,96
média	0,94	0,93	0,94

Tabela 10: Resultados para *use_relation* sobre dados de teste.

#	Seleção	Ordenação	Redundância	Dice	MASI	Exatidão
1	mais frequente	por frequência	não	0.73	0.49	0.34
2	mais frequente	por frequência	sim	0.58	0.22	0.00
3	mais frequente	gulosa	não	0.28	0.11	0.00
4	mais frequente	gulosa	sim	0.21	0.05	0.00
5	aleatória	por frequência	não	0.73	0.49	0.34
6	aleatória	por frequência	sim	0.58	0.22	0.00
7	aleatória	gulosa	não	0.25	0.10	0.00
8	aleatória	gulosa	sim	0.20	0.05	0.00

Tabela 11: Resultados.

Referências

- Akkersdijk, S., M. Langenbach, F. Loch, e M. Theune. 2011. The Thumbs Up! Twente system for GIVE 2.5. Em *Generation Challenges Session at ENLG-2011*, pp. 312–317.
- Aziz, Wilker Ferreira, Thiago Alexandre Salgueiro Pardo, e Ivandré Paraboni. 2008. An experiment in Spanish-Portuguese statistical machine translation. *Advances in Artificial Intelligence-SBIA 2008*, LNAI 5249:248–257.
- Baltaretu, Adriana Alexandra, Emiel Kraemer, e Alfons Maes. 2013. Factors influencing the choice of relatum in referring expressions generation: animacy vs. position. Em *CogSci workshop on the production of referring expressions: bridging the gap between cognitive and computational approaches to reference (PRE-CogSci-2013)*, pp. 1–6.
- Barclay, Michael e Antony Galton. 2008. An influence model for reference object selection in spatially locative phrases. *Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, LNCS 5248:216–232.
- Belz, A. e A. Gatt. 2007. The attribute selection for GRE challenge: Overview and evaluation results. Em *UCNLG+MT: Language Generation and Machine Translation*.
- Braunias, J., U. Boltz, M. Drager, B. Fersing, e O. Nikitina. 2010. The GIVE-2 challenge: Saarland NLG system. Em *INLG-2010*.
- Byron, Donna, Alexander Koller, Kristina Striegnitz, Justine Cassell, Robert Dale, Johanna Moore, e Jon Oberlander. 2009. Report on the first NLG challenge on generating instructions in virtual environments (GIVE). Em *ENLG-2009*.
- Cuevas, R. e Ivandré Paraboni. 2008. A machine learning approach to portuguese pronoun resolution. *Advances in Artificial Intelligence-IBERAMIA 2008*, LNAI 5290:262–271.
- Dale, R. e N. J. Haddock. 1991. Content determination in the generation of referring expressions. *Computational Intelligence*, 7:252–265.
- Dale, R. e E. Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19.
- Dale, Robert e Jette Viethen. 2009. Referring expression generation through attribute-based heuristics. Em *Proceedings of the 12th European Workshop on Natural Language Generation, ENLG '09*, pp. 58–65. Association for Computational Linguistics.
- de Lucena, Diego Jesus, Ivandré Paraboni, e Daniel Bastos Pereira. 2010. From semantic properties to surface text: The generation of domain object descriptions. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 14(45):48–58.
- de Novaes, Eder Miranda e Ivandré Paraboni. 2012. Portuguese text generation using factored language models. *Journal of the Brazilian Computer Society*, 19(2):135–146.
- Dice, L. R. 1945. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.
- Engelhardt, Paul E., S. B. Demiral, e Fernanda Ferreira. 2011. Over-specified referring expressions impair comprehension: An ERP study. *Brain and Cognition*.
- Ferreira, Thiago Castro e Ivandré Paraboni. 2014. Classification-based referring expression generation. *Lecture Notes in Computer Science*, 8403:481–491.
- Gardent, Claire. 2002. Generating minimal definite descriptions. Em *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, pp. 96–103, Stroudsburg, PA, USA. Association for Computational Linguistics.

- Gargett, A., K. Garoufi, Alexander Koller, e K. Striegnitz. 2010. The GIVE-2 corpus of giving instructions in virtual environments. Em *LREC-2010*.
- Garoufi, K. e A. Koller. 2011. The Potsdam NLG systems at the GIVE-2.5 challenge. Em *Generation Challenges Session at ENLG-2011*, pp. 307–311.
- Gatt, A., A. Belz, e E. Kow. 2008. The TUNA challenge 2008: Overview and evaluation results. Em *INLG-2008*, pp. 198–206.
- Gatt, A., A. Belz, e E. Kow. 2009. The TUNAREG challenge 2009: Overview and evaluation results. Em *ENLG-2009*, pp. 174–182.
- Gatt, Albert, Ilka van der Sluis, e Kees van Deemter. 2007. Evaluating algorithms for the generation of referring expressions using a balanced corpus. Em *11th European Workshop on Natural Language Generation (ENLG-07)*.
- Grice, H. P. 1975. Logic and conversation. Em Peter Cole e Jerry L. Morgan, editores, *Syntax and semantics*, volume 3. New York: Academic Press.
- Kelleher, John D. e Fintan J. Costello. 2009. Applying computational models of spatial prepositions to visually situated dialog. *Computational Linguistics*, 35(2):271–306, June, 2009.
- Koller, A., D. Byron, J. Cassell, R. Dale, K. Striegnitz, J. Moore, e J. Oberlander. 2009. The software architecture for the first challenge on generating instructions in virtual environments. Em *EACL-2009*.
- Koller, Alexander, Konstantina Garoufi, Maria Staudte, e Matthew Crocker. 2012. Enhancing referential success by tracking hearer gaze. Em *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial)*, pp. 30–39.
- Koller, Alexander, Kristina Striegnitz, Andrew Gargett, Donna Byron, Justine Cassell, Robert Dale, Johanna Moore, e Jon Oberlander. 2010. Report on the second NLG challenge on generating instructions in virtual environments (GIVE-2). Em *INLG-2010*.
- Krahmer, E. e Kees van Deemter. 2012. Computational generation of referring expressions: A survey. *Computational Linguistics*, 38(1):173–218.
- Krahmer, Emiel e Mariët Theune. 2002. Efficient context-sensitive generation of referring expressions. Em *Information sharing: Reference and presupposition in language generation and interpretation*, volume 143. CSLI Publications, California, pp. 223–263.
- Paraboni, Ivandré. 1997. Uma arquitetura para a resolução de referências pronominais possessivas no processamento de textos em língua portuguesa. Tese de Mestrado, PUCRS.
- Paraboni, Ivandré. 2000. An algorithm for generating document-deictic references. Em *Procs. of the INLG-2000 workshop Coherence in Generated Multimedia*, pp. 27–31, Mitzpe Ramon.
- Paraboni, Ivandré e Kees van Deemter. 1999. Issues for the generation of document deixis. Em *Procs. of workshop on Deixis, Demonstration and Deictic Belief in Multimedia Contexts, in association with the 11th European Summers School in Logic, Language and Information (ESSLLI-99)*, pp. 44–48.
- Paraboni, Ivandré e Kees van Deemter. 2013. Reference and the facilitation of search in spatial domains. *Language and Cognitive Processes*, online.
- Passonneau, Rebecca. 2006. Measuring agreement on set-valued items (MASI) for semantic and pragmatic annotation. Em *LREC-2006*.
- Pereira, D. B. e I. Paraboni. 2008. Statistical surface realisation of portuguese referring expressions. Em *Advances in Natural Language Processing*, volume LNAI 5221. Springer-Verlag, pp. 383–392.
- Pereira, Daniel Bastos e Ivandré Paraboni. 2007. A language modelling tool for statistical NLP. Em *5th Workshop on Information and Human Language Technology (TIL-2007)*, pp. 1679–1688, Rio de Janeiro. Sociedade Brasileira de Computação.
- Schutte, N. e N. Dethlefs. 2010. The Dublin-Bremen system for the GIVE-2 challenge. Em *INLG-2010*.
- Striegnitz, K., A. Denis, A. Gargett, K. Garoufi, A. Koller, e M. Theune. 2011. Report on the second second challenge on generating instructions in virtual environments (GIVE-2.5). Em *Generation Challenges Session at ENLG-2011*, pp. 270–279.
- Teixeira, Caio V. M., Ivandré Paraboni, Adriano S. R. da Silva, e Alan K. Yamasaki. 2014. Generating relational descriptions involving mutual disambiguation. *Lecture Notes in Computer Science*, 8403:492–502.

- Tenbrink, T. 2005. Identifying objects on the basis of spatial contrast: An empirical study. Em C. Freksa, M. Knauff, B. Krieg-Bruckner, B. Nebel, e T. Thomas Barkowsky, editoras, *Spatial Cognition IV: Reasoning, Action, Interaction. International Conference Spatial Cognition 2004*. Springer, pp. 124–146.
- Viethen, H. A. E. 2010. *The Generation of Natural Descriptions: Corpus-based Investigations of Referring Expressions in Visual Domains*. Tese de doutoramento, Macquarie University.
- Viethen, J. e R. Dale. 2011. GRE3D7: A corpus of distinguishing descriptions for objects in visual scenes. Em *UCNLG+Eval: Language Generation and Evaluation Workshop*, pp. 12–22, Edinburgh, Scotland.
- Witten, I. H., E. Frank, e M. A. Hall. 2011. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufman Publishers, Burlington, MA, 3rd edition.