

Generating Customizable Natural Language Descriptions

A. Costa, and I. Paraboni

Abstract—In the Natural Language Generation field, the task of Reference Expression Generation (REG) consists of selecting the semantic content that should compose the linguistic description of a discourse object, as in ‘the man in black’ or ‘the boy in the centre’. REG algorithms tend to use predefined parameters to model qualitative and quantitative aspects of the description to be generated, and are often difficult to adapt to new domains or applications. In addition, obtaining linguistic examples in all possible reference situations is also often impracticable, which limits the use of machine learning approaches in this task. Based on these observations, this work proposes an extension of a standard REG algorithm, in which certain aspects of a reference strategy are customized without necessarily resorting to training data. The proposal was tested by making use of three REG domains, and its results compared favourably against a standard REG approach.

Index Terms—Natural Language Processing, Text Generation, Definite Descriptions.

I. INTRODUÇÃO

SISTEMAS de Geração de Língua Natural (GLN em Português, ou NLG no original em inglês) [1] - que tipicamente produzem descrições textuais a partir de uma entrada de dados não linguística - são empregados quando o uso de texto predefinido não é considerado suficiente, ou seja, quando é necessária uma maior variação linguística de saída, ou quando deseja-se maior proximidade em relação ao desempenho humano. Sistemas deste tipo possuem uma ampla gama de aplicações de transformação de dados estruturados em língua natural, facilitando a interpretação de grandes volumes de dados, e oferecendo suporte à modelagem de agentes virtuais como tutores, personagens de jogos e outros.

Na arquitetura de um sistema de GLN, uma das etapas fundamentais trata da seleção de conteúdo semântico, ou seja, da tarefa de computar quais os significados a serem realizados em língua natural. Considere-se por exemplo a cena na Fig. 1, na qual uma aplicação de GLN (e.g., associada a um jogo, guia virtual etc.) poderia gerar instruções do tipo ‘Agora converse com o rapaz com os braços à cabeça’.

Decidir qual informação incluir em uma instrução deste tipo pode, em certo sentido, parecer trivial, já que boa parte da informação a ser representada poderia em muitos casos ser considerada fixa ou predefinida pela aplicação. Assim, poderíamos simplesmente fazer uso de um *template* do tipo ‘Agora converse com *X*’, onde *X* corresponde à porção variável do texto a ser gerado. Grande parte do desafio



Fig. 1. Exemplo de contexto com objetos referenciáveis e1..e4.

computacional se concentra, entretanto, em como descrever *X*.

Uma descrição de *X* como neste exemplo envolve o uso de uma expressão de referência que permita distinguir o objeto-alvo em questão dos demais elementos contextuais. Para objetos que possuem nomes próprios, ou que foram recentemente citados no discurso, produzir uma expressão de referência é relativamente simples: basta fazer uso do nome adequado (e.g., ‘João’) ou de uma forma pronominal compatível com a saliência do objeto-alvo na estrutura do discurso (e.g., ‘ele’). Por outro lado, para todas as situações que não atendem a estes critérios (ou seja, para descrever objetos sem um nome próprio conhecido, e que não foram citados recentemente), a estratégia referencial a ser adotada costuma ser o uso de uma descrição definida (e.g., ‘o homem de blusão escuro’) ou, em caso de ambiguidade, indefinida (e.g., ‘um rapaz mais ao centro’).

A determinação do conteúdo semântico a ser expresso na forma de descrições definidas constitui a tarefa computacional de seleção de conteúdo referencial, ou Geração de Expressões de Referência (GER em Português, ou REG do original em inglês) [2]. De forma simplificada, esta tarefa objetiva decidir o que dizer, ou seja, quais propriedades semânticas de um objeto-alvo devem ser realizadas em uma descrição linguística [3]. GER é uma ativa linha de pesquisa em GLN, e será também o tema do presente trabalho.

Descrições definidas são dependentes do contexto da comunicação, e no exemplo acima poderiam variar em virtude do ângulo ou perspectiva do usuário ou ouvinte, de propriedades dinâmicas do objeto-alvo (e.g., se o rapaz cruzar os braços, mover-se, ou sofrer oclusão por parte de outro objeto ou personagem da cena), e podem ainda depender do que foi dito anteriormente (tanto pelo sistema como pelo usuário, se for o caso), dentre muitas outras possibilidades. Assim, a tarefa

A. C. Mariotti was with the Universidade de São Paulo, São Paulo, Brazil e-mail: drehmariotti@gmail.com.

I. Paraboni is with the Universidade de São Paulo, São Paulo, Brazil e-mail: ivandre@usp.br.

de GER não é facilmente implementável com uso de *templates* de texto predefinido, e pode representar um obstáculo inclusive para métodos de aprendizagem de máquina (em especial, se a construção de um corpus de treinamento de ampla cobertura não for factível).

Como objetivo mais básico de um algoritmo de GER, espera-se que a descrição gerada seja livre de ambiguidade, mas este não é o principal desafio computacional em problemas deste tipo. Tão ou mais importante do que evitar ambiguidade é decidir o que exatamente constitui uma descrição adequada do ponto de vista humano, ou seja, decidir quantas e quais propriedades do objeto-alvo ela deve expressar de modo a obter uma descrição semelhante àquela que um locutor humano produziria nas mesmas circunstância.

Tendo em vista o objetivo de modelar o comportamento humano da forma mais fiel possível, algoritmos de GER tendem a utilizar parâmetros ou regras representando diversos aspectos qualitativos e quantitativos do tipo de descrição a ser gerada [3]. No entanto, soluções deste tipo podem ser de difícil adaptação a novos domínios ou aplicações e, ao mesmo tempo, a obtenção de exemplos linguísticos de todas situações de referência possíveis para treinamento de modelos de aprendizagem de máquina é muitas vezes inviável [4].

Com base nestas observações, e objetivando obter mais flexibilidade e precisão em algoritmos de GER sem a coleta de grandes volumes de exemplos de treinamento, este trabalho apresenta uma proposta de extensão do algoritmo de GER padrão apresentado em [3], na qual passa a ser possível customizar certos aspectos da estratégia de referência. A proposta foi avaliada com base em diferentes corpus de descrições atômicas (i.e., sem uso de relações espaciais entre objetos, conforme será discutido a seguir), e apresenta resultados superiores aos da abordagem original sem customização.

II. CONCEITOS BÁSICOS

A. A Seleção de Conteúdo Referencial para GER

A tarefa de seleção de conteúdo para GER foi formalizada em uma série de estudos culminando na proposta da abordagem Incremental em [3], que possivelmente é a mais tradicional da área. Embora universalmente conhecida como um algoritmo, entretanto, esta abordagem não representa uma solução completa para o problema de GER, tratando apenas de estabelecer a entrada e saída esperadas, e apresentando uma estratégia geral de seleção de conteúdo motivada por estudos da área de Psicologia.

A seguir são descritos de forma simplificada os conceitos mais importantes desta abordagem como forma de ilustrar o problema computacional de GER e alguns de seus desafios, tomando como exemplo prático mais uma vez o contexto da Fig. 1. A entrada de um algoritmo de GER típico é um contexto C formado pelo objeto-alvo ou referente r que se deseja descrever e por um grupo de objetos denominados distraidores dos quais deseja-se distinguir o alvo r . Tanto o alvo como os distraidores são representados por conjuntos de propriedades semânticas na forma de pares (atributo-valor), como em (*hair-dark*), representando o fato de que um determinado objeto do domínio possui a propriedade de 'ter

cabelos escuros'. Propriedades intrínsecas a um objeto são ditas atômicas (e.g., a propriedade de ter olhos castanhos), e propriedades que representam uma relação com outros objetos do contexto são ditas relacionais (e.g., a propriedade de estar próximo de uma determinada pessoa, de ser mais jovem do que ela etc.).

Uma possível representação do contexto de referência ilustrado pela cena da Fig. 1 é apresentada na Tabela I a seguir. Neste exemplo, nomes de atributos e valores foram mantidos em inglês apenas para enfatizar a distinção entre seleção de conteúdo (semântico) e realização superficial (texto, não tratado pelo algoritmo), ou seja, para enfatizar o fato de que representam conceitos, e não palavras.

TABELA I
PROPRIEDADES SEMÂNTICAS (PARES ATRIBUTO-valor) REPRESENTANDO O CONTEXTO DA FIG. 1

Entidade	Gender	Arms	Hair	Direction
<i>e1</i>	<i>male</i>	<i>crossed</i>	<i>blonde</i>	<i>back</i>
<i>e2</i>	<i>female</i>	<i>crossed</i>	<i>dark</i>	<i>left</i>
<i>e3</i>	<i>male</i>	<i>on head</i>	<i>dark</i>	<i>front</i>
<i>e4</i>	<i>female</i>	<i>crossed</i>	<i>dark</i>	<i>right</i>

Por simplicidade, neste exemplo consideramos apenas quatro propriedades atômicas de cada um dos objetos do tipo pessoa observados na cena. Na prática, entretanto, esta base de conhecimento poderia ser estendida com a inclusão de muitos outros objetos referenciáveis (e.g., mesas, cadeiras, partes do corpo de cada indivíduo, peças de vestuário etc.) e com a definição de diversas propriedades relacionais. Por exemplo, o fato de um objeto estar próximo a *e3* poderia ser representado por uma propriedade relacional do tipo (*near-e3*). Nota-se entretanto que propriedades relacionais não são limitadas a este caso (talvez mais óbvio) de relação espacial, podendo representar qualquer tipo de relação entre dois objetos. Por exemplo, a observação de que uma determinada pessoa é mais jovem do que *e3* poderia ser representada por uma propriedade do tipo (*younger.than-e3*).

A pesquisa em GER e áreas correlatas apresenta inúmeras variações e extensões do problema básico aqui discutido, as quais estariam fora do escopo da presente discussão. Questões importantes da área incluem a geração de descrições mínimas [5] e a modelagem de propriedades relacionais [6], dentre muitas outras. Uma revisão detalhada da área e de suas abordagens é apresentada em [7]. Na seção a seguir será tratado exclusivamente do algoritmo Incremental apresentado em [3].

B. A Abordagem Incremental

O algoritmo Incremental em [3] é ilustrado pela Fig. 2. Dado um contexto C e um objeto-alvo r que se deseja identificar, a seleção de conteúdo para GER consiste em produzir um (sub)conjunto L de propriedades de r tal que L seja capaz de distinguir r de todos os distraidores em C .

Em linhas gerais, o algoritmo Incremental funciona da seguinte forma. As propriedades são incluídas (linha 7) em L incrementalmente (de onde provém o nome do algoritmo) segundo uma ordem preferencial P predefinida (linha 4), desde

```

1 MakeReferringExpression( $r, C, P$ )
2    $L \leftarrow \{\}$ 
3    $C \leftarrow C - \{r\}$ 
4   for  $A_i \in P$  do
5      $V \leftarrow value(r, A_i)$ 
6     if  $|rulesOut(\langle A_i, V \rangle)| > 0$  then
7        $L \leftarrow L \cup \langle A_i, V \rangle$ 
8        $C \leftarrow C - rulesOut(\langle A_i, V \rangle)$ 
9     end
10    if  $|C| == 0$  then
11      return  $L$ 
12    end
13  end
14  return  $L$ 

```

Fig. 2. O algoritmo Incremental - adaptado de [3].

que contribuam (linhas 5-6) para a desambiguação do referente (i.e., desde que excluam pelo menos um objeto do contexto C). A cada seleção, os objetos distraidores que não possuem a propriedade selecionada são descartados com a atualização do contexto C (8).

O procedimento é repetido (4-13) até que um conjunto único (i.e., livre de ambiguidade) de propriedades seja obtido, ou seja, até que o conjunto de distraidores em C esteja vazio (10), ou até que todas as propriedades possíveis em P tenham sido consideradas (13). No primeiro caso, L poderia ser posteriormente realizada como uma descrição definida. No segundo caso L permaneceria ambígua, e poderia ser realizada, por exemplo, como uma descrição indefinida.

Um exemplo prático de funcionamento deste algoritmo pode ser descrito da seguinte forma. Suponha-se uma ordem de preferência $P = \{gender, arms, hair, direction\}$, e o objeto-alvo $r = e2$. Neste caso, o algoritmo consideraria primeiro o atributo *gender*, o qual seria incluído na descrição resultante L porque seu valor *female* elimina dois elementos do contexto ($e1$ e $e3$, cujo atributo *gender* possui o valor *male*). A seguir, o atributo *arms* é considerado para inclusão, mas este não é selecionado porque seu valor (*crossed*) não descarta o único elemento restante no contexto ($e4$), que também possui este mesmo valor. O atributo *hair* é também desconsiderado por este mesmo motivo, e finalmente o atributo *direction* é incluído porque seu valor (*left*) descarta $e4$ (cuja direção é à direita). Assim, a expressão resultante $L = \{gender-female, direction-left\}$ descreve de forma única o objeto-alvo $e2$, e poderia ser subsequentemente realizada, por exemplo, como ‘a moça voltada para a esquerda’. De forma análoga, $e1$ poderia ser descrito como ‘o rapaz de braços cruzados’, $e3$ poderia ser descrito como ‘o rapaz com os braços à cabeça’ e $e4$ como ‘a moça voltada para a direita’.

Apesar de evidentemente simples, a estratégia básica de GER ilustrada pelo algoritmo Incremental e outros tende a apresentar resultados suficientemente positivos em domínios restritos e/ou aplicações em que a variação da língua a ser gerada não seja uma questão crucial. No entanto, esta estratégia não contempla diversas questões que passaram a ser consideradas importantes em modelos de GER em anos

recentes, como a da naturalidade da expressão (ou *humanlikeness*) [8]. Em especial, observa-se que a ordem preferencial P - cuja definição é deixada em aberto em [3] justamente por ser dependente de domínio - tem grande impacto sobre o tipo de expressão produzida pelo algoritmo. A depender de como a lista P for ordenada, o algoritmo pode produzir descrições muito distintas. No exemplo anterior, se fosse considerada a ordem $P = \{gender, hair, arms, direction\}$, $e2$ continuaria sendo descrito como ‘a moça voltada para a esquerda’. No entanto, esta ordenação levaria $e1$ a ser descrito como ‘o rapaz loiro’, e $e3$ como ‘o rapaz de cabelos pretos’.

Além do papel crucial do parâmetro P sobre o funcionamento do algoritmo, observa-se que o problema computacional de GER apresenta também uma série de questões referentes ao volume de informação incluído na descrição. Em especial, observa-se que locutores humanos normalmente não se limitam a dizer o que é estritamente necessário para a identificação do objeto-alvo. Ao contrário, muitas das expressões que produzimos são do tipo superespecificado, ou seja, contendo propriedades redundantes que poderiam em tese ser omitidas sem prejuízo à identificação. A superespecificação é inerente à produção de língua natural [9], e se torna especialmente problemática para um algoritmo de GER que tenha como objetivo reproduzir o comportamento humano, uma vez que os critérios para inclusão de informação redundante são pouco conhecidos e, conseqüentemente, de difícil formalização. Enquanto atributos altamente discriminatórios (como a informação de gênero no exemplo anterior) são facilmente selecionados por muitos tipos de algoritmos de GER, o mesmo não pode ser dito de atributos que não sejam estritamente necessários para a desambiguação (como o fato de a pessoa em questão estar de pé, que é uma informação redundante que locutores humanos tendem a incluir com frequência em suas descrições).

Com base nestas considerações, diversas extensões do algoritmo Incremental, cada uma tipicamente enfocando um aspecto específico do fenômeno de referência, têm sido propostas. Iniciativas deste tipo incluem a execução do algoritmo Incremental em chamadas recursivas para geração de expressões relacionais [10], o reuso de propriedades recentemente selecionadas para maior coesão em diálogos automatizados [11], e o cômputo de listas de preferência P individuais para cada locutor de um córpus de modo a obter descrições mais personalizadas [12], dentre outras.

C. Córpis para GER

A evolução da pesquisa em GLN levou ao desenvolvimento de uma série de conjuntos de dados dedicados ao estudo e desenvolvimento de técnicas de GER em anos recentes. A construção destes recursos é normalmente implementada na forma de experimentos controlados envolvendo participantes humanos engajados na tarefa de produzir descrições a partir de imagens como a da Fig. 1 anterior. Após a coleta e anotação dos dados, o resultado final - ou córpus para GER - é assim formado pela representação semântica dos contextos de referência e das descrições neles produzidas. Alguns exemplos de córpus construídos de acordo com esta metodologia, e que serão utilizados para treinamento e teste da proposta

descrita na próxima seção, são descritos brevemente a seguir. Em todos os casos, a tarefa dos participantes recrutados em cada experimento consistia em descrever um dado objeto-alvo apresentado em destaque em cada cena.

Um exemplo proeminente de cópua para GER é o resultado do projeto TUNA [13], um conjunto de descrições definidas produzidas por locutores humanos para apoio ao estudo de fenômenos de referência e algoritmos de GER. O cópua TUNA contempla situações de referência em dois domínios: peças de mobília (*Furniture*) e fotos de pessoas (*People*). Um exemplo destes domínios é ilustrado na Fig. 3. Nestes exemplos, o objeto-alvo da imagem esquerda (domínio *Furniture*) poderia ser descrito como, e.g., ‘a cadeira verde’, e o da direita (*People*) como, e.g., ‘o homem de barba branca, olhando para a esquerda’.

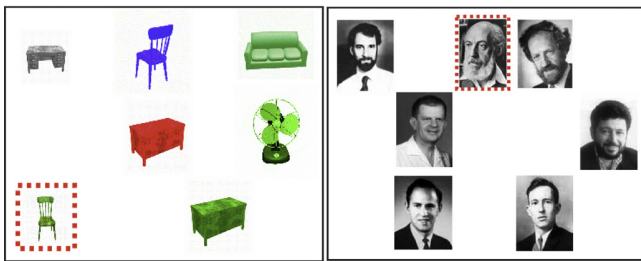


Fig. 3. Imagens de estímulo dos cópua TUNA *Furniture* (esquerda) e TUNA *People* (direita), adaptado de [14].

Mais recentemente, o cópua b5-ref [4] contempla imagens realistas com potencial de explicitar diferenças entre traços de personalidade [15] dos locutores que as descrevem. Os contextos de referência considerados fazem uso de imagens extraídas da base *Face Place* [16] de fotografias humanas validadas para diversos tipos de emoções e características físicas (as imagens *Face Place* são cortesia de Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University). Um exemplo de imagem de estímulo deste tipo é ilustrado pela Fig. 4, na qual o objeto-alvo poderia ser descrito como, e.g., ‘A moça loira que parece chateada’.

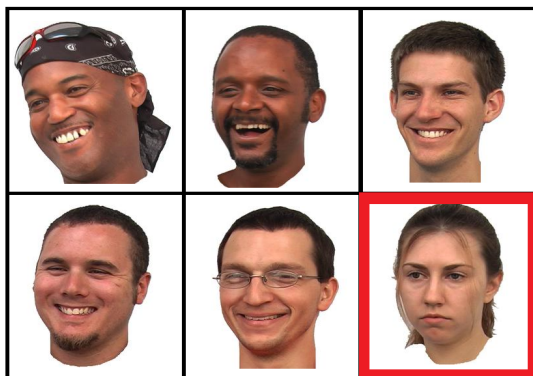


Fig. 4. Imagens de estímulo do cópua b5-ref [4].

Descrições do tipo TUNA *Furniture*, TUNA *People* e b5-ref serão utilizadas na avaliação do algoritmo proposto na próxima seção. Diversos outros conjuntos de dados semelhantes são

entretanto disponibilizados na literatura em GER e áreas correlatas, contemplando fenômenos linguísticos diversos como o uso de relações espaciais [17], [18], a questão saliência relativa [19], [20], situações de referência de alto realismo [21], [22] e em mundos virtuais interativos [23], [24], dentre muitos outros.

III. GER CUSTOMIZÁVEL

Nesta seção descrevemos a proposta de um novo algoritmo de GER customizável – aqui denominado IAL – que, assim como muitos outros estudos da área de REG (cf. Section II-B), propõe uma extensão da abordagem Incremental em [3]. O objetivo específico desta proposta é o de gerar descrições mais próximas das que seriam produzidas por locutores humanos por meio de uma seleção de conteúdo mais precisa – tanto em termos qualitativos como quantitativos – do que seria obtido pelo algoritmo Incremental original, e sem recorrer a grandes volumes de dados de treinamento. Para este fim, IAL contempla duas modificações sobre o algoritmo Incremental: o acréscimo de um novo parâmetro de entrada (*level*, de onde provém o ‘L’ do nome do algoritmo) que passa a controlar a condição de parada do algoritmo, e o uso de listas de propriedades preferenciais (i.e., o parâmetro *P* do algoritmo original) customizadas por contexto, e não apenas para o domínio como um todo. Estas duas extensões são discutidas individualmente a seguir.

A. Geração de Descrições de Tamanho Customizável

A abordagem Incremental em [3] considera como condição de parada a obtenção de uma descrição livre de ambiguidade (ou, na impossibilidade de obter este resultado, o esgotamento da lista de propriedades a considerar). Conforme já discutido, entretanto, esta estratégia pode não refletir o comportamento de locutores humanos, que frequentemente acrescentam à descrição mais propriedades do que seria estritamente necessário para a desambiguação. Considerando-se que este comportamento pode ser ao menos em parte influenciado pelo contexto (i.e., em certas situações pode haver mais tendência à superespecificação do que em outras), a primeira extensão proposta no presente trabalho é de caráter quantitativo, e consiste na manipulação de condição de parada do algoritmo Incremental de modo a gerar descrições do tamanho médio esperado para uma determinada situação.

De forma mais específica, propomos utilizar um parâmetro de entrada adicional *level* para representar o número médio de propriedades superespecificadas (ou redundantes) que tende a ser acrescentado em descrições de um dado contexto. O valor de *level* pode ser estimado de diversas formas e, havendo disponibilidade de dados de treinamento no domínio em questão, pode ser obtido de forma trivial a partir de um número reduzido de exemplos linguísticos fornecidos, conforme proposto em [25]. Este método é o adotado na avaliação discutida na próxima seção, que é baseada em dados de cópua de GER existentes. Alternativamente, este parâmetro poderia também ser estimado de forma empírica, isso é, simplesmente testando-se diferentes valores até obter o tipo de descrição desejada.

B. Seleção de Propriedades Customizável

A segunda proposta de extensão da abordagem Incremental trata de aspectos qualitativos da estratégia de referência, partindo da observação de que diferentes contextos de um mesmo domínio podem favorecer diferentes tipos de propriedades. Considere-se por exemplo a cena da Fig. 4 anterior, relativa ao domínio b5-ref. Neste contexto, o atributo gênero é altamente saliente em virtude de seu valor único no contexto (i.e., só há uma moça na cena) e, conseqüentemente, tende a ser selecionado com mais frequência. No entanto, observamos que o domínio b5-ref possui diversas outras cenas deste tipo, porém apresentando número variado de pessoas de ambos os sexos. Assim, o uso da informação de gênero pode ser maior ou menor em diferentes situações de referência e, em última análise, em uma cena em que todas as pessoas são do mesmo sexo a utilidade prática de uma referência ao gênero do indivíduo torna-se extremamente reduzida.

Considerando que o algoritmo Incremental contempla o uso de uma lista de preferências fixa para cada domínio, uma extensão natural desta abordagem consiste em utilizar listas de propriedades preferenciais individualizadas para cada contexto do domínio, conforme discutido a seguir.

C. O Algoritmo IAL

Tomadas em conjunto, as duas extensões propostas nas seções anteriores – o controle quantitativo do grau de superespecificação, e o uso (qualitativo) de preferências específicas para cada contexto – são a base de um algoritmo IAL, ilustrado na Fig. 5 a seguir. Neste código, entretanto, observamos que o uso de listas P individualizadas por contexto é apenas pressuposto, porém não representado de forma explícita por questões de brevidade.

```

1 MakeReferringExpression( $r, C, P, level$ )
2    $L \leftarrow \{\}$ 
3    $C \leftarrow C - \{r\}$ 
4   for  $A_i \in P$  do
5      $V \leftarrow value(r, A_i)$ 
6     if  $|rulesOut(\langle A_i, V \rangle)| > 0$  then
7        $L \leftarrow L \cup \langle A_i, V \rangle$ 
8        $C \leftarrow C - rulesOut(\langle A_i, V \rangle)$ 
9     end
10    if  $|L| == level$  then
11      return  $L$ 
12    end
13  end
14  if  $|L| < level$  then
15    for  $A_i \in P$  do
16       $V \leftarrow value(r, A_i)$ 
17       $L \leftarrow L \cup \langle A_i, V \rangle$ 
18      if  $|L| == level$  then
19        return  $L$ 
20      end
21    end
22  end
23  return  $L$ 

```

Fig. 5. O algoritmo Incremental de nível customizável (IAL).

Assim como o algoritmo Incremental original, IAL recebe como entrada (linha 1) o objeto-alvo r a descrever, o conjunto de objetos do contexto de referência C , a lista de propriedades preferenciais P (neste caso específica para cada contexto) e, adicionalmente, o parâmetro $level$ previamente estimado para aquele contexto conforme discutido. A saída do algoritmo é a lista de propriedades L (inicialmente vazia, cf. linha 2) que forma a descrição de r .

O algoritmo é dividido em duas etapas principais. A primeira etapa (linhas 2-13) objetiva a construção de uma descrição básica, e possivelmente incompleta de r . Este procedimento é semelhante ao adotado pelo algoritmo Incremental, exceto pela condição de parada (linha 10). No algoritmo Incremental, propriedades eram incluídas em L até a descrição se tornar livre de ambigüidade (ou seja, até o conjunto de distraidores C se tornar vazio). No caso de IAL, pelo contrário, a seleção se encerra quando a quantidade de propriedades em L atinge o valor $level$ estipulado.

Uma vez concluída esta primeira etapa, segue-se um procedimento de superespecificação (linhas 14-22) que objetiva expandir (se necessário) a descrição L até atingir o tamanho médio $level$ esperado para o contexto em questão. Uma vez atingido este limite, ou uma vez esgotadas as propriedades possíveis a considerar, a descrição L é retornada e o algoritmo encerra.

A possível superioridade de uma abordagem customizável como IAL sobre um algoritmo orientado à produção de descrições breves como o Incremental [3] é facilmente observada em situações de referência como a do contexto ilustrado na Fig. 4 anterior. Considere-se para este fim que o objeto-alvo destacado em vermelho (a única mulher da cena) possuía, simplificada, apenas as propriedades *gender-female*, *smile-no*, *hairColour-blonde* e *ponyTail-yes*, e considere-se uma ordem preferencial para este contexto na forma $P = \{gender, smile, hairColour, ponyTail\}$. Neste caso, observamos comportamentos distintos por parte dos dois algoritmos como segue:

A abordagem Incremental sempre produziria a mesma descrição $L = \{gender-female\}$, que poderia ser realizado como, e.g., ‘a mulher’, ‘a moça’ etc., e que seria o suficiente para eliminar todos os distraidores deste contexto (que são homens).

A abordagem IAL, por outro lado, exploraria o fato de que o tamanho médio das descrições deste alvo, tal qual observado na porção de treinamento do cópulus, é de $level = 3$ propriedades, ou seja, levaria em conta que locutores humanos na verdade são altamente redundantes [9] e que, nesta cena específica, tendem a acrescentar em média duas propriedades além do necessário para a desambiguação. Assim, IAL produziria $L = \{gender-female, smile-no, hairColour-blonde\}$, que poderia ser realizado como, e.g., ‘a mulher loira que não está sorrindo’, ‘a loira emburrada’ etc. Estas descrições, embora sendo mais longas e redundantes do que seria estritamente necessário para identificar o objeto-alvo nesta cena, são em média mais próximas das descrições observadas nos exemplos reais do cópulus, o que em princípio permite que a abordagem customizada IAL seja (semanticamente) mais próxima do desempenho humano do que a abordagem Incremental.

IV. AVALIAÇÃO

O algoritmo IAL proposto na seção anterior foi testado com base em três domínios de GER, e seus resultados foram comparados com o algoritmo Incremental original em [3], aqui denominado Incr. Detalhes desta avaliação são discutidos a seguir.

A. Conjuntos de Dados

Para avaliação da proposta, foram utilizadas as descrições disponibilizadas pelos corpú TUNA *Furniture* e *People* [13] e b5-ref [4] conforme discutido na seção II-C. No caso do domínio TUNA, apenas os dados correspondente a descrições no singular foram considerados, uma vez que a geração de descrições de conjuntos constituiu um problema computacional bastante distinto do problema básico de GER. Em especial, observa-se que o cômputo de descrições coletivas como ‘as cadeiras’ em referência aos dois objetos deste tipo na Fig. 3 requer, no caso geral, o uso de formalismos lógicos específicos que fogem ao propósito fundamental de algoritmos como a abordagem Incremental em [3]. Uma proposta que trata especificamente deste tipo de problema é discutida em [26]. A Tabela II ilustra a quantidade de descrições, o número de sujeitos participantes e o número de contextos (cenas) em cada domínio.

TABELA II
CONJUNTOS DE DADOS CONSIDERADOS

Domínio	Descrições	Sujeitos	Contextos
TUNA-Furniture	420	60	7
TUNA-People	360	60	6
b5-ref	1900	154	12

Cada conjunto de dados foi dividido em porções de treinamento (80%) e teste (20%) de modo pseudoaleatório, objetivando garantir que os conjuntos de treinamento e teste incluíssem ao menos um exemplo de descrição de cada cena e de cada participante.

B. Procedimento

A porção de treinamento de cada corpú foi utilizada para estimar os parâmetros dos algoritmos IAL e Incr da seguinte forma. No caso de IAL, foram computadas listas de preferências P individualizadas e um parâmetro *level* para cada contexto. Para Incr, foi computada uma lista de preferências fixa por domínio. Em todos os casos, as listas de preferências são ordenadas pela frequência dos atributos observados em cada conjunto de treinamento, começando pelo atributo mais frequente.

Cada uma das situações de referência dos conjuntos de testes (formada por um objeto-alvo, distraidores e suas respectivas propriedades) foi submetida a ambos algoritmos com os devidos parâmetros de entrada, e a descrição resultante em cada caso foi comparada com a descrição original (humana) disponibilizada nos dados de teste. Seguindo a prática da área [7], [8], esta comparação foi realizada medindo-se o coeficiente Dice [27], que representa o grau de sobreposição entre dois conjuntos (no caso, entre o conjunto de propriedades

gerada e o conjunto original constante do corpú). Coeficientes Dice variam entre 0 e 1, em que o valor 0 indica total dissemelhança entre a descrição do sistema e a versão humana, e o valor 1 indica total coincidência.

V. RESULTADOS

A Tabela III sumariza os coeficientes Dice médios obtidos pelo baseline Incremental (Incr) e pelo modelo proposto (IAL) para cada conjunto de teste. Além disso, são apresentados também resultados do teste não paramétrico de Wilcoxon [28] comparando os coeficientes obtidos pelas duas alternativas. Os melhores resultados para cada domínio são destacados em negrito.

TABELA III
COEFICIENTES DICE MÉDIOS

Domínio	Incr	IAL	W	z	p
TUNA-Furniture	0,82	0,95	-15378	-11,36	< 0,0001
TUNA-People	0,62	0,94	-55725	-15,71	< 0,0001
b5-ref	0,59	0,67	78	3,04	0,0012

Os resultados da Tabela III sugerem que a abordagem customizada IAL apresenta resultados superiores à abordagem Incremental original. Esta diferença é estatisticamente significativa para os três domínios de teste considerados.

VI. CONSIDERAÇÕES FINAIS

Este artigo enfocou o problema de geração automática de texto, tratando especificamente da questão da seleção de conteúdo para geração de expressões de referência. O trabalho propôs uma extensão de um algoritmo padrão da área, em que aspectos qualitativos e quantitativos da estratégia de referência são customizáveis a partir de um pequeno volume de dados de treinamento. A solução proposta foi testada em diferentes corpú para GER, e seus resultados são comparados favoravelmente à abordagem tradicional.

O presente estudo abre espaço para uma série de oportunidades de melhoria e futuras extensões. Em especial, destacamos que o algoritmo proposto, assim como a abordagem Incremental original, lida apenas com propriedades atômicas, isto é, não considera a geração de descrições relacionais como em ‘o homem ao lado da moça loira’. Consequentemente, a presente proposta foi avaliada apenas com base nos conjuntos de teste de descrições atômicas TUNA e b5-ref. Uma extensão relacional do algoritmo IAL proposto, e os respectivos testes com base em conjuntos de dados deste tipo, é deixada como trabalho futuro.

Além disso, destacamos também que a presente discussão concentrou-se exclusivamente no problema da seleção de conteúdo semântico, sem considerar a etapa subsequente em uma arquitetura de GLN, ou seja, a etapa de realização superficial, que trataria da representação do conteúdo selecionado em uma língua-alvo específica. Um estudo desta natureza também é deixado como trabalho futuro.

AGRADECIMENTOS

Este trabalho contou com apoio FAPESP nro. 2016/14223-0 e da Universidade de São Paulo.

REFERÊNCIAS

- [1] E. Reiter and R. Dale, *Building natural language generation systems*. New York, NY, USA: Cambridge University Press, 2000.
- [2] K. van Deemter, *Computational Models of Referring. A Study in Cognitive Science*. Cambridge, Massachusetts, USA: MIT Press, 2016.
- [3] R. Dale and E. Reiter, "Computational interpretations of the Gricean maxims in the generation of referring expressions," *Cognitive Science*, vol. 19, 1995.
- [4] I. Paraboni, D. S. Monteiro, and A. G. J. Lan, "Personality-dependent referring expression generation," in *Text, Speech and Dialogue (TSD-2017) Lecture Notes in Artificial Intelligence vol. 10415*. Prague, Czech Republic: Springer-Verlag, 2017, pp. 20–28.
- [5] R. Dale, "Cooking up referring expressions," in *Proceedings of ACL-2002*. Philadelphia, PA, USA: Association for Computational Linguistics, 2002, pp. 68–75.
- [6] E. Krahmer, S. van Erk, and A. Verleg, "Graph-based generation of referring expressions," *Computational Linguistics*, vol. 29, no. 1, pp. 53–72, 2003.
- [7] E. Krahmer and K. van Deemter, "Computational generation of referring expressions: A survey," *Computational Linguistics*, vol. 38, no. 1, pp. 173–218, 2012.
- [8] A. Gatt, A. Belz, and E. Kow, "The TUNA challenge 2008: Overview and evaluation results," in *Fifth International Natural Language Generation Conference (INLG-2008)*. Salt Fork, Ohio, USA: The Association for Computational Linguistics, 2008, pp. 198–206.
- [9] T. Pechmann, "Incremental speech production and referential overspecification," *Linguistics*, vol. 27, no. 1, pp. 98–110, 1989.
- [10] I. Paraboni, "An algorithm for generating document-deictic references," in *Proc. of workshop Coherence in Generated Multimedia, associated with First Int. Conf. on Natural Language Generation (INLG-2000), Mitzpe Ramon*, 2000, pp. 27–31.
- [11] S. Gupta and A. Stent, "Automatic evaluation of referring expression generation using corpora," in *1st workshop on using corpora in NLG*, Birmingham, UK, 2005.
- [12] B. Bohnet, "The fingerprint of human referring expressions and their surface realization with graph transducers," in *5th International Natural Language Generation Conference*. Salt Fork, Ohio, USA: Association for Computational Linguistics, 2008, pp. 207–210.
- [13] A. Gatt, I. van der Sluis, and K. van Deemter, "Evaluating algorithms for the generation of referring expressions using a balanced corpus," in *Proceedings of ENLG-07*. Schloss Dagstuhl, Germany: Association for Computational Linguistics, 2007.
- [14] R. Koolen, A. Gatt, M. Goudbeek, and E. Krahmer, "Factors causing overspecification in definite descriptions," *Journal of Pragmatics*, vol. 43, pp. 3231–3250, 2011.
- [15] B. B. C. Silva and I. Paraboni, "Learning personality traits from Facebook text," *IEEE Latin America Transactions*, vol. 16, no. 4, pp. 1256–1262, 2018.
- [16] G. Righi, J. J. Peissig, and M. J. Tarr, "Recognizing disguised faces," *Visual Cognition*, vol. 20, no. 2, pp. 143–169, 2012.
- [17] R. Dale and J. Viethen, "Referring expression generation through attribute-based heuristics," in *12th European Workshop on Natural Language Generation*, ser. ENLG '09. Athens, Greece: Association for Computational Linguistics, 2009, pp. 58–65.
- [18] J. Viethen and R. Dale, "GRE3D7: A corpus of distinguishing descriptions for objects in visual scenes," in *UCNLG+Eval: Language Generation and Evaluation Workshop*. Edinburgh, UK: Association for Computational Linguistics, July 2011, pp. 12–22.
- [19] C. V. M. Teixeira, I. Paraboni, A. S. R. da Silva, and A. K. Yamasaki, "Generating relational descriptions involving mutual disambiguation," in *Computational Linguistics and Intelligent Text Processing (CICLing-2014), Lecture Notes in Computer Science 8403*. Kathmandu, Nepal: Springer, 2014, pp. 492–502.
- [20] I. Paraboni, M. Galindo, and D. Iacovelli, "Stars2: a corpus of object descriptions in a visual domain," *Language Resources and Evaluation*, vol. 51, no. 2, pp. 439–462, 2017.
- [21] M. Mitchell, K. van Deemter, and E. Reiter, "Natural reference to objects in a visual domain," in *Proceedings of INLG-2010*. Dublin, Ireland: The Association for Computational Linguistics, 2010.
- [22] S. Kazemzadeh, V. Ordonez, M. Matten, and T. Berg, "ReferItGame: Referring to objects in photographs of natural scenes," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, 2014, pp. 787–798.
- [23] A. Gargett, K. Garoufi, A. Koller, and K. Striegnitz, "The GIVE-2 corpus of giving instructions in virtual environments," in *Proceedings of LREC-2010*. Valletta, Malta: ELRA, 2010.
- [24] D. dos Santos Silva and I. Paraboni, "Generating spatial referring expressions in interactive 3D worlds," *Spatial Cognition & Computation*, vol. 15, no. 03, pp. 186–225, 2015.
- [25] R. Koolen, E. Krahmer, and M. Theune, "Learning preferences for referring expression generation: Effects of domain, language and algorithm," in *Proceedings of the Seventh International Natural Language Generation Conference*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2012, pp. 3–11.
- [26] M. Stone, "On identifying sets," in *Proceedings of the First International Conference on Natural Language Generation - Volume 14*, ser. INLG '00. Stroudsburg, PA, USA: Association for Computational Linguistics, 2000, pp. 116–123.
- [27] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.
- [28] F. Wilcoxon, "Individual comparisons by ranking methods."



André C. Mariotti é graduado em Sistemas de Informação pela Universidade de São Paulo (EACH-USP) e mestre em ciências pela mesma unidade. Atualmente trabalha como engenheiro de machine learning no Santander, responsável pelo desenvolvimento de modelos que atendem as áreas de negócio do banco.



Ivandré Paraboni é doutor em ciências da computação pela University of Brighton (Inglaterra, 2003) e realizou estágio de pós-doutorado junto à University of Aberdeen (Escócia, 2012). É professor associado em regime de dedicação integral à docência e pesquisa junto à Universidade de São Paulo (EACH-USP).