



UNIVERSIDADE DE SÃO PAULO

Escola de Artes, Ciências e Humanidades

Relatório Técnico PPgSI-002/2016

Efeitos da superespecificação referencial sobre a identificação de objetos em domínios visuais

Alex Gwo Jen Lan, Matheus Mendes de Sant'Ana, Ivandré Paraboni

Março - 2016

O conteúdo do presente relatório é de única responsabilidade dos autores.

Série de Relatórios Técnicos

PPgSI-EACH-USP

Rua Arlindo Bétio, 1000 – Ermelino Matarazzo

03828-000 – São Paulo, SP.

TEL: (11) 3091-8197

<http://www.each.usp.br/ppgsi>

Efeitos da superespecificação referencial sobre a identificação de objetos em domínios visuais

Alex Gwo Jen Lan, Matheus Mendes de Sant'Ana, Ivandré Paraboni¹

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo
São Paulo – SP, Brazil

{alex.lan,matheus.ana,ivandre}@usp.br

Resumo. *Este relatório apresenta um estudo do impacto da superespecificação de expressões de referência sobre o tempo de inspeção de objetos distraidores em domínio visuais. Para este fim, foi realizado um experimento envolvendo participantes humanos engajados em uma tarefa de identificação de objetos com base em descrições mínimas ou superespecificadas com uso de uma propriedade de maior ou menor dificuldade de reconhecimento. Resultados sugerem que, dependendo do tipo de propriedade considerada, a superespecificação pode tanto facilitar como dificultar a interpretação da expressão de referência.*

1. Introdução

Sistemas de geração de língua natural (GLN) tratam da produção de descrições textuais a partir de uma entrada de dados geralmente não linguística [Reiter e Dale 2000]. Dentre as diversas subtarefas desempenhadas por um sistema de GLN está a geração de expressões de referência (GER) [Krahmer e van Deemter 2012]. GER consiste em produzir descrições linguísticas de objetos de que trata o discurso, e pode de certa forma ser vista como a tarefa computacional ‘simétrica’ à interpretação referencial [Paraboni 1997, Paraboni e de Lima 1998, Cuevas e Paraboni 2008].

GER é uma ativa linha de pesquisa em GLN, tendo sido inclusive objeto de duas séries de competições (ou *shared tasks*) recentes: a série de desafios TUNA [Gatt et al. 2009] e GIVE [Byron et al. 2007], que pode ser estudada inclusive a partir de corpus dedicados [Eugenio et al. 2000, Gorniak e Roy 2004, Viethen e Dale 2006, Gatt et al. 2007, Guhe 2009, Dale e Viethen 2009, Mitchell et al. 2010, Gargett et al. 2010, Viethen e Dale 2011, Clarke et al. 2013, FitzGerald et al. 2013, Kazemzadeh et al. 2014, Teixeira et al. 2014, Altamirano et al. 2015, Paraboni et al. 2015, 2016].

A tarefa de GER possui reflexos nas três camadas da arquitetura de GLN tradicional, a saber: macroplanejamento [Paraboni e van Deemter 1999, 2002b], microplanejamento [Krahmer e van Deemter 2012] e realização superficial [Pereira e Paraboni 2007, 2008, de Novais e Paraboni 2012]. Neste trabalho enfocamos apenas a questão da seleção de conteúdo implementada como uma das tarefas de microplanejamento.

A seleção de conteúdo em GER conta com uma ampla gama de algoritmos de propósito geral [Dale e Haddock 1991, Dale e Reiter 1995, Paraboni 2000, Paraboni e van Deemter 2002a, Krahmer e Theune 2002, Krahmer et al. 2003, de Lucena et al. 2010, Ferreira e Paraboni 2014b,a]. Dependendo do objetivo da referência, podemos dizer que a expressão gerada precisa ser psicologicamente plausível, embora muitos outros tipos de objetivos tenham sido discutidos na literatura [Krahmer e van Deemter 2012].

Algoritmos de GER tipicamente seguem critérios de preferência por determinados tipos de propriedades [Pechmann 1989], mas estudos mais recentes demonstram que a tarefa é na realidade mais complexa, e estas preferências podem ser efetivamente redefinidas [van Deemter et al. 2012, Tarenskeen et al. 2014, van Gompel et al. 2014]. De modo geral, estas expressões devem fazer referência única a um ou mais objetos-alvo do contexto em questão, e é normalmente desejável que sejam breves [Dale 2002, Gardent 2002, Areces et al. 2011] e consistentes com a máxima de brevidade [Grice 1975]. Assim, apenas propriedades ditas discriminatórias [Olson 1970] costumam ser consideradas, embora esta visão também tenha sido contestada face à tendência humana de superespecificar [Engelhardt e Ferreira 2006, Koolen et al. 2011, 2013, van Gompel et al. 2012].

Diante desta diversidade de critérios, um dos problemas fundamentais do projeto de algoritmos de seleção de conteúdo é a questão de como avaliá-los. Dentre diversas possibilidades de avaliação intrínseca [Gatt et al. 2009, dos Santos Silva e Paraboni 2015] e extrínseca [Gargett et al. 2010] com uso de métricas de seleção de conteúdo [Dice 1945, Passonneau 2006] ou realização superficial [Papineni et al. 2002, NIST 2002], uma técnica em especial tem sido alvo de atenção da comunidade científica em anos recentes: o uso de técnicas de rastreamento do olhar (RT) [Duchowski 2002, Morimoto e Mimica 2005, Hansen e Ji 2010]. Com uso de técnicas de RT torna-se possível, por exemplo, medir diretamente o impacto de uma instrução linguística sobre o foco de atenção do ouvinte, determinando assim se um algoritmo produz resultados que exigem maior ou menor tempo de interpretação para reconhecimento de certas propriedades.

De especial interesse para o presente trabalho, observamos que não há consenso na literatura sobre o impacto da superespecificação referencial na identificação de um objeto-alvo. Por um lado, diversos estudos em GER e áreas correlatas defende a ideia de que o acréscimo de informações além do estritamente necessário para identificação é um facilitador [Arts et al. 2011, Paraboni 2003]. Por outro lado, estudos como Engelhardt et al. [2011] sugerem que descrições superespecificadas podem na verdade dificultar a identificação do objeto-alvo. O experimento descrito nas próximas seções é uma tentativa de conciliar estas duas visões aparentemente conflitantes com o uso de técnicas de RT para medir o tempo de identificação de objetos a partir de diferentes tipos de expressões mínimas e superespecificadas.

2. Experimento

O experimento realizado objetivou investigar se a superespecificação de expressões de referência com uso de propriedades que exigem menos ou mais esforço de reconhecimento afeta o tempo de identificação de um objeto. Para este fim, será considerado um domínio visual simples contendo objetos do tipo estrela de 7 ou 8 pontas, e contendo rótulos alfanuméricos. A Fig. 1 exemplifica uma cena deste tipo.

O presente experimento parte do pressuposto de que decidir se uma estrela possui 7 ou 8 pontas é *mais difícil* do que decidir se o rótulo representa uma letra ou número. Mais especificamente, partimos do pressuposto de que um sujeito humano precisaria de *mais tempo* examinando o objeto para determinar seu formato do que para determinar o seu tipo de rótulo. Embora esta seja uma suposição razoável neste contexto, a confirmação deste pressuposto será tratada juntamente com a análise de dados descrita na seção 3.

O experimento contempla quatro condições que fazem uso de expressões mínimas

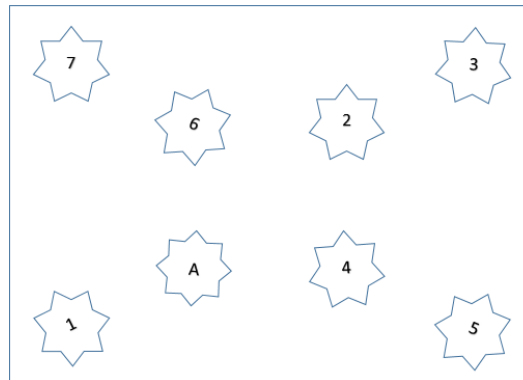


Figura 1. Um domínio visual simples representando estrelas de 7/8 pontas contendo letras ou números.

(aqui denominadas *min*) ou superespecificadas (*overspec*). Estas expressões fazem uso de propriedades do tipo H (de ‘hard’, mais difíceis de reconhecer, representando o número de pontas da estrela) e E (de ‘easy’, mais fáceis de reconhecer, representando o tipo de rótulo). Assim, as condições são denominadas *min.H*, *overspec.HE*, *min.E* e *overspec.EH* conforme a composição de cada expressão de referência. Por exemplo, uma expressão como ‘a estrela de 7 pontas que contém uma letra’ representa a condição difícil-fácil *overspec.HE*, enquanto que ‘a letra na estrela de 7 pontas’ representa a condição fácil-difícil *overspec.EH*.

A Fig. 2 exemplifica estas condições por meio de quatro imagens de estímulo utilizadas no presente experimento. Nesta representação, o objeto-alvo foi destacado de modo a facilitar a compreensão dos exemplos, mas este destaque não ocorre nas imagens reais do experimento.

As condições *min* (no lado esquerdo da figura) fazem uso de uma descrição mínima composta pela propriedade dita mais ‘difícil’ (i.e., o número de pontas da estrela, que pode ser 7 ou 8), ou pela propriedade mais ‘fácil’ (o tipo do rótulo, que pode ser letra ou número). Em ambos os casos, esta propriedade é suficiente para descrever o objeto-alvo de forma única.

As condições *overspec* (lado direito da figura) são idênticas às condições *min*, exceto pela expressão de referência. Nestes casos, foi acrescentada à descrição original a segunda propriedade possível (fácil ou difícil, conforme o caso) que não é necessária para desambiguação, embora também seja suficiente para descrever o objeto-alvo de forma única.

As quatro condições foram testadas com uso de um experimento envolvendo sujeitos humanos em uma tarefa de identificação de objetos em cenas visuais a partir de instruções contendo uma descrição definida não ambígua. Nesta tarefa, os participantes eram orientados a ler a instrução fornecida e identificar o objeto na cena enquanto um dispositivo de rastreamento de olhar media o tempo gasto examinando cada objeto da tela, de modo a estabelecer o impacto de estratégias variadas de superespecificação na resolução destas expressões.

Uma vez que no presente experimento a identificação do objeto-alvo inclui a decisão consciente do sujeito de concluir a tarefa ou não (e se for o caso pressionar espaço para

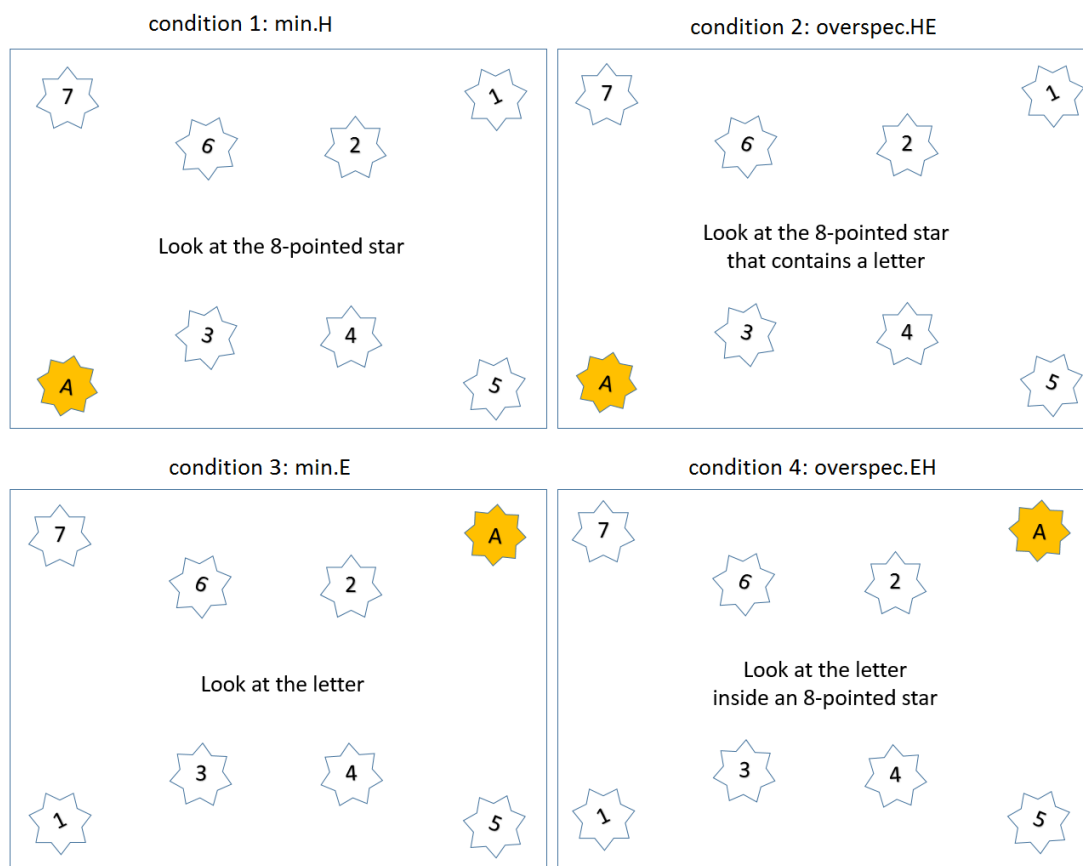


Figura 2. Condições críticas do experimento, envolvendo descrições mínimas (esquerda) e superespecificadas (direita).

prosseguir, conforme será discutido na seção 2.2), o tempo de reconhecimento do objeto-alvo pode ser pouco representativo do real esforço de identificação. Por este motivo, no presente experimento será considerado o tempo médio de inspeção dos objetos distraídos em cada uma das cenas de estímulo.

A superespecificação de uma expressão de referência é frequentemente associada à ideia de facilitar a busca ou a identificação do referente [Paraboni et al. 2006, Arts et al. 2011], embora em alguns casos tenha sido observado o efeito oposto [Engelhardt et al. 2011]. Neste trabalho tentaremos conciliar estas duas visões aparentemente contraditórias defendendo a ideia de que a superespecificação pode tanto facilitar como prejudicar a identificação de um objeto *dependendo do tipo de propriedade que for acrescentada* testando duas hipóteses de pesquisa $h1$ e $h2$ conforme descrito a seguir.

A primeira hipótese trata do caso de superespecificação mediante uso de uma propriedade de fácil reconhecimento:

h1: O acréscimo de uma propriedade de fácil reconhecimento a uma expressão de referência mínima reduz o tempo médio de inspeção de objetos distraídos.

A hipótese $h1$ prevê que a superespecificação referencial facilita a identificação do objeto tal qual proposto em estudos como Arts et al. [2011], Paraboni e van Deemter [2014]. Esta hipótese será testada comparando-se o tempo médio gasto inspecionando objetos distraídos em situações em que é fornecida uma descrição mínima contendo

apenas a propriedade mais difícil (condição *min.H*) ou superespecificada por uma propriedade fácil (*overspec.HE*). Espera-se que este tempo seja menor na condição *overspec.HE* do que na condição *min.H*.

A segunda hipótese trata do caso de superespecificação mediante uso de uma propriedade de difícil reconhecimento:

h2: O acréscimo de uma propriedade de difícil reconhecimento a uma expressão de referência mínima aumenta o tempo médio de inspeção de objetos distraidores.

A hipótese *h2* estabelece que a superespecificação referencial pode prejudicar a identificação do objeto, e é aparentemente conflitante com *h1*. Efeitos deste tipo foram entretanto observados em estudos como Engelhardt et al. [2011], e serão retomados na seção 4. Esta hipótese será testada comparando-se o tempo médio gasto examinando objetos distraidores em situações em que é fornecida uma descrição mínima contendo apenas a propriedade mais fácil (condição *min.E*) ou superespecificada por uma propriedade dita difícil (*overspec.EH*). Espera-se que este tempo seja maior na condição *overspec.EH* do que na condição *min.E*.

2.1. Sujeitos

10 estudantes de Sistemas de Informação da USP-EACH, que responderam a um convite pessoal e concordaram em participar de forma voluntária. Os participantes tinham em média 20,6 anos de idade (mínimo=19 e máximo=23) e eram predominantemente do sexo masculino (8, ou 80%). Todos participantes eram brasileiros nativos.

2.2. Procedimento

O experimento consistiu em exibir cada um dos estímulos ao participante em ordem aleatória, e rastrear seu olhar do momento em que a expressão de referência era exibida no centro da tela, até a fixação sobre o objeto-alvo em questão. Ao pressionar a tecla de espaço, o participante indicava a conclusão da identificação do objeto, e o estímulo seguinte era exibido. Nenhuma menção era feita ao papel dos objetos distraidores (que são o verdadeiro interesse do presente experimento, conforme discutido na seção 2), levando os participantes a crer que o foco do experimento era realmente encontrar o objeto-alvo.

A participação no experimento foi realizada de forma individual em horário pré-agendado. Após a calibração do dispositivo de rastreamento de olhar, os participantes recebiam instruções sobre como proceder e tinham algum tempo para se familiarizar com a tarefa utilizando algumas imagens práticas semelhantes aos estímulos reais do experimento. Após a sessão prática, seguia-se o experimento propriamente dito.

Os dados coletados - inicialmente na forma de um fluxo de coordenadas da tela - foram processados para extrair a sequência de objetos examinados pelo participante e o tempo de inspeção de cada um, começando pelo primeiro objeto visualizado na tela, e terminando no momento em que o objeto-alvo é visualizado pela primeira vez. Assim, quando o sujeito continua inspecionando elementos na tela depois de ter visualizado o alvo, o restante da sequência de busca é desprezado. Esta medida foi necessária porque alguns participantes optaram por verificar objetos adicionais antes de confirmar a seleção do alvo já identificado, o que causava distorção no tempo medido.

A exceção à regra de truncamento da sequência no momento em que o alvo era encontrado é o caso em que o alvo foi o primeiro objeto visualizado, mas mesmo assim o participante optou por fazer uma inspeção dos demais elementos da tela antes de retornar ao alvo inicial. Nestes casos optamos por desconsiderar a visualização inicial do alvo e tratar o restante da sequência de busca normalmente. Esta medida foi tomada para evitar que os dados se tornassem excessivamente esparsos, já que na descoberta do alvo no início da sequência não haveria passagem por nenhum distraidor. Nos casos em que um objeto é inspecionado múltiplas vezes, o tempo de reconhecimento em questão é definido como a soma de todas as visitas individuais, mas sempre limitado ao conjunto de distraidores visitados até a primeira ocorrência do alvo.

2.3. Materiais

O experimento fez uso de software desenvolvido especificamente para este fim e imagens de estímulo. O aplicativo exibia as instruções gerais, as imagens de estímulo em ordem aleatória e a instrução no centro da tela, fazia o monitoramento do olhar do participante e armazenava a sequência de objetos visualizados e seus respectivos tempos.

As imagens de estímulo eram semelhantes às ilustradas na Fig. 2 anterior, contendo oito estrelas cada. Um experimento com estas características poderia em princípio ser realizado com o uso de imagens de objetos individuais, e não necessariamente levando em conta cenas contendo múltiplos objetos como nos exemplos anteriores. Esta opção em nosso experimento está entretanto ligada ao futuro reuso dos dados coletados para investigação de questões ligadas à trajetória de busca pelo objeto-alvo, a qual não será discutida no presente trabalho.

Nas condições críticas do experimento o objeto-alvo sempre se localiza em um dos quatro cantos mais distantes do centro da tela de modo a aumentar as chances de que ao menos um distraidor seja inspecionado antes da identificação. Imagens do tipo *filler* foram utilizados para contemplar cenas com objetos-alvo nas demais posições possíveis. O ângulo de inclinação dos objetos foi definido de forma aleatória com o objetivo de aumentar a dificuldade de identificação do formato do objeto, mas não possui nenhum significado especial no experimento.

Também como forma de aumentar as chances de que algum distraidor seja inspecionado antes de encontrar o alvo, cada uma das quatro condições críticas foi apresentada duas vezes, alternando o uso de letras/números e formatos de 7/8 pontas para maior variedade, totalizando assim oito imagens críticas para o experimento. A estas oito imagens foram acrescentados 16 *fillers*, totalizando assim 24 imagens de estímulo.

3. Resultados

A Tabela 1 apresenta o tempo médio de inspeção de objetos distraidores nas condições críticas do experimento.

Os itens marcados como 'na' indicam casos em que o sujeito não chegou a examinar nenhum distraidor naquela condição (por exemplo, tendo encontrado o alvo de imediato) ou não encontrou o objeto-alvo esperado (caso em que talvez não estivesse tratando a tarefa com a devida atenção).

Na análise a seguir, utilizamos testes do tipo ANOVA de fator único sobre o tempo médio de inspeção de objetos distraidores. Antes de discutir o teste de hipóteses própria-

Tabela 1. Tempo médio de inspeção de objetos distraidores

sub.	h1				h2			
	min.H		overspec.HE		min.E		overspec.EH	
	mean	sd	mean	sd	mean	sd	mean	sd
1	316.1	155.8	305.3	64.7	178.5	36.6	277.2	63.9
2	363.3	90.9	300.1	37.4	177.0	41.5	633.2	670.1
3	262.4	116.8	190.8	66.8	203.4	38.7	210.2	27.8
4	334.0	75.5	199.5	48.2	166.5	21.4	191.0	na
5	409.7	86.5	403.7	126.9	177.3	na	400.2	25.7
6	390.5	7.8	193.0	123.0	351.5	60.5	358.0	na
7	482.3	88.7	192.3	37.2	319.1	44.6	461.8	264.7
8	199.3	23.7	167.4	2.6	319.8	97.3	378.8	177.1
9	321.0	121.2	323.6	12.1	454.5	10.6	489.8	60.5
10	454.5	160.5	265.9	102.0	230.3	20.8	341.9	5.6
	353.3	85.8	254.1	77.6	257.8	98.0	374.2	133.3

mente dito, entretanto, trataremos de verificar os pressupostos básicos do experimento no que diz respeito ao tempo de reconhecimento de propriedades ditas fáceis e difíceis, e da eventual diferença entre reconhecer letras x números ou objetos de 7 x 8 pontas.

Conforme discutido na seção 2, consideramos que o tempo de reconhecimento da propriedade formato da estrela (7 x 8 pontas) requer mais tempo do que o reconhecimento do seu tipo de rótulo (letra x número). Para verificar a validade desta premissa, comparamos o tempo médio de inspeção de distraidores nas condições *min.E* e *min.H*. Conforme ilustrado na Tab. 1, o tempo médio de inspeção de objetos com base no formato (*min.H*) é maior do que o tempo médio de inspeção destes mesmos objetos com base no tipo de rótulo (*min.E*). A diferença é significativa ($F(1,18)=5.37$, $MSE=8486.81$ $p=0.032399$). Reconhecer o número de pontas (7/8) requer mais tempo do que reconhecer o tipo de rótulo (letra/número). Isso confirma a premissa básica do experimento.

Realizamos também comparações individualizadas dos tempos de reconhecimento de letras x números, e dos tempos de reconhecimento de formatos de 7 x 8 pontas. Em ambos os casos, não houve diferença significativa entre o uso de uma propriedade e outra. O tempo gasto para reconhecer uma letra não difere significativamente do tempo gasto para reconhecer um número, e o tempo gasto para reconhecer um objeto de 7 pontas não difere significativamente do reconhecimento de um objeto de 8 pontas.

Uma vez confirmadas estas premissas, passamos ao teste de hipóteses propriamente dito. Com relação à hipótese *h1*, observou-se que a adição de uma propriedade de menor esforço (*overspec.HE*) reduz o tempo médio de inspeção dos objetos distraidores em relação à expressão mínima original (*min.H*). Esta diferença é significativa ($F(1,18)=7.35$, $MSE=6692.43$ $p=0.014326$). Este resultado oferece suporte à hipótese *h1*.

Com relação à hipótese *h2*, observou-se o efeito oposto, ou seja: a adição de uma propriedade de maior esforço (*overspec.EH*) aumenta o tempo médio de inspeção dos objetos distraidores em relação à expressão mínima original (*min.E*). Esta diferença também é significativa ($F(1,18)=4.95$ $MSE=13689.99$, $p=0.039134$). Este resultado oferece suporte à hipótese *h2*.

4. Considerações

Este relatório descreveu um experimento sobre o uso de expressões de referência superespecificadas em domínio visuais. Os resultados sugerem que (a) o acréscimo de propriedades ditas fáceis auxilia a identificação do objeto, sendo portanto consistente com estudos como Arts et al. [2011] e, por outro lado, que (b) o acréscimo de propriedades ditas difíceis prejudica a identificação, o que é consistente com estudos como Engelhardt et al. [2011]. Embora nenhuma destas duas conclusões (a-b) seja individualmente inédita, acreditamos que o presente experimento tenha demonstrado, possivelmente pela primeira vez, a ligação entre estas duas observações utilizando evidência extraída a partir do rastreamento de olhar.

Como trabalho futuro, esperamos incorporar estes resultados em um algoritmo de GER e ampliar o escopo desta investigação analisando as consequências práticas para modelos computacionais deste tipo.

Agradecimentos

Ao Programa de Educação Tutorial do Ministério da Educação e à FAPESP pelo apoio financeiro, ao Prof. Carlos Hitoshi Morimoto (IME-USP) pelo equipamento utilizado, e ao prof. Flávio Coutinho (EACH-USP) pelo apoio técnico. Os autores são também gratos aos voluntários que participaram deste experimento.

Referências

- Altamirano, R., Ferreira, T. C., Paraboni, I., e Benotti, L. (2015). Zoom: a corpus of natural language descriptions of map locations. Em *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics ACL 2015*, páginas 69–75.
- Areces, C., Figueira, S., e Gorín, D. (2011). Using logic in the generation of referring expressions. Em *Proceedings of the 6th International Conference on Logical Aspects of Computational Linguistics (LACL 2011)*, páginas 17–32, Montpellier. Springer.
- Arts, A., Maes, A., Noordman, L. G. M., e Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, 43(1):361–374.
- Byron, D., Koller, A., Oberlander, J., Stoia, L., e Striegnitz, K. (2007). Generating instructions in virtual environments (GIVE): A challenge and evaluation testbed for NLG. Em *Workshop on Shared Tasks and Comparative Evaluation in Natural Language Generation*.
- Clarke, A. D. F., Elsner, M., e Rohde, H. (2013). Where's Wally: The influence of visual salience on referring expression generation. *Frontiers in Psychology*, 4(329).
- Cuevas, R. R. M. e Paraboni, I. (2008). A machine learning approach to portuguese pronoun resolution. *Advances in Artificial Intelligence-IBERAMIA 2008*, LNAI 5290:262–271.
- Dale, R. (2002). Cooking up referring expressions. Em *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, páginas 68–75.
- Dale, R. e Haddock, N. J. (1991). Content determination in the generation of referring expressions. *Computational Intelligence*, 7(4):252–265.
- Dale, R. e Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263.
- Dale, R. e Viethen, J. (2009). Referring expression generation through attribute-based heuristics. Em *Proceedings of ENLG-2009*, páginas 58–65.

- de Lucena, D. J., Paraboni, I., e Pereira, D. B. (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 Novais, E. M. e Paraboni, I. (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.
- dos Santos Silva, D. e Paraboni, I. (2015). Generating spatial referring expressions in interactive 3D worlds. *Spatial Cognition & Computation*, 15(03):186–225.
- Duchowski, A. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470.
- Engelhardt, P. E., Demiral, S. B., e Ferreira, F. (2011). Over-specified referring expressions impair comprehension: An ERP study. *Brain and Cognition*, 77(2):304–314.
- Engelhardt, P. E. e Ferreira, K. B. F. (2006). Do speakers and listeners observe the Gricean maxim of quantity? *Journal of Memory and Language*, 54(4):554–573.
- Eugenio, B. D., Jordan, P. W., Thomason, R. H., e Moore, J. D. (2000). The agreement process: An empirical investigation of human-human computer-mediated collaborative dialogues. *International Journal of Human-Computer Studies*, 53(6):1017–1076.
- Ferreira, T. C. e Paraboni, I. (2014a). Classification-based referring expression generation. *Lecture Notes in Computer Science*, 8403:481–491.
- Ferreira, T. C. e Paraboni, I. (2014b). Referring expression generation: taking speakers' preferences into account. *Lecture Notes in Artificial Intelligence*, 8655:539–546.
- FitzGerald, N., Artzi, Y., e Zettlemoyer, L. (2013). Learning distributions over logical forms for referring expression generation. Em *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, páginas 1914–1925. Association for Computational Linguistics.
- Gardent, C. (2002). Generating minimal definite descriptions. Em *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, páginas 96–103.
- Gargett, A., Garoufi, K., Koller, A., e Striegnitz, K. (2010). The GIVE-2 corpus of giving instructions in virtual environments. Em *Proceedings of LREC-2010*.
- Gatt, A., Belz, A., e Kow, E. (2009). The TUNA challenge 2009: Overview and evaluation results. Em *Proceedings of the 12nd European Workshop on Natural Language Generation*, páginas 174–182.
- Gatt, A., van der Sluis, I., e van Deemter, K. (2007). Evaluating algorithms for the generation of referring expressions using a balanced corpus. Em *Proceedings of ENLG-07*.
- Gorniak, P. e Roy, D. (2004). Grounded semantic composition for visual scenes. *Journal of Artificial Intelligence Research*, 21:429–470.
- Grice, H. P. (1975). Logic and conversation. Em Cole, P. e Morgan, J. L., editores, *Syntax and semantics*, volume 3. New York: Academic Press.
- Guhe, M. (2009). Generating referring expressions with a cognitive model. Em *Production of Referring Expressions: Bridging the gap between computational and empirical approaches to reference*.
- Hansen, D. W. e Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *TPAMI*, 32(3):478–500.
- Kazemzadeh, S., Ordonez, V., Matten, M., e Berg, T. (2014). ReferItGame: Referring to objects in photographs of natural scenes. Em *Proceedings of the 2014 Conference*

- on *Empirical Methods in Natural Language Processing (EMNLP)*, páginas 787–798. Association for Computational Linguistics.
- Koolen, R., Gatt, A., Goudbeek, M., e Krahmer, E. (2011). Factors causing overspecification in definite descriptions. *Journal of Pragmatics*, 43(13):3231–3250.
- Koolen, R., Goudbeek, M., e Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, 37(2):395–411.
- Krahmer, E. e Theune, M. (2002). Efficient context-sensitive generation of referring expressions. Em van Deemter, K. e Kibble, R., editores, *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, páginas 223–264. CSLI Publications, Stanford, CA.
- Krahmer, E. e van Deemter, K. (2012). Computational generation of referring expressions: A survey. *Computational Linguistics*, 38(1):173–218.
- Krahmer, E., van Erk, S., e Verleg, A. (2003). Graph-based generation of referring expressions. *Computational Linguistics*, 29(1):53–72.
- Mitchell, M., van Deemter, K., e Reiter, E. (2010). Natural reference to objects in a visual domain. Em *Proceedings of INLG-2010*. The Association for Computer Linguistics.
- Morimoto, C. H. e Mimica, M. R. M. (2005). Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98:4–24.
- NIST (2002). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics.
- Olson, D. R. (1970). Language and thought: aspects of a cognitive theory of semantics. *Psychological Review*, 77(4):257–273.
- Papineni, S., Roukos, T., Ward, W., e Zhu, W. (2002). Bleu: a method for automatic evaluation of machine translation. Em *Proceedings of ACL-2002*, páginas 311–318.
- Paraboni, I. (1997). Uma arquitetura para a resolução de referências pronominais possessivas no processamento de textos em língua portuguesa. Master's thesis, PUCRS, Porto Alegre.
- Paraboni, I. (2000). An algorithm for generating document-deictic references. Em *Procs. of workshop Coherence in Generated Multimedia, associated with First Int. Conf. on Natural Language Generation (INLG-2000)*, Mitzpe Ramon, páginas 27–31.
- Paraboni, I. (2003). *Generating references in hierarchical domains: the case of Document Deixis*. Tese de Doutorado, University of Brighton.
- Paraboni, I. e de Lima, V. L. S. (1998). Possessive pronominal anaphor resolution in portuguese written texts. Em *Proceedings of the 17th international conference on Computational linguistics-Volume 2*, páginas 1010–1014.
- Paraboni, I., Galindo, M., e Iacovelli, D. (2016). Stars2: a corpus of object descriptions in a visual domain. *Language Resources and Evaluation*.
- Paraboni, I., Galindo, M. R., e Iacovelli, D. (2015). Generating overspecified referring expressions: the role of discrimination. Em *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015*, páginas 76–82.
- Paraboni, I., Masthoff, J., e van Deemter, K. (2006). Overspecified reference in hierarchical domains: measuring the benefits for readers. Em *Proc. of INLG-2006*, páginas 55–62, Sydney.
- Paraboni, I. e van Deemter, K. (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 (esslli99)*, páginas 44–48.
- Paraboni, I. e van Deemter, K. (2002a). Generating easy references: the case of document deixis. Em *INLG-2002, New York*, páginas 113–119.
- Paraboni, I. e van Deemter, K. (2002b). Towards the generation of document-deictic references. Em *Information sharing: reference and presupposition in language generation and interpretation*, páginas 329–352. CSLI Publications.
- Paraboni, I. e van Deemter, K. (2014). Reference and the facilitation of search in spatial domains. *Language, Cognition and Neuroscience*, 29(8):1002–1017.
- Passonneau, R. (2006). Measuring agreement on set-valued items (masi) for semantic and pragmatic annotation. Em *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*.
- Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, 27(1):98–110.
- Pereira, D. B. e Paraboni, I. (2007). A language modelling tool for statistical NLP. Em *5th Workshop on Information and Human Language Technology (TIL-2007). Anais do XXVII Congresso da SBC*, páginas 1679–1688, Rio de Janeiro.
- Pereira, D. B. e Paraboni, I. (2008). Statistical surface realisation of portuguese referring expressions. *Advances in Natural Language Processing*, LNAI 5221:383–392.
- Reiter, E. e Dale, R. (2000). *Building natural language generation systems*. Cambridge University Press, New York, NY, USA.
- Tarenskeen, S., Broersma, M., e Geurts, B. (2014). Referential overspecification: Colour is not that special. Em *RefNet Workshop on Psychological and Computational Models of Reference Comprehension and Production*, Edinburgh, Scotland.
- Teixeira, C. V. M., Paraboni, I., da Silva, A. S. R., e Yamasaki, A. K. (2014). Generating relational descriptions involving mutual disambiguation. *Lecture Notes in Computer Science*, 8403:492–502.
- van Deemter, K., van Gompel, A. G. R., e Krahmer, E. (2012). Toward a computational psycholinguistics of reference production. *Topics in Cognitive Science*, 4:166–183.
- van Gompel, R., Gatt, A., Krahmer, E., e Deemter, K. V. (2014). Testing computational models of reference generation as models of human language production: The case of size contrast. Em *RefNet Workshop on Psychological and Computational Models of Reference Comprehension and Production*, Edinburgh, Scotland.
- van Gompel, R., Gatt, A., Krahmer, E., e van Deemter, K. (2012). PRO: A computational model of referential overspecification. Em *Proceedings of AMLAP-2012*.
- Viethen, J. e Dale, R. (2006). Algorithms for generating referring expressions: Do they do what people do? Em *Proceedings of INLG-2006*, páginas 63–70.
- Viethen, J. e Dale, R. (2011). GRE3D7: A corpus of distinguishing descriptions for objects in visual scenes. Em *Proceedings of UCNLG+Eval-2011*, páginas 12–22.