

Pronomes E-type e resolução de anáfora: investigações computacionais

Daniel Alves da Silva Lopes Diniz
UNICAMP

Um pronome anafórico (isto é, em uma leitura não-dêitica) parece poder, à primeira vista, ter como antecedente ou uma expressão-R ou um sintagma nominal quantificado. Neste último caso, é uma variável linguística tanto sintática quanto semanticamente ligada, de modo que corresponde a uma variável lógica ligada por um quantificador existencial ou universal¹. Tomem-se, no entanto, as seguintes sentenças:

- (1) a. Every man who owns a donkey feeds it. b. If a man owns a donkey, he feeds it.

“It” não pode correferir com “a donkey”, simplesmente por não haver referente a retomar. Deve funcionar, assim, como uma variável ligada (descartando, aqui, uma leitura dêitica). Analogamente, a sentença (2-a) deve expressar ou a proposição (2-b) ou a proposição (2-c):

- (2) a. If Pedro owns a donkey he is rich.
b. $\forall x(\text{DONKEY}(x) \wedge \text{OWNS}(\text{Pedro}, x) \rightarrow \text{RICH}(\text{Pedro}))$
c. $\exists x(\text{DONKEY}(x) \wedge \text{OWNS}(\text{Pedro}, x) \rightarrow \text{RICH}(\text{Pedro}))$

Explicar como um tipo de sintagma que sempre pareceu expressar existência pode expressar universalidade teria consequências contraintuitivas, de modo que parece melhor supor que a sentença (2-a) de fato expressa existência, como em (2-c). Observe-se, no entanto, que em (2-c) o quantificador existencial não tem escopo sobre os termos lógicos que correspondem à segunda oração subordinada de (2-a). Analogamente, portanto, o quantificador existencial relacionado a “a donkey” em (1-a) não terá escopo sobre a oração “he feeds it”. Isso, no entanto, deve ser impossível porque faz com que “it” não seja devidamente ligado, de modo a carecer de referente (KAMP, 2002, p. 190–191). Esse problema de escopo pode ser apresentado também de modo mais sintático: mesmo após ser alçado, “a donkey” não c-comanda “it” em (1-a), de modo que não pode ligá-lo (HEIM, 1990, p. 140). Nesse sentido, Evans (1980) argumenta que, em certos contextos, pronomes com leitura anafórica que, contudo, não correferem com uma expressão-R nem são variáveis ligadas, constituem a categoria dos “pronomes E-type”.

Esse fenômeno suscitou problemas sintático-semânticos que receberam dois tratamentos na literatura: i) na abordagem composicional da semântica formal, considerou-se que pronomes E-type podem ser substituídos por descrições correferentes;

¹ O fato de que a Lógica e a Linguística empregam “variável”, “ligação” e “quantificador” para nomear objetos e fenômenos análogos, mas distintos, implica em uma certa confusão terminológica cuja dissipação pode exigir alguma redundância.

e ii) na Discourse Representation Theory (DRT), conforme Kamp (2002), os indivíduos referenciados e as proposições expressas em uma sequência de sentenças são adicionados ou removidos de estruturas conjuntísticas conforme o discurso é processado, o que permite um tratamento uniforme dos vários tipos de pronome.

Pretende-se aqui discutir os métodos, motivações teóricas e resultados de Diniz (2016), em que intentou-se implementar em Python um algoritmo de resolução de anáforas que determinasse o antecedente do pronome de terceira pessoa plural em contextos E-type ora pressupondo uma dessas semânticas, ora a outra, para verificar se a diferença de estratégias afetaria a eficiência do algoritmo, em que medida e por quê. Ressalte-se que algoritmos desse tipo usualmente visam obter uma taxa de acerto tão próxima daquela do falante nativo quanto possível, adotando para tanto heurísticas com pouca inspiração linguística e métodos estatísticos, sem preocupação com reproduzir o modo com que o falante executa a mesma tarefa. Nesse sentido, essa proposta procurou também investigar quais as dificuldades e eventuais vantagens em empregar métodos explicitamente linguísticos, isto é, em almejar não meramente uma alta taxa de acerto, mas, idealmente, uma representação plausível do processamento linguístico real.

Assim, 30.126 sentenças de 19 textos do Corpus Histórico do Português Tycho Brahe (CHPTB) foram analisadas com um programa escrito em Python 3 que importa várias ferramentas da biblioteca Natural Language Toolkit (NLTK). Embora seja diacrônico e fenômenos de mudança não tenham sido objeto de análise, o CHPTB foi escolhido por ser um dos poucos em português anotados sintaticamente.

Após serem convertidas em objetos estruturados hierarquicamente a partir de suas eti-quetas sintáticas, as sentenças do CHPTB foram binarizadas automaticamente por meio de um método de NLTK. Embora essa binarização por vezes desconecte constituintes relacionados ou conecte constituintes independentes, Diniz (2016) considera que as imprecisões decorrentes desse procedimento são justificadas pelo fato de que a binarização é condição necessária para detecção de pronomes E-type, de modo que o método de NLTK para esse fim afigura-se como uma primeira abordagem razoável. O programa então prosseguiu procurando a palavra “eles” nas estruturas binárias; ao encontrar uma ocorrência desse pronome, varria os constituintes à esquerda e acima, respectivamente, da posição do pronome, em busca de sintagmas etiquetados morfologicamente como nominais plurais.

Do total de sentenças, 234 continham pronomes anafóricos do tipo procurado. Conseguiu-se fazer com que o programa detectasse relações de dominância, irmandade e c-comando entre sintagmas e selecionasse, de uma lista de possíveis antecedentes para uma dada ocorrência de “eles”, aquele com a maior probabilidade de sê-lo. Não se pôde, no entanto, implementar as semânticas composicional e DRT como estratégias relevantes para os procedimentos de decisão realizados pelo programa.

Considerando que Diniz (2016) situa-se em uma interface pouco explorada, especialmente em um âmbito que não o da Ciência da Computação, entre Sintaxe Gerativa, Semântica Formal e Linguística Computacional e discute o emprego de métodos incomuns na área de resolução de anáfora, parecem promissores mesmo esses resultados ainda preliminares e aquém do inicialmente planejado.

REFERÊNCIAS

- DINIZ, D. A. S. L. Aspectos semânticos e computacionais de resolução de pronomes E-type em um corpus histórico do português. 48 f. Monografia (Bacharel em Linguística) — Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2016. Disponível em: <http://www.bibliotecadigital.unicamp.br/document/?code=74301>. Acesso em: 28 jul. 2017.
- EVANS, G. Pronouns. *Linguistic Inquiry*, MIT Press, v. 11, n. 2, p. 337–362, 1980. Disponível em: <http://www.jstor.org/stable/4178164>. Acesso em: 11 mar. 2016.
- HEIM, I. R. E-type pronouns and donkey anaphora. *Linguistics and Philosophy*, Springer, v. 13, n. 2, p. 137–177, 1990. Disponível em: <http://www.jstor.org/stable/25001382>. Acesso em: 12 set. 2016.
- KAMP, H. A theory of truth and semantic representation. In: PARTEE, B. H.; PORTNER, P. (Ed.). *Formal Semantics: The essential readings*. Oxford: Blackwell Publishers Ltd, 2002. cap. 8, p. 189–222. Disponível em: <http://dx.doi.org/10.1002/9780470758335.ch8>. Acesso em: 13 set. 2016.