A Proof System for Dialogical Anaphora Resolution

Davide Catta[1], Richard Moot[2], Christian Retoré[2] and Symon Stevens-Guille[3]

[1] Télécom Paris / CNRS LTCI - UMR5141
[2] CNRS, LIRMM, Univ Montpellier
[3] Ohio State University

Abstract

We present a proof-theoretic account of anaphora resolution, namely a sequent calculus corresponding to dialogical games where two players argue to find the reference of some anaphor.

Keywords

logic, proof theory, formal semantics, anaphora resolution, dialogical games

1. Introduction

In our ongoing work, we propose a proof-theoretic account of anaphora resolution, focusing on pronominal reference to entities. Our account is motivated by the view that knowledge of possible anaphoric dependencies in natural language should be considered part of semantic competence rather than merely the byproduct of extralinguistic mechanisms. Consequently, our account uses the standard tools of contemporary semantic theory, namely formal logic, to model possible resolutions of anaphoric expressions. Unlike most previous work in this tradition [1, 2], we resolve anaphora in the proof theory of our logic, not by checking the model-theoretic truth of some oracle's chosen indexing or reuse of variables. In [3] two of the authors of the present work introduced a dialogical argumentation framework for solving anaphoric dependencies in sentences. This framework introduced a new quantifier (the anaphoric quantifier $\mathscr{A}$) whose formal meaning is defined by the way in which a formula having $\mathscr{A}$ as main connective can be attacked and defended in a dialogical logic [4, 5] inspired framework. Here we present a sequent calculus formalism in which the two rules for the quantifier $\mathscr{A}$ precisely capture those of the above-mentioned dialogical system above.

In Section 2 we introduce the linguistic phenomenon to be modelled — anaphoric reference to entities by pronouns. Section 3 introduces the sequent calculus whose rules implicitly define the meaning of our new quantifier. Section 4 illustrates applications of the logic with commented proofs of the expected readings of pronouns. In particular, we give examples of specific patterns of reference which have attracted the attention of linguists and logicians in the literature: donkey sentences [6] and bathroom sentences.[1] Section 5 briefly discusses soundness for our logic, and Section 6 concludes.

6th Workshop on Advances in Argumentation in Artificial Intelligence (AI³ 2022), November 28, 2022
© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

[1]Roberts [7] attributes the bathroom sentence to Barbara Partee while noting a similar example in [8]

CEUR-WS.org/Vol-3354/short3.pdf

2. Anaphora, pronouns and inferences

Anaphora is the linguistic phenomenon whereby the interpretation of one occurrence of an expression depends on the interpretation of an occurrence of another expression. This characterisation allows us to recognise that in the two sentences “Gertrude eats an apple. It is delicious” the pronoun “it” is anaphoric. The interpretation of the pronoun “it” is the same as the interpretation of the noun “an apple”. Let us call the linguistic expression on which the interpretation of an anaphor depends the antecedent. Humans are generally able to resolve anaphora correctly, i.e., find an appropriate antecedent for an anaphoric expression. Anaphora resolution is a difficult task in natural language processing. We consider anaphora resolution to be an inferential problem. If, for example, the assertion of the two sentences “(A) someone does not smile” and “(B) he has a headache” is justified, then we can conclude that the sentence “(C) someone has a headache and does not smile” is justified. On the contrary, if the assertion of the two sentences “(D) not everyone smiled” and “(B) he had a headache” is justified, then we *cannot* conclude that the assertion of the sentence “‘(C) someone has a headache and does not smile” is justified. In our framework, this will be represented by the fact that the logical representation of (A), together with the one of (B), implies the logical representation of (C). That is: the following formula should be derivable.

$$(1) \quad \overbrace{\exists x_1 \neg \text{smile}(x_1)}^{(A)} \wedge \overbrace{\mathscr{A} y \, \text{has-headache}(y)}^{(B)} \Rightarrow \overbrace{\exists z(\neg \text{smile}(z) \wedge \text{has-headache}(z))}^{(C)}$$

On the contrary, the representation of (D) together with the one of (B) *does not* imply the logical representation of (C). Thus, the following formula should **not** be derivable

$$(2) \quad \overbrace{\neg(\forall x_1 \text{smile}(x_1))}^{(D)} \wedge \overbrace{\mathscr{A} y \, \text{has-headache}(y)}^{(B)} \Rightarrow \overbrace{\exists z(\neg \text{smile}(z) \wedge \text{has-headache}(z))}^{(C)}$$

Remark that we represent the pronoun “he” by a variable bound by an occurrence of the quantifier $\mathscr{A}$.

3. The sequent calculus SAC

We consider a standard first order multisorted language (the reader can consult [9, Chapter 3] for definitions) in which the set of sorted terms only contains variables and constants. Formulae are generated from a set of atomic formulae using the the usual connectives and quantifiers of first order logic $\neg, \Rightarrow, \wedge, \vee, \forall, \exists$, and the anaphoric quantifier $\mathscr{A}$. That is: we add the following formation rule to the usual ones of multisorted first-order logic: if A is a formula and $x^{\mathfrak{s}}$ is a variable of atomic sort s, then $\mathscr{A} x^{\mathfrak{s}} A$ is a formula (in which any occurrence of $x^{\mathfrak{s}}$ is bound).

A sequent $\Gamma \vdash \Delta$, is an expression where Γ and Δ are finite (possibly empty) multisets of formulae. We use Greek capital letters $\Gamma, \Delta, \Pi, \Sigma, \ldots$ to denote arbitrary multisets of formulas.

A sequent calculus is a formalism to construct formal deductive arguments. The arguments, called derivations or proofs, are obtained through the application of inference rules. Inference rules have a (possibly empty) list of sequents as premise and a sequent as conclusion. Proofs in the sequent calculus are trees of sequents that are constructed from a given set of rules.

Table 1
The sequent calculus SAC

$$\frac{}{A \vdash A}\ \mathsf{Ax}$$

$$\frac{\Gamma, A, A \vdash \Delta}{\Gamma, A \vdash \Delta}\ \mathsf{C}^{\mathsf{L}} \qquad \frac{\Gamma \vdash A, A, \Delta}{\Gamma \vdash A, \Delta}\ \mathsf{C}^{\mathsf{R}} \qquad \frac{\Gamma \vdash \Delta, A}{\Gamma, \neg A \vdash \Delta}\ \neg^{\mathsf{L}} \qquad \frac{\Gamma, A \vdash \Delta}{\Gamma \vdash \Delta, \neg A}\ \neg^{\mathsf{R}}$$

$$\frac{\Gamma \vdash A, \Delta \qquad \Sigma, B \vdash \Pi}{\Gamma, \Sigma, A \Rightarrow B \vdash \Delta, \Pi}\ \Rightarrow^{\mathsf{L}} \qquad \frac{\Gamma, A \vdash B, \Delta}{\Gamma \vdash A \Rightarrow B, \Delta}\ \Rightarrow^{\mathsf{R}} \qquad \frac{\Gamma, A, B \vdash \Delta}{\Gamma, A \wedge B \vdash \Delta}\ \wedge^{\mathsf{L}}$$

$$\frac{\Gamma \vdash A, \Delta \qquad \Sigma \vdash B, \Pi}{\Gamma, \Sigma \vdash A \wedge B, \Delta, \Pi}\ \wedge^{\mathsf{R}} \qquad \frac{\Gamma, A \vdash \Delta \qquad \Sigma, B \vdash \Pi}{\Gamma, \Sigma, A \vee B \vdash \Delta, \Pi}\ \vee^{\mathsf{L}} \qquad \frac{\Gamma \vdash A, B, \Delta}{\Gamma \vdash A \vee B, \Delta}\ \vee^{\mathsf{R}}$$

$$\frac{\Gamma, A[\mathrm{k}^{\mathfrak{s}}/x^{\mathfrak{s}}] \vdash \Delta}{\Gamma, \forall x^{\mathfrak{s}} A \vdash \Delta}\ \forall^{\mathsf{L}} \qquad \frac{\Gamma \vdash A[\mathrm{k}^{\mathfrak{s}}/x^{\mathfrak{s}}], \Delta}{\Gamma \vdash \forall x^{\mathfrak{s}} A, \Delta}\ \forall^{\mathsf{R}} \qquad \frac{\Gamma, A[\mathrm{k}^{\mathfrak{s}}/x^{\mathfrak{s}}] \vdash \Delta}{\Gamma, \exists x^{\mathfrak{s}} A \vdash \Delta}\ \exists^{\mathsf{L}} \qquad \frac{\Gamma \vdash A[\mathrm{k}^{\mathfrak{s}}/x^{\mathfrak{s}}], \Delta}{\Gamma \vdash \exists x^{\mathfrak{s}} A, \Delta}\ \exists^{\mathsf{R}}$$

$$\frac{\Gamma_1, A[\mathrm{k_1}^{\mathfrak{s}}/x^{\mathfrak{s}}] \vdash \Delta_1 \qquad \cdots \qquad \Gamma_n, A[\mathrm{k}_n{}^{\mathfrak{s}}/x^{\mathfrak{s}}] \vdash \Delta_n}{\Gamma_1, \ldots, \Gamma_n, \mathscr{A} x^{\mathfrak{s}} A \vdash \Delta_1, \ldots, \Delta_n}\ \mathscr{A}^{\mathsf{L}} \qquad \frac{\Gamma \vdash \Delta, A[\mathrm{k}_1^{\mathfrak{s}}/x^{\mathfrak{s}}], \ldots, A[\mathrm{k}_n^{\mathfrak{s}}/x^{\mathfrak{s}}]}{\Gamma \vdash \Delta, \mathscr{A} x^{\mathfrak{s}} A}\ \mathscr{A}^{\mathsf{R}}$$

Definition 1. *Given the rules in Table 1, a prederivation in* SAC *(Sequent Anaphoric Calculus) is a tree where each local subtree is an instance of one of the rules subject to the following conditions: in the premises of* $\exists^{\mathsf{L}}$ *and* $\forall^{\mathsf{R}}$*,* $k^{\mathfrak{s}}$ *does not appear in* Γ, Δ*. The number of premises of* $\mathscr{A}^{\mathsf{L}}$ *is at least one. In* $\mathscr{A}^{\mathsf{R}}$*, the constant* $k_1^{\mathfrak{s}}, \ldots, k_n^{\mathfrak{s}}$ *are all and only the constants of sort* $\mathfrak{s}$ *appearing in* Γ*.*

The first condition is the usual condition on quantifiers in first-order logic. For the $\mathscr{A}^{\mathsf{R}}$, we require that the constants $k_i^{\mathfrak{s}}$ *does* appear, but they must appear in Γ on the left-hand side of the turnstile. For the $\mathscr{A}^{\mathsf{L}}$ we require for the moment only that the rule has at least one premise, but we will refine this when moving from prederivations to derivations.

We say that a constant k appears negatively in a sequent $\Gamma \vdash \Delta$ if k appears in some formula in Γ. Let $\mathscr{D}$ be a prederivation, and $\mathcal{P} = x_0, \ldots, x_n$ a path in $\mathscr{D}$ where x_0 is the root of $\mathscr{D}$ and x_i is the mother of x_{i+1} for all $i < n$. We say a constant k^s appears negatively in $\mathcal{P}$ if k_s appears negatively in one of the path's sequents.

Definition 2. *A prederivation is a* derivation *if, whenever* $\Gamma_1, B(k_1^s), \ldots, \Gamma_n, B(k_n^s)$ *are the* n *premises of an* $\mathcal{A}^{\mathsf{L}}$ *rule then* $k_1^{\mathfrak{s}}, \ldots k_n^{\mathfrak{s}}$ *are* all and only the constants of sort $\mathfrak{s}$ *which appear negatively in the path from the root of the derivation* $\mathscr{D}$ *to the conclusion* $\Gamma_1, \ldots, \Gamma_n, \mathcal{A}\, x^s B \vdash \Delta_1, \ldots \Delta_n$ *of the rule.*

Notation we use $x_1, x_2, x_3, \ldots$ for variables that are bounded by a universal or an existential quantifier and $y_1, y_2, y_3 \ldots$ for variables that are bounded by an anaphoric quantifier.

4. Linguistic examples

Bathroom sentence Consider the sentence "Either there is no bathroom in this house or *it* is upstairs". Clearly the pronoun "it" refers to "the bathroom" i.e., we expect that 3 below entails 4.

(3) $(\neg\exists x_1 \mathit{Bathroom}(x_1)) \vee \mathscr{A} y_1 \mathit{Upstairs}(y_1)$

(4) $(\neg\exists x_1 \mathit{Bathroom}(x_1)) \vee (\exists x_2(\mathit{Bathroom}(x_2) \wedge \mathit{Upstairs}(x_2))$

The proof below shows how we derive the entailment $3 \vdash 4$. The condition on the $\mathscr{A}^{\mathsf{L}}$ rule is satisfied trivially, since k_1 appears directly in the conclusion of the rule (in the formula coloured blue) and k_1 is the only constant k_i which appears in the path from the conclusion of the rule to the root of the proof.

$$
\begin{prooftree}
\AxiomC{}
\RightLabel{Ax}
\UnaryInfC{$\neg\exists x_1 B(x_1) \vdash \neg\exists x_1 B(x_1)$}
\AxiomC{}
\RightLabel{Ax}
\UnaryInfC{$U(\mathrm{k}_1) \vdash U(\mathrm{k}_1)$}
\AxiomC{}
\RightLabel{Ax}
\UnaryInfC{$B(\mathrm{k}_1) \vdash B(\mathrm{k}_1)$}
\RightLabel{$\wedge^{\mathsf{R}}$}
\BinaryInfC{$U(\mathrm{k}_1), B(\mathrm{k}_1) \vdash (B(\mathrm{k}_1) \wedge U(\mathrm{k}_1))$}
\RightLabel{$\exists^{\mathsf{R}}$}
\UnaryInfC{$U(\mathrm{k}_1), B(\mathrm{k}_1) \vdash \exists x_2(B(x_2) \wedge U(x_2))$}
\RightLabel{$\mathscr{A}^{\mathsf{L}}$}
\UnaryInfC{$\mathscr{A} y_1 U(y_1), \color{blue}{B(\mathrm{k}_1)} \vdash \exists x_2(B(x_2) \wedge U(x_2))$}
\RightLabel{$\exists^{\mathsf{L}}$}
\UnaryInfC{$\mathscr{A} y_1 U(y_1), \exists x_1 B(x_1) \vdash \exists x_2(B(x_2) \wedge U(x_2))$}
\RightLabel{$\vee^{\mathsf{L}}$}
\BinaryInfC{$\neg\exists x_1 B(x_1)) \vee \mathscr{A} y_1 U(y), \exists x_1 B(x_1) \vdash \neg\exists x_1 B(x_1), \exists x_2(B(x_2) \wedge U(x_2))$}
\RightLabel{$\neg^{\mathsf{R}}$}
\UnaryInfC{$\neg\exists x_1 B(x_1)) \vee \mathscr{A} y_1 \mathsf{U}(y), \exists x_1 B(x_1) \vdash \neg\exists x_1 B(x_1), \exists x_2(B(x_2) \wedge U(x_2))$}
\RightLabel{$\Rightarrow^{\mathsf{R}}$}
\UnaryInfC{$\neg\exists x_1 B(x_1)) \vee \mathscr{A} y_1 U(y) \vdash \neg\exists x_1 B(x_1), \neg\exists x_1 B(x_1), \exists x_2(B(x_2) \wedge U(x_2))$}
\RightLabel{C^{R}}
\UnaryInfC{$\neg\exists x_1 B(x_1)) \vee \mathscr{A} y_1 U(y) \vdash \neg\exists x_1 B(x_1), \exists x_2(B(x_2) \wedge U(x_2))$}
\RightLabel{$\vee^{\mathsf{R}}$}
\UnaryInfC{$(\neg\exists x_1 B(x_1)) \vee \mathscr{A} y_1 U(y) \vdash (\neg\exists x_1 B(x_1)) \vee (\exists x_2(B(x_2) \wedge U(x_2)))$}
\end{prooftree}
$$

Donkey sentences Consider the sentence "If a monster fights Guts then he kills it". Here the pronoun "he" refers to Guts since Guts is an (imaginary) human being while the pronoun "it" refer to "a monster". Moreover, the pronoun "it" has a universal reading: what the sentence means is "Any monster that fights Guts will be killed by Guts". Thus we expect that 5 below implies 6.

(5) $[\exists x_1^{nh}(\mathit{Monster}(x_1) \wedge \mathit{Fight}(\mathrm{g}^m, x_1))] \Rightarrow \mathscr{A} y_1^m \mathscr{A} y_2^{nh} \mathit{Kill}(y_1, y_2)$

(6) $\forall x_2^{nh}[\mathit{Monster}(x_2) \wedge \mathit{Fight}(\mathrm{g}^m, x_2) \Rightarrow \mathit{Kill}(\mathrm{g}^m, x_2)]$

The proof below shows how we derive the entailment $5 \vdash 6$. The condition on the first $\mathscr{A}^{\mathsf{L}}$-rule (the one colored in red) is satisfied trivially, since g^m is the only constant of sort m (male human) which appears in the path from the conclusion to the root. The condition on the other $\mathscr{A}^{\mathsf{L}}$ is also trivially satisfied: $\mathrm{k}_1{}^{nh}$ is the only constant of sort nh (non-human) appearing in the path proof.

$$
\begin{prooftree}
\AxiomC{}
\RightLabel{Ax}
\UnaryInfC{$M(\mathrm{k}_1) \wedge F(\mathrm{g}^m, \mathrm{k}_1) \vdash M(\mathrm{k}_1) \wedge F(\mathrm{g}^m, \mathrm{k}_1)$}
\UnaryInfC{$M(\mathrm{k}_1) \wedge F(\mathrm{g}^m, \mathrm{k}_1) \vdash \exists x_1{}^{nh}(M(x_1) \wedge F(\mathrm{g}^m, x_1))$}
\AxiomC{}
\RightLabel{Ax}
\UnaryInfC{$K(\mathrm{g}^m, \mathrm{k}_1) \vdash K(\mathrm{g}^m, \mathrm{k}_1)$}
\RightLabel{$\mathscr{A}^{\mathsf{L}}$}
\UnaryInfC{$\mathscr{A} y_2^{nh} K(\mathrm{g}^m, y_2) \vdash K(\mathrm{g}^m, \mathrm{k}_1)$}
\RightLabel{$\color{red}{\mathscr{A}^{\mathsf{L}}}$}
\UnaryInfC{$\mathscr{A} y_1{}^m \mathscr{A} y_2^{nh} K(y_1, y_2) \vdash K(\mathrm{g}^m, \mathrm{k}_1)$}
\RightLabel{$\Rightarrow^{\mathsf{L}}$}
\BinaryInfC{$[\exists x_1^{nh}(M(x_1) \wedge F(\mathrm{g}^m, x_1))] \Rightarrow \mathscr{A} y_1 \mathscr{A} y_2 K(y_1, y_2), M(\mathrm{k}_1) \wedge F(\mathrm{g}^m, \mathrm{k}_1) \vdash K(\mathrm{g}^m, \mathrm{k}_1)$}
\RightLabel{$\Rightarrow^{\mathsf{R}}$}
\UnaryInfC{$[\exists x_1^{nh}(M(x_1) \wedge F(\mathrm{g}^m, x_1))] \Rightarrow \mathscr{A} y_1 \mathscr{A} y_2 K(y_1, y_2) \vdash M(\mathrm{k}_1) \wedge F(\mathrm{g}^m, \mathrm{k}_1) \Rightarrow K(\mathrm{g}^m, \mathrm{k}_1)$}
\RightLabel{$\forall^{\mathsf{R}}$}
\UnaryInfC{$[\exists x_1^{nh}(M(x_1) \wedge F(\mathrm{g}^m, x_1))] \Rightarrow \mathscr{A} y_1 \mathscr{A} y_2 K(y_1, y_2) \vdash \forall x_2[M(x_2) \wedge F(\mathrm{g}^m, x_2) \Rightarrow K(\mathrm{g}^m, x_2)]$}
\end{prooftree}
$$

The reader can check that the formula in 1 is derivable, while there is no derivation of the formula in 2

5. Relation to LK and soundness

Let the expression $\mathscr{A}^{\{k_1^s,\ldots,k_p^s\}}x.A[x]$ ($\mathscr{A}$ is indexed by a finite set of constants) stand for the classical formula $A[k_1^s] \vee \cdots \vee A[k_p^s]$. The $\mathscr{A}$ rules of SAC, adding the proper indices, become derivable as particular cases of $\vee$ rules. Our calculus SAC consists of a proper subset of classical proofs in LK sequent calculus (more precisely LK without the rules of weakening) with $\mathscr{A}^K.A[x]$ instead of the disjunction $\vee_{k\in K} A[k]$ [2] in which all superscripts are thereafter masked: a proof in SAC is the projection of a proof in LK. Because provable sequents in LK are true in all models, so are ours with the meaning of $\mathscr{A}x.A[x]$: there is a finite set of constants $K = \{k_1^s, \ldots, k_p^s\}$ such that the formula $A[k_1^s] \vee \cdots \vee A[k_p^s]$ is true. This establishes the soundness of SAC.

6. Conclusions and future work

We have introduced SAC, a sequent calculus for anaphora resolution, which corresponds to the resolution with argumentative dialogues of [3]. We illustrated how it can be use to solve several standard puzzles concerning anaphoric dependencies in formal semantics. In the future, as SAC enjoys soundness (cf. previous section), we intend to establish completeness by defining a model as a family of relevant logic models in order to express the global conditions on proofs.

References

[1] H. Kamp, A theory of truth and semantic representation, in: Truth, interpretation and information, Foris Dordrecht, 1984, pp. 1–41.

[2] J. Groenendijk, M. Stokhof, Dynamic predicate logic, Linguist Philos (1991) 39–100.

[3] D. Catta, S. Stevens-Guille, Lorenzen won the game, Lorenz did too: Dialogical logic for anaphora and ellipsis resolution, in: A. Silva, R. Wasserman, R. de Queiroz (Eds.), WoLLIC 2021, volume 13038 of *LNCS*, Springer, 2021.

[4] P. Lorenzen, Logik und agon, Atti Del XII Congresso Int. di Filosofia 4 (1958) 187–194.

[5] P. Lorenzen, K. Lorenz, Dialogische Logik, Wissenschaftliche Buchgesellschaft, 1978.

[6] P. T. Geach, Reference and generality: An examination of some medieval and modern theories, Mind 73 (1964) 575–583.

[7] C. Roberts, Modal subordination and pronominal anaphora in discourse, Linguist Philos 12 (1989) 683–721.

[8] G. Evans, Pronouns, quantifiers, and relative clauses (l), Can. J. Philos 7 (1977) 467–536.

[9] R. Moot, C. Retoré, The logic of categorial grammars: a deductive account of natural language syntax and semantics, volume 6850 of *LNCS*, Springer, 2012.

[2] From the rules of classical logic one clearly has $\mathscr{A}^K x.A[x] \vdash \mathscr{A}^{K\cup J} x.A[x]$.