Monadic Second-Order Logic and Transitive Closure Logics over Trees

Hans-Jörg Tiede
Department of Mathematics and Computer Science
Illinois Wesleyan University
and
Stephan Kepser
Collaborative Research Centre 441
University of Tübingen
Tübingen, Germany

WoLLIC'2006

Mathematical Linguistics

- Classifying the complexity of natural languages and linguistic theories.
- Testing linguistic theories.
- Complexity measures: automata, time/space, learnability.
- Problem: some linguistic theories are not easily analyzed wrt these measures directly.

Model Theoretic Syntax

- Research paradigm in mathematical linguistics.
- Capturing the descriptive complexity of grammar formalisms.
- Treat trees as (finite) models.
- Represent grammars as formulas.
- A formula φ of some logic $\mathcal{L}$ specifies a set of trees:

$$\{\mathcal{M} \mid \mathcal{M} \models \varphi\}$$

- Descriptive complexity of a grammar $\mathcal{G}$: the least expressive logic in which $\mathcal{G}$ can formalized.

Why Logics of Trees?

- Linguists are interested in strong generative capacity.
- Many linguistic theories are formulated as well-formedness condition of trees.
- Logics of strings don't have strong generative capacity.
- Complexity: regular tree languages have CFLs as their yields.

Properties of Logics

- Expressive/natural
- Decidable
- Correspond to grammar formalism/language family
- Tension between these requirements:
  - lack of closure/decision properties leads to unnatural/undecidable logics

Logics

- Modal Logics (Moss and Tiede, 2006)
- Monadic Second Order Logic (Rogers, 1998)
- Both are limited to regular tree languages (context-free string languages).
- Decidable

Achievements and Challenges

- Formalizing constraint-based grammars (e.g. GB: English is CF)
- Capturing non-CF phenomena:
  - Multi-dimensional trees (Rogers)
  - Two-step approach (Mönnich)
  - Quantification over certain functions (Langholm)
  - All extend MSO syntactically
- Problems with Extensions
  - Indirection, but decidable (Rogers, Mönnich)
  - Logically odd, corresponds to language class (Langholm)

Proposal: Transitive Closure Logic

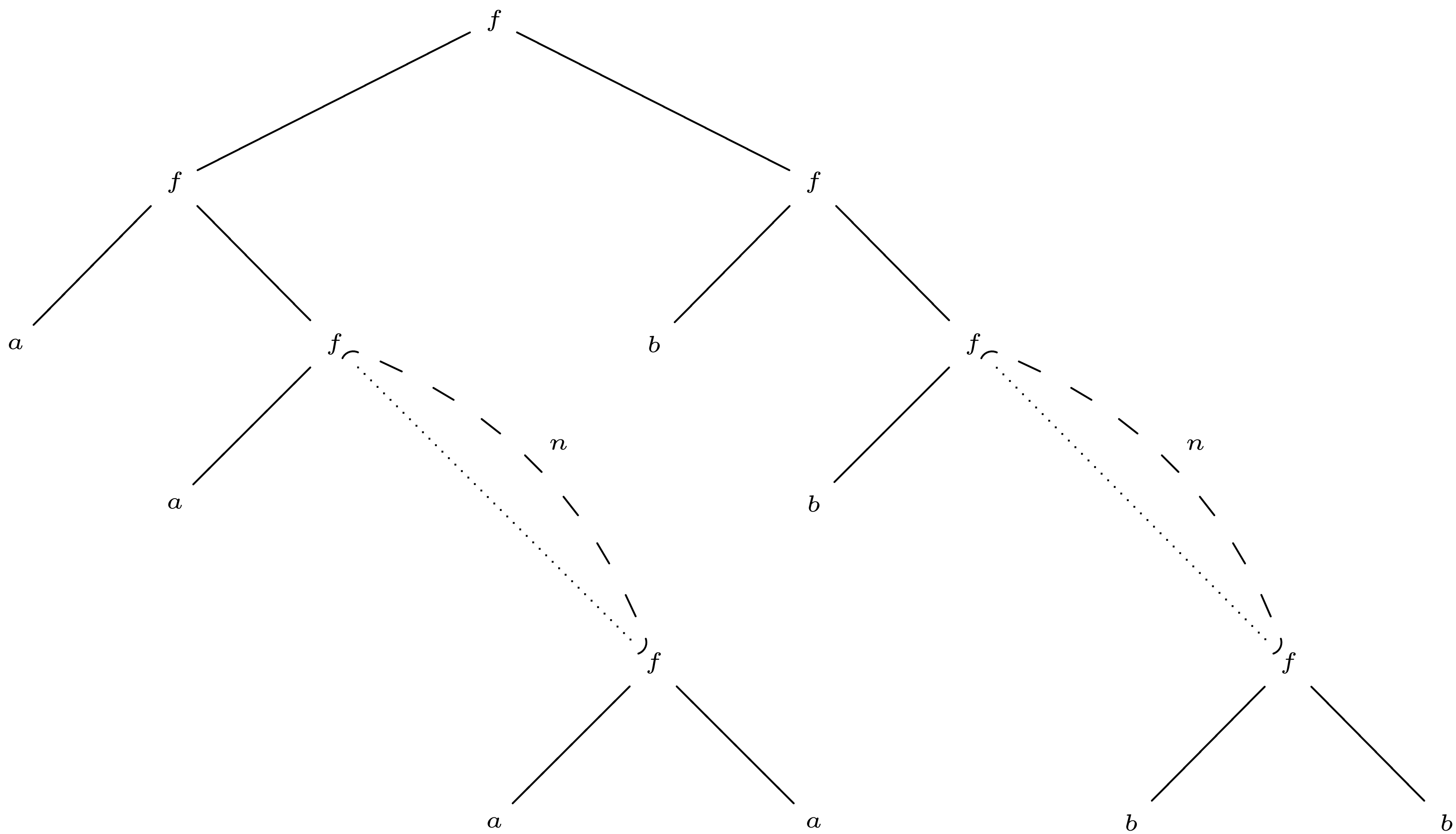
- History: Immerman
  - FO(DTC) = LOGSPACE
  - FO(TC) = NLOGSPACE
- TC is not FO-definable.
- Syntax: $\varphi ::= [(D)TC_{\bar{x},\bar{y}}\varphi](\bar{s},\bar{t})$
- Semantics:

$$
\begin{array}{rcl}
\mathcal{M} & \models & [(D)TC_{\bar{x},\bar{y}}\varphi]\bar{s},\bar{t} \\
& \Leftrightarrow & \\
(\bar{s}^{\mathcal{M}},\bar{t}^{\mathcal{M}}) & \in & (D)TC\{(\bar{a},\bar{b}) \mid \mathcal{M} \models \varphi[\bar{a},\bar{b}]\}.
\end{array}
$$

FO((D)TC) and MSO

- FO((D)TC^n): restrict tuple width to n.
- **Theorem** Every regular tree language is FO(DTC)-definable.
- **Proposition** FO(TC^1) $\leq$ MSO. Strictness is an open problem (equal over strings).
- **Proposition** MSO $<$ FO(DTC). (Whether MSO $\leq$ FO(TC^2) is open.)
- **Proposition** There exists a non-regular tree language that can be defined in FO(DTC^2).
- **Proof** We can define the predicate "the distance from x_1 to y_1 on a right branch is the same as the distance from x_2 to y_2":

$$[DTC_{(x_1,x_2),(y_1,y_2)}(S_2(x_1,y_1) \wedge S_2(x_2,y_2))]$$


f
f
f
a
f
b
f
a
n
b
n
f
f
a
a
b
b

Subtree Isomorphism

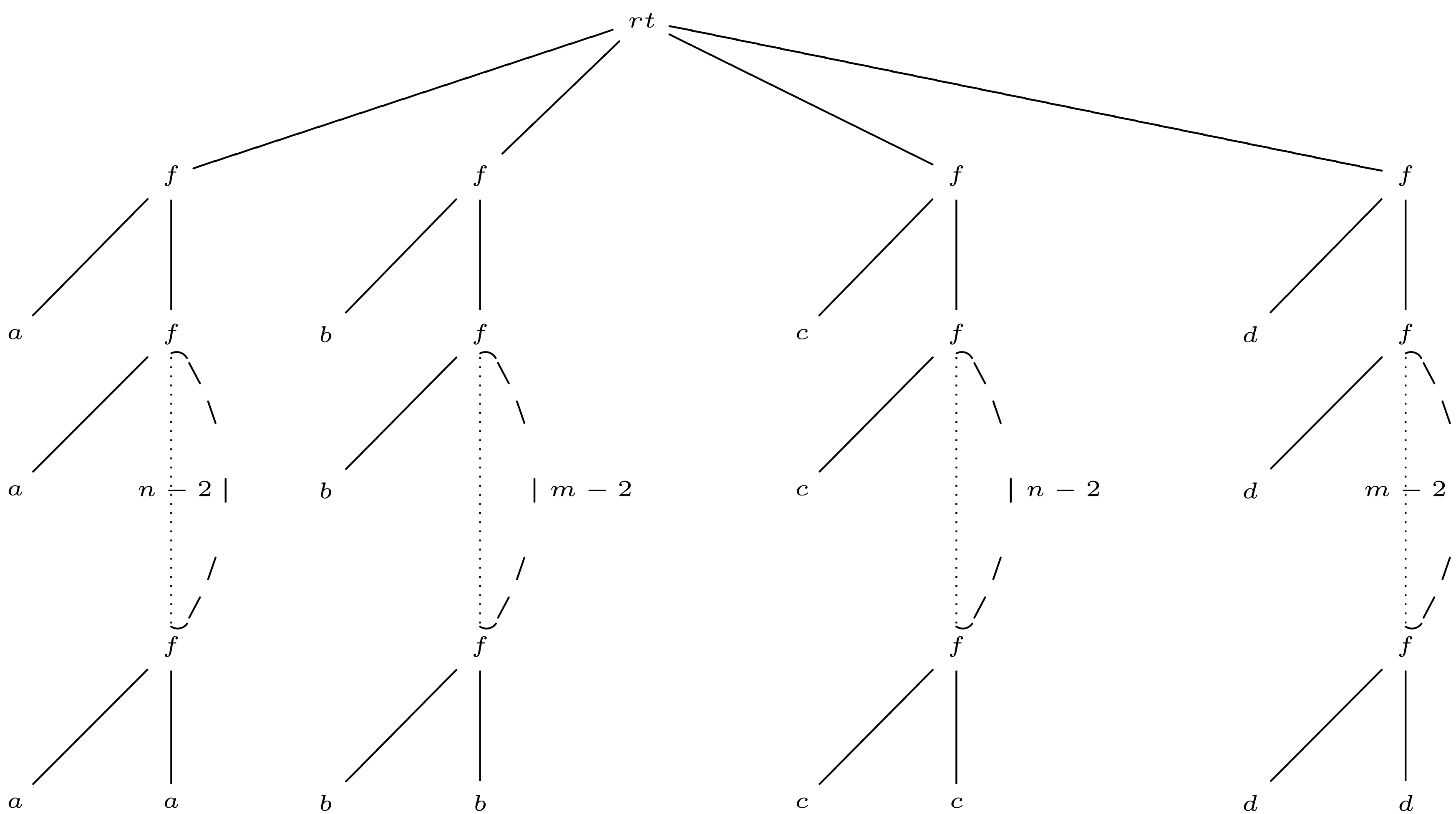
- **Theorem** (Rogers) MSO + Subtree Isomorphism is undecidable.
- **Observation** FO + Subtree Isomorphism is undecidable (with constants).
- **Proposition** Subtree Isomorphism is FO(TC^2)-definable:

$$[TC_{(x_1,x_2)(y_1,y_2)} \bigvee_{i=1}^{r} (S_i(x_1,y_1) \wedge S_i(x_2,y_2))]$$

- **Corollary** FO(TC^2) is undecidable over finite trees.
- **Corollary** FO(DTC^2) is undecidable over the finite trees.

Application: Cross-Serial Dependencies

- Shieber: Swiss German is not a CFL.
- $\mathrm{NP}_1\mathrm{NP}_2\mathrm{NP}_3\mathrm{V}_1\mathrm{V}_2\mathrm{V}_3$
- $a^n b^m c^n d^m$



rt
f
f
f
f
a
f
b
f
c
f
d
f
a
$n-2$
b
$m-2$
c
$n-2$
d
$m-2$
f
f
f
f
a
a
b
b
c
c
d
d

Conclusion

- $FO(DTC^2)$
  - is a natural, expressive, (semantically) minimal extension of MSO.
  - can be used to describe non-CF properties of natural languages.
  - is undecidable.
  - corresponds to tree-walking pebble automata (Engelfriet [et al.]).
  - is not known to correspond to any grammar formalism.