

Logics of subwords

Philippe Schnoebelen

LSV, CNRS & ENS Cachan

26 oct 2017

THE FIRST-ORDER LOGIC OF SUBWORDS

We consider $\text{FO}(A^*; \preceq)$ formulas, like

$$\forall u, v, w: u \preceq v \wedge v \preceq w \implies u \preceq w \quad (\varphi_1)$$

$$\forall u: ab \preceq u \wedge ba \preceq u \implies aa \preceq u \vee bb \preceq u \quad (\varphi_2)$$

$$\exists u: abcd \preceq u \wedge bcde \preceq u \wedge abcde \not\preceq u \quad (\varphi_3)$$

$$\forall u, v: \exists w: \left(\begin{array}{l} u \preceq w \wedge v \preceq w \\ \wedge \forall t: [u \preceq t \wedge v \preceq t \implies w \preceq t] \end{array} \right) \quad (\varphi_4)$$

$$\exists u_1, \dots, u_n \in a^+b^+: \bigwedge_{1 \leq i < j \leq n} u_i \perp u_j \quad (\varphi_{5,n})$$

NB: Whether $A^* \models \varphi$ may depend on A

VALIDITY PROBLEM FOR LOGICS OF WORDS

Problem: Given A and a sentence φ in $\text{FO}(A^*; \leq)$, is φ true?

$\text{FO}(A^*; \leq_{\text{prefix}})$ —even $\text{MSO}(A^*, \leq_{\text{prefix}})$ — is decidable (Rabin 1969)

$\text{FO}(A^*; \cdot)$: undecidable (Quine, 1946) but the Σ_1 fragment is decidable: cf. word equations (Makanin, 1977; Büchi & Senger, 1986/7)

$\text{FO}(A^*; \leq_{\text{infix}})$: undecidable (Kuske 2006)

What about $\text{FO}(A^*; \leq)$?

VALIDITY PROBLEM FOR LOGICS OF WORDS

Problem: Given A and a sentence φ in $\text{FO}(A^*; \leq)$, is φ true?

$\text{FO}(A^*; \leq_{\text{prefix}})$ —even $\text{MSO}(A^*, \leq_{\text{prefix}})$ — is decidable (Rabin 1969)

$\text{FO}(A^*; \cdot)$: undecidable (Quine, 1946) but the Σ_1 fragment is decidable: cf. word equations (Makanin, 1977; Büchi & Senger, 1986/7)

$\text{FO}(A^*; \leq_{\text{infix}})$: undecidable (Kuske 2006)

What about $\text{FO}(A^*; \leq)$?

WHAT ABOUT $\text{FO}(A^*; \leq)$?

Comon & Treinen, 1994: small extension $\text{FO}(A^*; \leq, p_a)$ (with prefixing function $p_a : u \mapsto a \cdot u$) is undecidable, even the Σ_4 fragment, on a 3-letter alphabet

Kuske, 2006: $\text{FO}(A^*; \leq)$ undecidable, even the Σ_3 fragment on a 2-letter alphabet. And the Σ_1 fragment is decidable

Kudinov, Selivanov & Yartseva, 2010: $\text{FO}(A^*; \leq)$ is computably isomorphic to $\text{FO}(\omega; +, \times)$, aka first-order arithmetic

Karandikar & Schnoebelen, 2015: The Σ_2 fragment is undecidable, even over a “small” fixed alphabet, and eventually a 2-letter alphabet

Karandikar & Schnoebelen, 2016: The FO^2 fragment is decidable with elementary complexity

Halfon, Schnoebelen & Zetsche, 2017: For the Σ_1 fragment, constants matter

FO(A^* ; \leq) WITH OR WITHOUT CONSTANTS?

Unlike “ $\forall u, v, w : u \leq v \leq w \implies u \leq w$ ”, some formulas use **constants**, e.g., “ $ab \leq u \wedge ba \leq u \implies (aa \leq u \vee bb \leq u)$ ”

Same for “ $x \in a^+b^+$ ”, short for “ $ab \leq x \wedge ba \not\leq x \wedge c \not\leq x \wedge \dots$ ”

This is FO(A^* ; \leq) vs. FO(A^* ; \leq, w_1, w_2, \dots)

Anyway, constant words can be defined in FO(A^* ; \leq):

$\psi_e(u) \stackrel{\text{def}}{\equiv} \forall x : u \leq x$ defines “ $u = \varepsilon$ ”

$\psi_l(v) \stackrel{\text{def}}{\equiv} \forall x : x \leq v \implies (\psi_e(x) \vee v \leq x)$ defines “ v is a letter or ε ”

NB: We can state “ $|A| = n$ ” and “ $|A| = \aleph_0$ ” in FO(A^* ; \leq)

Defining words: we can define e.g., “ $v \doteq a_1 a_1 a_2 a_1 a_3$ ” without using constants but this is defined “modulo automorphisms of the (A^* ; \leq) structure”.

FO(A^* ; \leq) WITH OR WITHOUT CONSTANTS?

Unlike “ $\forall u, v, w : u \leq v \leq w \implies u \leq w$ ”, some formulas use **constants**, e.g., “ $ab \leq u \wedge ba \leq u \implies (aa \leq u \vee bb \leq u)$ ”

Same for “ $x \in a^+b^+$ ”, short for “ $ab \leq x \wedge ba \not\leq x \wedge c \not\leq x \wedge \dots$ ”

This is FO(A^* ; \leq) vs. FO(A^* ; \leq, w_1, w_2, \dots)

Anyway, constant words can be defined in FO(A^* ; \leq):

$\psi_e(u) \stackrel{\text{def}}{\equiv} \forall x : u \leq x$ defines “ $u = \varepsilon$ ”

$\psi_l(v) \stackrel{\text{def}}{\equiv} \forall x : x \leq v \implies (\psi_e(x) \vee v \leq x)$ defines “ v is a letter or ε ”

NB: We can state “ $|A| = n$ ” and “ $|A| = \aleph_0$ ” in FO(A^* ; \leq)

Defining words: we can define e.g., “ $v \doteq a_1a_1a_2a_1a_3$ ” without using constants but this is defined “modulo automorphisms of the (A^* ; \leq) structure”.

FO($A^*; \leq$) WITH OR WITHOUT CONSTANTS?

Unlike “ $\forall u, v, w : u \leq v \leq w \implies u \leq w$ ”, some formulas use **constants**, e.g., “ $ab \leq u \wedge ba \leq u \implies (aa \leq u \vee bb \leq u)$ ”

Same for “ $x \in a^+b^+$ ”, short for “ $ab \leq x \wedge ba \not\leq x \wedge c \not\leq x \wedge \dots$ ”

This is FO($A^*; \leq$) vs. FO($A^*; \leq, w_1, w_2, \dots$)

Anyway, constant words can be defined in FO($A^*; \leq$):

$\psi_e(u) \stackrel{\text{def}}{\equiv} \forall x : u \leq x$ defines “ $u = \varepsilon$ ”

$\psi_l(v) \stackrel{\text{def}}{\equiv} \forall x : x \leq v \implies (\psi_e(x) \vee v \leq x)$ defines “ v is a letter or ε ”

NB: We can state “ $|A| = n$ ” and “ $|A| = \aleph_0$ ” in FO($A^*; \leq$)

Defining words: we can define e.g., “ $v \doteq a_1 a_1 a_2 a_1 a_3$ ” without using constants but this is defined “modulo automorphisms of the ($A^*; \leq$) structure”.

Fragment 1:

Solving Subword Constraints

SUBWORD CONSTRAINTS

“Subword Constraints” \equiv the Σ_1 -fragment

$$abc \not\leq u \wedge u \leq v \wedge u \not\leq baa \wedge \dots \wedge v \perp w$$

How do we compute a set of solutions?

Recall: “The Σ_1 fragment is **decidable**” (in fact NP-complete)

Yes but this was about the logic **without constants**!

Ok but “constants can be defined within the logic”, no?

Well, we defined ε by a Π_1 formula ...

Bottom Line: we don't really know whether the Σ_1 fragment of $\text{FO}(A^*; \leq, w_1, w_2, \dots)$ is decidable

Thm. (Halfon, Schnoebelen & Zetsche, 2017): the fragment is **undecidable** (Surprise! Surprise!)

SUBWORD CONSTRAINTS

“Subword Constraints” \equiv the Σ_1 -fragment

$$abc \not\leq u \wedge u \leq v \wedge u \not\leq baa \wedge \dots \wedge v \perp w$$

How do we compute a set of solutions?

Recall: “The Σ_1 fragment is **decidable**” (in fact NP-complete)

Yes but this was about the logic **without constants**!

Ok but “constants can be defined within the logic”, no?

Well, we defined ε by a Π_1 formula ...

Bottom Line: we don't really know whether the Σ_1 fragment of $\text{FO}(A^*; \leq, w_1, w_2, \dots)$ is decidable

Thm. (Halfon, Schnoebelen & Zetsche, 2017): the fragment is **undecidable** (Surprise! Surprise!)

SUBWORD CONSTRAINTS

“Subword Constraints” \equiv the Σ_1 -fragment

$$abc \not\leq u \wedge u \leq v \wedge u \not\leq baa \wedge \dots \wedge v \perp w$$

How do we compute a set of solutions?

Recall: “The Σ_1 fragment is **decidable**” (in fact NP-complete)

Yes but this was about the logic **without constants**!

Ok but “constants can be defined within the logic”, no?

Well, we defined ε by a Π_1 formula ...

Bottom Line: we don't really know whether the Σ_1 fragment of $\text{FO}(A^*; \leq, w_1, w_2, \dots)$ is decidable

Thm. (Halfon, Schnoebelen & Zetsche, 2017): the fragment is **undecidable** (Surprise! Surprise!)