Rankers over Infinite Words*

(Extended Abstract)

Luc Dartois[1], Manfred Kufleitner[2], and Alexander Lauser[2]

[1] ENS Cachan, France
`ldartois@dptinfo.ens-cachan.fr`
[2] FMI, Universität Stuttgart, Germany
{`kufleitner`, `lauser`}`@fmi.uni-stuttgart.de`

Abstract. We consider the fragments FO^2, $\Sigma_2 \cap \mathrm{FO}^2$, $\Pi_2 \cap \mathrm{FO}^2$, and Δ_2 of first-order logic $\mathrm{FO}[<]$ over finite and infinite words. For all four fragments, we give characterizations in terms of rankers. In particular, we generalize the notion of a ranker to infinite words in two possible ways. Both extensions are natural in the sense that over finite words they coincide with classical rankers, and over infinite words they both have the full expressive power of FO^2. Moreover, the first extension of rankers admits a characterization of $\Sigma_2 \cap \mathrm{FO}^2$ while the other leads to a characterization of $\Pi_2 \cap \mathrm{FO}^2$. Both versions of rankers yield characterizations of the fragment $\Delta_2 = \Sigma_2 \cap \Pi_2$. As a byproduct, we also obtain characterizations based on unambiguous temporal logic and unambiguous interval temporal logic.

1 Introduction

We consider fragments of two-variable first-order logic FO^2. Formulas are interpreted over words which may be finite or infinite. Over finite words only, a large number of different characterizations of FO^2 is known, see e.g. [8] or [2] for an overview. Some of the characterizations have been generalized to infinite words in [3]. We continue this line of work. For this paper the main difference between finite word models and infinite word models is the following: Over finite words, FO^2 and the fragment $\Delta_2 = \Sigma_2 \cap \Pi_2$ have the same expressive power [9], whereas Δ_2 is a strict subclass of FO^2 over infinite words. Moreover, in the case of infinite words, FO^2 is incomparable to Σ_2 and Π_2. By definition, Σ_2 is the class of formulas in prenex normal form with two blocks of quantifiers starting with a block of existential quantifiers, and Π_2 is the class of negations of Σ_2-formulas. Here and throughout the paper, we identify a logical fragment with the class of languages definable in the fragment.

An important concept in this paper are rankers which have been introduced by Immerman and Weis [10] in order to give a combinatorial characterization of quantifier alternation within FO^2 over finite words. Casually speaking, a *ranker*

* The last two authors acknowledge the support by the German Research Foundation (DFG) for the project DI 435/5-1.

is a sequence of instructions of the form *"go to the next a-position"* and *"go to the previous a-position"* for some letters a. For every word, a ranker is either undefined or it determines a unique position. We generalize rankers to infinite words in two possible ways. The main difference to finite words is that we have to define the semantics of *"go to the last a-position"* if there are infinitely many occurrences of the letter a. The first solution is to say that the position is undefined. The second approach is to stay at an infinite position. For example, if a word has infinitely many a-positions but only two b-positions, then in the first semantics *"go to the last a-position and from there, go to the previous b-position"* would be undefined while in the second semantics it would determine the last b-position. By delaying the interpretation of instructions until some letter with finite occurrence is met, the second semantics is reminiscent of the *lazy evaluation* principle, and we therefore call it *lazy rankers*. If we want to emphasize that we use the first semantics, then we often use the term *eager ranker*. The language $L(r)$ generated by a ranker r consists of all words on which r is defined. A *ranker language* is a Boolean combination of languages of the form $L(r)$.

In both ways, rankers admit natural combinatorial characterizations of the first-order fragments FO^2 (Theorems 1 and 5) and Δ_2 (Theorem 3) over finite and infinite words. Moreover, the eager semantics yields a characterization of $\Sigma_2 \cap \mathrm{FO}^2$ (Theorem 2) while lazy rankers lead to a characterization of $\Pi_2 \cap \mathrm{FO}^2$ (Theorem 4). We note that the decidability results for the first-order fragments lead to decidability results for the respective ranker fragments [3].

It turns out that unambiguous temporal logic [4] and unambiguous interval temporal logic [5] allow natural intermediate characterizations on the way from first-order logic to rankers. In particular, this yields temporal logic counterparts of the first-order fragments. Moreover, we show that it is possible to convert formulas in unambiguous interval temporal logic into equivalent formulas in unambiguous temporal logic, without introducing new negations (Propositions 1 and 2). This also leads to a new characterization of FO^2 over finite words in terms of restricted ranker languages (Corollary 1).

Due to lack of space, most proofs are omitted. For complete proofs, we refer to the full version of this paper [1].

2 Preliminaries

In the following Γ denotes a finite alphabet. For $A \subseteq \Gamma$, we denote by A^* the set of finite words over A. The set of infinite words is A^ω and $A^\infty = A^* \cup A^\omega$ is the set of finite and infinite words. The empty word is ε and we have $\{\varepsilon\} = \emptyset^\infty$. For a word α and a position x of the word, $\alpha(x)$ is the x-th letter of α. By $|\alpha| \in \mathbb{N} \cup \{\infty\}$ we denote the *length* of α. Therefore $\alpha = \alpha(1) \cdots \alpha(|\alpha|)$ if α is finite and $\alpha = \alpha(1)\alpha(2)\cdots$ if α is infinite. We call $\mathrm{alph}(\alpha)$ the *alphabet* of α, i.e., the set of letters occurring in α. For $a \in \Gamma$, a position labeled by a is called an *a-position*. By $\mathrm{im}(\alpha)$ we mean the *imaginary* alphabet of α, i.e., the set of letters occurring infinitely often in α. For $A \subseteq \Gamma$, the set of words with imaginary alphabet A is denoted by A^{im}. In particular, $\Gamma^* = \emptyset^{\mathrm{im}}$. A *monomial* (of *degree* k) is a language of the form $A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ for letters $a_i \in \Gamma$ and sets $A_i \subseteq \Gamma$. It is *unambiguous* if each word of the monomial has a unique factorization $u_1 a_1 \cdots u_k a_k \beta$ with $u_i \in A_i^*$ and $\beta \in A_{k+1}^\infty$. A *polynomial* is a finite union of monomials. It is called *unambiguous* if it is a finite union of unambiguous monomials.

We denote by $\mathrm{FO} = \mathrm{FO}[<]$ the first-order logic over words interpreted as labeled linear orders (without ∞). As atomic formulas, FO comprises $\top$ (for *true*), the unary predicate $\lambda(x) = a$ for $a \in \Gamma$, and the binary predicate $x < y$ for variables x and y. The idea is that variables range over the linearly ordered positions of a word, and $\lambda(x) = a$ means that x is an a-position. Apart from the Boolean connectives, we allow composition of formulas using existential quantification $\exists x\colon \varphi$ and universal quantification $\forall x\colon \varphi$ for $\varphi \in \mathrm{FO}$. The semantics is as usual. Every formula in FO can be converted into a semantically equivalent formula in prenex normal form by renaming variables and moving quantifiers to the front. This observation gives rise to the fragment Σ_2 (resp. Π_2) consisting of all FO-formulas in prenex normal form with only two blocks of quantifiers, starting with a block of existential quantifiers (resp. universal quantifiers). Note that the negation of a formula in Σ_2 is equivalent to a formula in Π_2 and vice versa. The fragments Σ_2 and Π_2 are both closed under conjunction and disjunction. Furthermore, FO^2 is the fragment of FO containing all formulas which use at most two different names for the variables. This is a natural restriction, since FO with three variables already has the full expressive power of FO. A *sentence* in FO is a formula without free variables. The *language defined by* φ, denoted by $L(\varphi)$, is the set of words $\alpha \in \Gamma^\infty$ for which φ is true. We frequently identify logical fragments with the classes of languages they define (as in the definition of the fragment $\Delta_2 = \Sigma_2 \cap \Pi_2$ for example).

Example 1. Consider the formulas

$$\varphi \;=\; \exists x \forall y\colon y \leq x \vee \lambda(y) \neq a \qquad \text{and} \qquad \psi \;=\; \forall x \exists y\colon y > x \wedge \lambda(y) = a.$$

The formula $\varphi \in \Sigma_2 \cap \mathrm{FO}^2$ states that after some position there is no a-position, i.e., $L(\varphi)$ contains all words with finitely many a-positions. Its negation $\psi \in \Pi_2 \cap \mathrm{FO}^2$ says that for all positions there is a greater a-position, i.e., $L(\psi)$ is set of all words α with $a \in \mathrm{im}(\alpha)$. Surprisingly, $L(\varphi)$ is not definable in Π_2, while $L(\psi)$ is not definable in Σ_2, cf. [3]. $\Diamond$

3 Rankers and Unambiguous Temporal Logics

For finite words, rankers have been introduced by Immerman and Weis [10]. They can be seen as a generalization of *turtle programs* used by Schwentick, Thérien, and Vollmer [7] for characterizing FO^2-definable languages over finite words. The main difference between rankers and turtle programs is that rankers either uniquely determine a position in a word or they are undefined, whereas turtle programs mainly distinguish between being defined and being undefined.

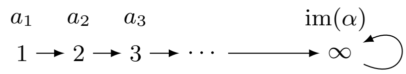

Fig. 1. Signature of $\alpha = a_1\, a_2\, a_3 \cdots$ over lazy rankers

Extending rankers with Boolean connectives yields unambiguous temporal logic (unambiguous TL). It is called *unambiguous* since each position considered by some formula in this logic is unique. Unambiguous TL has been introduced for Mazurkiewicz traces [4] which are a generalization of finite words.

All of our characterizations of first-order fragments rely on so-called unambiguous polynomials. A natural intermediate step from polynomials to temporal logic is interval temporal logic. Unambiguous interval temporal logic (unambiguous ITL) has been introduced by Lodaya, Pandya, and Shah [5] for finite words. They showed that over finite words it has the same expressive power as FO^2.

In this section, we generalize all three concepts (rankers, unambiguous TL, and unambiguous ITL) to infinite words. In fact, for each concept we shall give two natural generalizations. Surprisingly, it turns out that one of the two extensions can be used for the characterization of the first-order fragment $\Sigma_2 \cap \mathrm{FO}^2$ over Γ^∞ while the other yields a characterization of $\Pi_2 \cap \mathrm{FO}^2$. Moreover, both semantics can be used to describe FO^2 and Δ_2. In fact, for Δ_2 we use some fragment of rankers which conceals the difference between the two versions.

3.1 Rankers

A *ranker* is a finite word over the alphabet $\{\mathsf{X}_a, \mathsf{Y}_a \mid a \in \Gamma\}$. It can be interpreted as a sequence of instructions of the form X_a and Y_a. Here, X_a (for ne$\mathsf{X}$t-a) means "go to the next a-position" and Y_a (for Yesterday-a) means "go to the previous a-position". Below, we shall introduce a second variant of rankers called lazy rankers. If we want to emphasize the usage of this first version of rankers we refer to *eager rankers*. For a word α and a position $x \in \mathbb{N} \cup \{\infty\}$ we define

$$\begin{aligned}\mathsf{X}_a(\alpha, x) &= \min\{y \in \mathbb{N} \mid \alpha(y) = a \text{ and } y > x\},\\ \mathsf{Y}_a(\alpha, x) &= \max\{y \in \mathbb{N} \mid \alpha(y) = a \text{ and } y < x\}.\end{aligned}$$

As usual, we set $y < \infty$ for all $y \in \mathbb{N}$. The minimum and the maximum of $\emptyset$ as well as the maximum of an infinite set are undefined. In particular, $\mathsf{X}_a(\alpha, \infty)$ is always undefined and $\mathsf{Y}_a(\alpha, \infty)$ is defined if and only if $a \in \mathrm{alph}(\alpha) \setminus \mathrm{im}(\alpha)$. We extend this definition to rankers by setting $\mathsf{X}_a\, r(\alpha, x) = r(\alpha, \mathsf{X}_a(\alpha, x))$ and $\mathsf{Y}_a\, r(\alpha, x) = r(\alpha, \mathsf{Y}_a(\alpha, x))$, i.e., rankers are processed from left to right. If $r(\alpha, x)$ is defined for some non-empty ranker r, then $r(\alpha, x) \neq \infty$.

Next, we define another variant of rankers as finite words over the alphabet $\{\mathsf{X}^\ell_a, \mathsf{Y}^\ell_a \mid a \in \Gamma\}$. The superscript ℓ is derived from *lazy*, and such rankers are called *lazy rankers*, accordingly. The difference to eager rankers is that lazy rankers can point to an infinite position ∞. The idea is that the position ∞ is not reachable from any finite position and that it represents the behavior at infinity. We imagine that ∞ is labeled by all letters in $\mathrm{im}(\alpha)$ for words α. Therefore, it

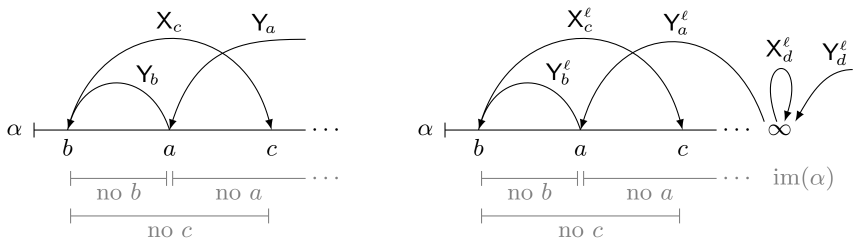

Fig. 2. An eager and a lazy ranker

is often adequate to set $\infty < \infty$, since the infinite position simulates a set of finite positions, see Fig. 1. For a word α and a finite position $x \in \mathbb{N}$ we define $\mathsf{X}^\ell_a(\alpha, x) = \mathsf{X}_a(\alpha, x)$ and $\mathsf{Y}^\ell_a(\alpha, x) = \mathsf{Y}_a(\alpha, x)$. For the infinite position we set

$$\mathsf{X}^\ell_a(\alpha, \infty) = \begin{cases} \infty & \text{if } a \in \mathrm{im}(\alpha) \\ \textit{undefined} & \text{else} \end{cases}$$

$$\mathsf{Y}^\ell_a(\alpha, \infty) = \begin{cases} \infty & \text{if } a \in \mathrm{im}(\alpha) \\ \mathsf{Y}_a(\alpha, \infty) & \text{else} \end{cases}$$

i.e., $\mathsf{Y}^\ell_a(\alpha, \infty)$ is undefined if $a \notin \mathrm{alph}(\alpha)$, and $\mathsf{Y}^\ell_a(\alpha, \infty) = \mathsf{Y}_a(\alpha, \infty)$ is a finite position if $a \in \mathrm{alph}(\alpha) \setminus \mathrm{im}(\alpha)$. As before, we extend this definition to rankers by setting $\mathsf{X}^\ell_a\, r(\alpha, x) = r(\alpha, \mathsf{X}^\ell_a(\alpha, x))$ and $\mathsf{Y}^\ell_a\, r(\alpha, x) = r(\alpha, \mathsf{Y}^\ell_a(\alpha, x))$. We denote by $\mathrm{alph}_\Gamma(r)$ the set of letters $a \in \Gamma$ such that r contains a modality using the letter a. It can happen that $r(\alpha, \infty) = \infty$ for some non-empty lazy ranker r. This is the case if and only if r is of the form $\mathsf{Y}^\ell_a\, s$ and $\mathrm{alph}_\Gamma(r) \subseteq \mathrm{im}(\alpha)$.

If the reference to the word α is clear from the context, then for eager and lazy rankers r we shorten the notation and write $r(x)$ instead of $r(\alpha, x)$.

An eager ranker r is an X-*ranker* if $r = \mathsf{X}_a\, s$ for some ranker s and $a \in \Gamma$, and it is a Y-*ranker* if r is of the form $\mathsf{Y}_a\, s$. Lazy X^ℓ-rankers and Y^ℓ-rankers are defined similarly. We proceed to define $r(\alpha)$, the position of α reached by the ranker r by starting "outside" the word α. The intuition is as follows. If r is an X-ranker or an X^ℓ-ranker, we imagine that we start at an outside position in front of α; if r is a Y-ranker or a Y^ℓ-ranker, then we start at a position behind α. Therefore, we define

$$\begin{aligned} r(\alpha) &= r(\alpha, 0) && \text{if } r \text{ is an } \mathsf{X}\text{-ranker or an } \mathsf{X}^\ell\text{-ranker}, \\ r(\alpha) &= r(\alpha, \infty) && \text{if } r \text{ is a } \mathsf{Y}\text{-ranker or a } \mathsf{Y}^\ell\text{-ranker}. \end{aligned}$$

On the left hand side of Fig. 2, a possible situation for the eager ranker $\mathsf{Y}_a\,\mathsf{Y}_b\,\mathsf{X}_c$ being defined on some word α is depicted. The right hand side of the same figure illustrates a similar situation for the lazy ranker $\mathsf{Y}^\ell_d\,\mathsf{X}^\ell_d\,\mathsf{Y}^\ell_a\,\mathsf{Y}^\ell_b\,\mathsf{X}^\ell_c$ with $d \in \mathrm{im}(\alpha)$ and $a \in \mathrm{alph}(\alpha) \setminus \mathrm{im}(\alpha)$. Note that the eager version of the same ranker is not defined on α since $d \in \mathrm{im}(\alpha)$.

For an eager or lazy ranker r the language $L(r)$ generated by r is the set of all words in Γ^∞ on which r is defined. A *(positive) ranker language* is a finite (positive) Boolean combination of languages of the form $L(r)$ for eager rankers r. A *(positive) lazy ranker language* is a finite (positive) Boolean combination of languages of the form $L(r)$ for lazy rankers r. Finally, a *(positive)* X*-ranker language* is a (positive) ranker language using only X-rankers. At the end of the next section, we extend rankers by some atomic modalities.

3.2 Unambiguous Temporal Logic

Our generalization of rankers allows us to define unambiguous temporal logic (unambiguous TL) over infinite words. As for rankers, we have an eager and a lazy variant. The syntax is given by:

$$\top \mid \neg\varphi \mid \varphi \vee \psi \mid \varphi \wedge \psi \mid \mathsf{X}_a\,\varphi \mid \mathsf{Y}_a\,\varphi \mid \mathsf{G}_{\bar{a}} \mid \mathsf{H}_{\bar{a}} \mid \mathsf{X}^\ell_a\,\varphi \mid \mathsf{Y}^\ell_a\,\varphi \mid \mathsf{G}^\ell_{\bar{a}} \mid \mathsf{H}^\ell_{\bar{a}}$$

for $a \in \Gamma$ and formulas φ, ψ in unambiguous TL. The atomic formulas are $\top$ (which is *true*), and the eager modalities $\mathsf{G}_{\bar{a}}$ (for Globally-no-a) and $\mathsf{H}_{\bar{a}}$ (for Historically-no-a), as well as the lazy modalities $\mathsf{G}^\ell_{\bar{a}}$ (for lazy-Globally-no-a) and $\mathsf{H}^\ell_{\bar{a}}$ (for lazy-Historically-no-a). We now define, when a word α with a position $x \in \mathbb{N} \cup \{\infty\}$ satisfies a formula φ in unambiguous TL, denoted by $\alpha, x \models \varphi$. The atomic formula $\top$ is true for all positions, and the semantics of the Boolean connectives is as usual. For $\mathsf{Z} \in \{\mathsf{X}_a, \mathsf{Y}_a, \mathsf{X}^\ell_a, \mathsf{Y}^\ell_a \mid a \in \Gamma\}$ we define

$$\alpha, x \models \mathsf{Z}\,\varphi \quad \text{iff} \quad \mathsf{Z}(x) \text{ is defined and } \alpha, \mathsf{Z}(x) \models \varphi.$$

The semantics of the atomic modalities is given by

$$\mathsf{G}_{\bar{a}} = \neg\,\mathsf{X}_a\,\top, \quad \mathsf{H}_{\bar{a}} = \neg\,\mathsf{Y}_a\,\top, \quad \mathsf{G}^\ell_{\bar{a}} = \neg\,\mathsf{X}^\ell_a\,\top, \quad \mathsf{H}^\ell_{\bar{a}} = \neg\,\mathsf{Y}^\ell_a\,\top.$$

In order to define when a word α models a formula φ, we have to distinguish whether φ starts with a future or with a past modality:

$$\begin{array}{llllll}
\alpha \models \mathsf{X}_a\,\varphi & \text{iff} & \alpha, 0 \models \mathsf{X}_a\,\varphi, & \alpha \models \mathsf{Y}_a\,\varphi & \text{iff} & \alpha, \infty \models \mathsf{Y}_a\,\varphi, \\
\alpha \models \mathsf{G}_{\bar{a}} & \text{iff} & \alpha, 0 \models \mathsf{G}_{\bar{a}}, & \alpha \models \mathsf{H}_{\bar{a}} & \text{iff} & \alpha, \infty \models \mathsf{H}_{\bar{a}}, \\
\alpha \models \mathsf{X}^\ell_a\,\varphi & \text{iff} & \alpha, 0 \models \mathsf{X}^\ell_a\,\varphi, & \alpha \models \mathsf{Y}^\ell_a\,\varphi & \text{iff} & \alpha, \infty \models \mathsf{Y}^\ell_a\,\varphi, \\
\alpha \models \mathsf{G}^\ell_{\bar{a}} & \text{iff} & \alpha, 0 \models \mathsf{G}^\ell_{\bar{a}}, & \alpha \models \mathsf{H}^\ell_{\bar{a}} & \text{iff} & \alpha, \infty \models \mathsf{H}^\ell_{\bar{a}}\,.
\end{array}$$

The modalities on the left are called *future modalities*, while the modalities on the right are called *past modalities*. The atomic modalities $\mathsf{G}_{\bar{a}}$ and $\mathsf{G}^\ell_{\bar{a}}$ differ only for the infinite position, but the semantics of $\mathsf{H}_{\bar{a}}$ and $\mathsf{H}^\ell_{\bar{a}}$ differs a lot: $\alpha \models \mathsf{H}_{\bar{a}}$ if and only if $a \in \mathrm{im}(\alpha)$ or $a \notin \mathrm{alph}(\alpha)$ whereas $\alpha \models \mathsf{H}^\ell_{\bar{a}}$ if and only if $a \notin \mathrm{alph}(\alpha)$. Every formula φ defines a language $L(\varphi) = \{\alpha \in \Gamma^\infty \mid \alpha \models \varphi\}$.

Finally, for $\mathcal{C} \subseteq \{\mathsf{X}_a, \mathsf{Y}_a, \mathsf{G}_{\bar{a}}, \mathsf{H}_{\bar{a}}, \mathsf{X}^\ell_a, \mathsf{Y}^\ell_a, \mathsf{G}^\ell_{\bar{a}}, \mathsf{H}^\ell_{\bar{a}}\}$ we define the following fragments of TL:

- TL$[\mathcal{C}]$ consists of all formulas using only $\top$, Boolean connectives, and temporal modalities in $\mathcal{C}$,
- TL$^+[\mathcal{C}]$ consists of all formulas using only $\top$, positive Boolean connectives (i.e., no negation), and temporal modalities in $\mathcal{C}$,

– $\mathrm{TL}_{\mathsf{X}}[\mathcal{C}]$ consists of all formulas using only $\top$, Boolean connectives, and temporal modalities in $\mathcal{C}$ such that all outmost modalities are future modalities,
– $\mathrm{TL}^+_{\mathsf{X}}[\mathcal{C}]$ consists of all formulas in $\mathrm{TL}^+[\mathcal{C}] \cap \mathrm{TL}_{\mathsf{X}}[\mathcal{C}]$.

Example 2. Consider the language $L \subseteq \Gamma^\infty$ consisting of all non-empty words with a as the first letter. This language is defined by each of following formulas:

$$\begin{aligned} \varphi_1 &= \mathsf{X}_a \top \wedge \bigwedge_{b \in \Gamma} \neg \mathsf{X}_a \mathsf{Y}_b \top && \in \mathrm{TL}_{\mathsf{X}}[\mathsf{X}_a, \mathsf{Y}_a], \\ \varphi_2 &= \bigwedge_{b \in \Gamma} \mathsf{X}_a \mathsf{H}_{\bar{b}} && \in \mathrm{TL}^+_{\mathsf{X}}[\mathsf{X}_a, \mathsf{H}_{\bar{a}}], \\ \varphi_3 &= \mathsf{X}_a \top \wedge \bigwedge_{b \in \Gamma \setminus \{a\}} \big(\mathsf{G}_{\bar{b}} \vee \mathsf{X}_b \mathsf{Y}_a \top \big) && \in \mathrm{TL}^+_{\mathsf{X}}[\mathsf{X}_a, \mathsf{Y}_a, \mathsf{G}_{\bar{a}}]. \end{aligned}$$ ◊

Inspired by the atomic logical modalities, we extend the notion of a ranker by allowing the atomic modalities $\mathsf{G}_{\bar{a}}$ and $\mathsf{H}_{\bar{a}}$ as well as $\mathsf{G}^\ell_{\bar{a}}$ and $\mathsf{H}^\ell_{\bar{a}}$. We call r a *ranker with atomic modality* $\mathsf{G}_{\bar{a}}$ ($\mathsf{H}_{\bar{a}}$, $\mathsf{G}^\ell_{\bar{a}}$, $\mathsf{H}^\ell_{\bar{a}}$, resp.) if $r = s\,\mathsf{G}_{\bar{a}}$ ($r = s\,\mathsf{H}_{\bar{a}}$, $r = s\,\mathsf{G}^\ell_{\bar{a}}$, $r = s\,\mathsf{H}^\ell_{\bar{a}}$, resp.) for some ranker s. In this setting, $r = \mathsf{G}_{\bar{a}}$ is an X-ranker, and $r = \mathsf{H}_{\bar{a}}$ is a Y-ranker. Similarly, $r = \mathsf{G}^\ell_{\bar{a}}$ is an X^ℓ-ranker, and $r = \mathsf{H}^\ell_{\bar{a}}$ is a Y^ℓ-ranker. Note that any ranker with some atomic modality is also a formula in unambiguous TL. We can therefore define the domain of an extended ranker r with some atomic modality by

$$r(\alpha, x) \text{ is defined} \quad \text{iff} \quad \alpha, x \models r.$$

If $r \in s\,\{\mathsf{G}_{\bar{a}}, \mathsf{H}_{\bar{a}}, \mathsf{G}^\ell_{\bar{a}}, \mathsf{H}^\ell_{\bar{a}} \mid a \in \Gamma\}$ is an extended ranker with $r(\alpha, x)$ being defined, then we set $r(\alpha, x) = s(\alpha, x)$, i.e., $r(\alpha, x)$ is the position reached after the execution of s. The reinterpretation of rankers as formulas also makes sense for a ranker $r \in \{\mathsf{X}_a, \mathsf{Y}_a, \mathsf{X}^\ell_a, \mathsf{Y}^\ell_a\}^*$ without atomic modality by identifying r with $r\top$ in unambiguous TL. This is justified since r is defined on α if and only if $\alpha \models r\top$.

Let $\mathcal{C} \subseteq \{\mathsf{G}_{\bar{a}}, \mathsf{H}_{\bar{a}}, \mathsf{G}^\ell_{\bar{a}}, \mathsf{H}^\ell_{\bar{a}}\}$. A language is a *ranker language with atomic modalities* $\mathcal{C}$ if it is a Boolean combination of languages $L(r)$ such that r is either a ranker without atomic modalities or a ranker with some atomic modality in $\mathcal{C}$. Similarly, the notions of lazy / positive / X-ranker languages are adapted to the use of atomic modalities.

3.3 Unambiguous Interval Temporal Logic

We extend unambiguous interval temporal logic (unambiguous ITL) to infinite words in such a way that it coincides with FO^2. Again, we have two extensions with this property, one being eager and one being lazy. The syntax of unambiguous ITL is given by Boolean combinations and:

$$\top \mid \varphi \,\mathsf{F}_a\, \psi \mid \varphi \,\mathsf{L}_a\, \psi \mid \mathsf{G}_{\bar{a}} \mid \mathsf{H}_{\bar{a}} \mid \varphi \,\mathsf{F}^\ell_a\, \psi \mid \varphi \,\mathsf{L}^\ell_a\, \psi \mid \mathsf{G}^\ell_{\bar{a}} \mid \mathsf{H}^\ell_{\bar{a}}$$

with $a \in \Gamma$ and formulas φ, ψ in unambiguous ITL. The name F_a derives from "First-a" and L_a from "Last-a". As in unambiguous temporal logic, the atomic formulas are $\top$, the eager modalities $\mathsf{G}_{\bar{a}}$ and $\mathsf{H}_{\bar{a}}$, and the lazy modalities $\mathsf{G}^{\ell}_{\bar{a}}$ and $\mathsf{H}^{\ell}_{\bar{a}}$. We now define, when a word α together with an interval $(x; y) = \{z \in \mathbb{N} \cup \{\infty\} \mid x < z < y\}$ satisfies a formula φ in unambiguous ITL, denoted by $\alpha, (x; y) \models \varphi$. Remember that we have set $\infty < \infty$. In particular $(\infty; \infty) = \{\infty\}$. The atomic formula $\top$ is true for all intervals, and the semantics of the Boolean connectives is as usual. The semantics of the binary modalities is as follows:

$$\alpha, (x; y) \models \varphi \, \mathsf{F}_a \, \psi \quad \text{iff} \quad \mathsf{X}_a(x) \text{ is defined, } \mathsf{X}_a(x) < y, \quad \alpha, \big(x; \mathsf{X}_a(x)\big) \models \varphi \text{ and } \alpha, \big(\mathsf{X}_a(x); y\big) \models \psi,$$

$$\alpha, (x; y) \models \varphi \, \mathsf{L}_a \, \psi \quad \text{iff} \quad \mathsf{Y}_a(y) \text{ is defined, } \mathsf{Y}_a(y) > x, \quad \alpha, \big(x; \mathsf{Y}_a(y)\big) \models \varphi \text{ and } \alpha, \big(\mathsf{Y}_a(y); y\big) \models \psi.$$

The semantics of F^{ℓ}_a and L^{ℓ}_a is defined analogously using X^{ℓ}_a and Y^{ℓ}_a, respectively. The semantics of the atomic modalities is given by

$$\mathsf{G}_{\bar{a}} = \neg(\top \, \mathsf{F}_a \, \top), \qquad \mathsf{H}_{\bar{a}} = \neg(\top \, \mathsf{L}_a \, \top),$$

$$\mathsf{G}^{\ell}_{\bar{a}} = \neg(\top \, \mathsf{F}^{\ell}_a \, \top), \qquad \mathsf{H}^{\ell}_{\bar{a}} = \neg(\top \, \mathsf{L}^{\ell}_a \, \top) \vee \bigvee_{b \in \Gamma} ((\top \, \mathsf{L}^{\ell}_b \, \top) \, \mathsf{F}^{\ell}_b \, \top).$$

In the definition of $\mathsf{H}^{\ell}_{\bar{a}}$, the disjunction on the right-hand side ensures that $\alpha, (\infty; \infty) \models \mathsf{H}^{\ell}_{\bar{a}}$ for every infinite word $\alpha \in \Gamma^{\omega}$ and every $a \in \Gamma$. It will turn out that the inability of specifying the letters not in $\mathrm{im}(\alpha)$ is crucial in the characterization of the fragment $\Pi_2 \cap \mathrm{FO}^2$. Observe that only for the interval $(\infty; \infty)$, there can be a b before the "first" b. Also note that for every finite interval, the formula $\mathsf{G}_{\bar{a}}$ is true if and only if $\mathsf{H}_{\bar{a}}$ is true and that $\mathsf{G}^{\ell}_{\bar{a}}$ is equivalent to $\neg(\top \, \mathsf{L}^{\ell}_a \, \top)$. Whether a word α models a formula φ in unambiguous ITL (i.e., $\alpha \models \varphi$) is defined by

$$\alpha \models \varphi \quad \text{iff} \quad \alpha, (0; \infty) \models \varphi.$$

The language defined by φ is $L(\varphi) = \{\alpha \in \Gamma^{\infty} \mid \alpha \models \varphi\}$.

Fig. 3 depicts the situation for the formula $(\varphi_1 \, \mathsf{F}_b \, \psi_1) \, \mathsf{L}_a \, (\varphi_2 \, \mathsf{F}_c \, \psi_2)$ being defined on α. The main difference to rankers and unambiguous TL is that there is no crossing over in unambiguous ITL, e.g., in the situation depicted on the left side of Fig. 2, the formula $(\top \, \mathsf{L}_b \, (\top \, \mathsf{F}_c \, \top)) \, \mathsf{L}_a \, \top$ is false even though the ranker $\mathsf{Y}_a \mathsf{Y}_b \mathsf{X}_c$ is defined.

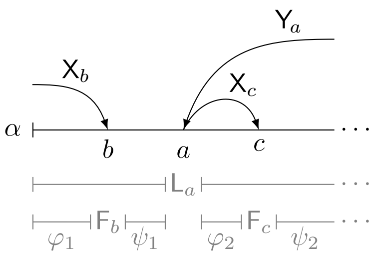

Fig. 3.

In unambiguous ITL, the modalities F_a, $\mathsf{G}_{\bar{a}}$, F^{ℓ}_a, $\mathsf{G}^{\ell}_{\bar{a}}$ are *future modalities* and L_a, $\mathsf{H}_{\bar{a}}$, L^{ℓ}_a, $\mathsf{H}^{\ell}_{\bar{a}}$ are *past modalities*. A formula φ is a *future formula* if in the parse tree of φ, every past modality occurs on the left branch of some future modality, i.e., if it is never necessary to interpret a past modality over an unbounded interval.

For $\mathcal{C} \subseteq \{\mathsf{F}_a, \mathsf{L}_a, \mathsf{G}_{\bar{a}}, \mathsf{H}_{\bar{a}}, \mathsf{F}^\ell_a, \mathsf{L}^\ell_a, \mathsf{G}^\ell_{\bar{a}}, \mathsf{H}^\ell_{\bar{a}}\}$ we define the following fragments of ITL:

- $\mathrm{ITL}[\mathcal{C}]$ consists of all formulas using only $\top$, Boolean connectives, and temporal modalities in $\mathcal{C}$,
- $\mathrm{ITL}^+[\mathcal{C}]$ consists of all formulas using only $\top$, positive Boolean connectives (i.e., no negation), and temporal modalities in $\mathcal{C}$,
- $\mathrm{ITL}_\mathsf{F}[\mathcal{C}]$ consists of all future formulas using only $\top$, Boolean connectives, and temporal modalities in $\mathcal{C}$,
- $\mathrm{ITL}^+_\mathsf{F}[\mathcal{C}]$ consists of all formulas in $\mathrm{ITL}^+[\mathcal{C}] \cap \mathrm{ITL}_\mathsf{F}[\mathcal{C}]$.

The proofs of the following two propositions give a procedure for converting unambiguous ITL formulas into unambiguous TL formulas without introducing new negations. A similar relativization technique as in our proof has been used by Lodaya, Pandya, and Shah [5] for the conversion of ITL over finite words into so-called *deterministic partially ordered two-way automata* (without the focus on not introducing negations).

Proposition 1. *We have the following inclusions:*

$$\begin{aligned}
\mathrm{ITL}[\mathsf{F}_a, \mathsf{L}_a] &\subseteq \mathrm{TL}[\mathsf{X}_a, \mathsf{Y}_a],\\
\mathrm{ITL}^+[\mathsf{F}_a, \mathsf{L}_a, \mathsf{G}_{\bar{a}}, \mathsf{H}_{\bar{a}}] &\subseteq \mathrm{TL}^+[\mathsf{X}_a, \mathsf{Y}_a, \mathsf{G}_{\bar{a}}, \mathsf{H}_{\bar{a}}],\\
\mathrm{ITL}^+[\mathsf{F}_a, \mathsf{L}_a, \mathsf{G}_{\bar{a}}] &\subseteq \mathrm{TL}^+[\mathsf{X}_a, \mathsf{Y}_a, \mathsf{G}_{\bar{a}}],\\
\mathrm{ITL}^+_\mathsf{F}[\mathsf{F}_a, \mathsf{L}_a, \mathsf{G}_{\bar{a}}, \mathsf{H}_{\bar{a}}] &\subseteq \mathrm{TL}^+_\mathsf{X}[\mathsf{X}_a, \mathsf{Y}_a, \mathsf{G}_{\bar{a}}],\\
\mathrm{ITL}_\mathsf{F}[\mathsf{F}_a, \mathsf{L}_a] &\subseteq \mathrm{TL}_\mathsf{X}[\mathsf{X}_a, \mathsf{Y}_a].
\end{aligned}$$

Proposition 2. *We have the following inclusions:*

$$\begin{aligned}
\mathrm{ITL}[\mathsf{F}^\ell_a, \mathsf{L}^\ell_a] &\subseteq \mathrm{TL}[\mathsf{X}^\ell_a, \mathsf{Y}^\ell_a],\\
\mathrm{ITL}^+[\mathsf{F}^\ell_a, \mathsf{L}^\ell_a, \mathsf{G}^\ell_{\bar{a}}, \mathsf{H}^\ell_{\bar{a}}] &\subseteq \mathrm{TL}^+[\mathsf{X}^\ell_a, \mathsf{Y}^\ell_a, \mathsf{G}^\ell_{\bar{a}}, \mathsf{H}^\ell_{\bar{a}}],\\
\mathrm{ITL}^+[\mathsf{F}^\ell_a, \mathsf{L}^\ell_a, \mathsf{H}^\ell_{\bar{a}}] &\subseteq \mathrm{TL}^+[\mathsf{X}^\ell_a, \mathsf{Y}^\ell_a, \mathsf{H}^\ell_{\bar{a}}].
\end{aligned}$$

4 Main results

We start this section with various ITL, TL, and ranker characterizations using the eager variants. We postpone characterizations in terms of the lazy fragments to Theorem 4 and Theorem 5.

Theorem 1. *For $L \subseteq \Gamma^\infty$ the following assertions are equivalent:*

1. *L is definable in* FO^2.
2. *L is definable in* $\mathrm{ITL}^+[\mathsf{F}_a, \mathsf{L}_a, \mathsf{G}_{\bar{a}}, \mathsf{H}_{\bar{a}}]$.
3. *L is definable in* $\mathrm{ITL}[\mathsf{F}_a, \mathsf{L}_a]$.
4. *L is definable in* $\mathrm{TL}[\mathsf{X}_a, \mathsf{Y}_a]$.
5. *L is definable in* $\mathrm{TL}^+[\mathsf{X}_a, \mathsf{Y}_a, \mathsf{G}_{\bar{a}}, \mathsf{H}_{\bar{a}}]$.
6. *L is a positive ranker language with atomic modalities $\mathsf{G}_{\bar{a}}$ and $\mathsf{H}_{\bar{a}}$.*
7. *L is a ranker language.*

Every FO^2-definable language is a Boolean combination of unambiguous monomials and languages of the form A^{im}, see [3]. The language A^{im} is definable by the formula

$$\bigwedge_{a \in A} (\top \mathsf{F}_a \top) \wedge \mathsf{H}_{\bar{a}} \quad \in \mathrm{ITL}^+[\mathsf{F}_a, \mathsf{H}_{\bar{a}}].$$

Hence, the following lemma provides the missing part in order to show that every language in FO^2 is definable in unambiguous ITL.

Lemma 1. *Every unambiguous monomial $L = A_1^* a_1 \cdots A_k^* a_k A_{k+1}^\infty$ is definable in $\mathrm{ITL}^+[\mathsf{F}_a, \mathsf{L}_a, \mathsf{G}_{\bar{a}}]$.*

Proof. We perform an induction on k. For $k = 0$ we have $L\big(\bigwedge_{a \notin A_1} \mathsf{G}_{\bar{a}}\big) = A_1^\infty$. Let $k \geq 1$. Since L is unambiguous, we have $\{a_1, \ldots, a_k\} \not\subseteq A_1 \cap A_{k+1}$; otherwise $(a_1 \cdots a_k)^2$ admits two different factorizations showing that L is not unambiguous. First, consider the case $a_i \notin A_1$ and let i be minimal with this property. Each word $\alpha \in L$ has a unique factorization $\alpha = u a_i \beta$ such that $a_i \notin \mathrm{alph}(u)$. Depending on whether the first a_i of α coincides with the marker a_i or not, we have

$$\begin{array}{lll} u \in A_1^* a_1 \cdots A_i^*, & & \beta \in A_{i+1}^* a_{i+1} \cdots A_k^* a_k A_{k+1}^\infty \quad \text{or} \\ u \in A_1^* a_1 \cdots A_j^*, & a_i \in A_j, & \beta \in A_j^* a_j \cdots A_k^* a_k A_{k+1}^\infty \end{array}$$

with $2 \leq j \leq i$. In both cases, since L is unambiguous, each expression containing u or β is unambiguous. Moreover, each of these expressions is strictly shorter than L. By induction, for each $2 \leq j \leq k$, there exist formulas $\varphi, \psi \in \mathrm{ITL}^+[\mathsf{F}_a, \mathsf{L}_a, \mathsf{G}_{\bar{a}}]$ such that $L(\varphi) = A_1^* a_1 \cdots A_j^\infty$ and $L(\psi) = A_j^* a_j \cdots A_k^* a_k A_{k+1}^\infty$. By the above reasoning, we see that L is the union of (at most i) languages of the form

$$\big(L(\varphi) \cap (\Gamma \setminus \{a_i\})^*\big)\, a_i\, L(\psi)$$

and each of them is defined by $\varphi \, \mathsf{F}_{a_i} \, \psi$.

For $a_i \notin A_{k+1}$ with i maximal, we consider the unique factorization $\alpha = u a_i \beta$ with $a_i \notin \mathrm{alph}(\beta)$ and, again, we end up with one of the two cases from above, with the difference that $1 \leq i < j \leq k$ in the second case. Inductively L is defined by a disjunction of formulas $\varphi \, \mathsf{L}_{a_i} \, \psi$. ☐

Theorem 2. *Let $L \subseteq \Gamma^\infty$. The following assertions are equivalent:*

1. *L is definable in Σ_2 and FO^2.*
2. *L is definable in $\mathrm{ITL}^+[\mathsf{F}_a, \mathsf{L}_a, \mathsf{G}_{\bar{a}}]$.*
3. *L is definable in $\mathrm{TL}^+[\mathsf{X}_a, \mathsf{Y}_a, \mathsf{G}_{\bar{a}}]$.*
4. *L is a positive ranker language with atomic modality $\mathsf{G}_{\bar{a}}$.*

Theorem 5 shows that the same characterizations for FO^2 hold using the lazy variants. Note that we cannot use lazy counterparts in the characterizations for $\Sigma_2 \cap \mathrm{FO}^2$, since for example $\mathsf{Y}_a^\ell \, \mathsf{X}_a^\ell$ is defined if and only if there are infinitely many a's, but this property is not Σ_2-definable.

Over finite words, the fragments FO^2 and Δ_2 coincide [9]. In particular, $\mathrm{FO}^2 \cap \Sigma_2 = \mathrm{FO}^2$ over finite words. Since finiteness of a word is definable in $\mathrm{FO}^2 \cap \Sigma_2$, we obtain the following corollary of Theorem 2.

Corollary 1. *A language $L \subseteq \Gamma^*$ of finite words is definable in* FO^2 *if and only if L is a positive ranker language with atomic modality* $\mathsf{G}_{\bar{a}}$.

Over infinite words, the fragment Δ_2 is a strict subclass of FO^2. The next theorem says that Δ_2 is basically FO^2 with the lack of past formulas and Y-rankers. Since eager future formulas and X-rankers coincide with their lazy counterparts, all of the characterizations in the next theorem could be replaced by their lazy pendants.

Theorem 3. *Let $L \subseteq \Gamma^\infty$. The following assertions are equivalent:*

1. *L is definable in Δ_2.*
2. *L is definable in* $\mathrm{ITL}_{\mathsf{F}}^+[\mathsf{F}_a, \mathsf{L}_a, \mathsf{G}_{\bar{a}}]$.
3. *L is definable in* $\mathrm{ITL}_{\mathsf{F}}[\mathsf{F}_a, \mathsf{L}_a]$.
4. *L is definable in* $\mathrm{TL}_{\mathsf{X}}[\mathsf{X}_a, \mathsf{Y}_a]$.
5. *L is definable in* $\mathrm{TL}_{\mathsf{X}}^+[\mathsf{X}_a, \mathsf{Y}_a, \mathsf{G}_{\bar{a}}]$.
6. *L is a positive* X*-ranker language with atomic modality* $\mathsf{G}_{\bar{a}}$.
7. *L is an* X*-ranker language.*

In the next theorem we give characterizations of the fragment $\Pi_2 \cap \mathrm{FO}^2$ in terms of the lazy variants of ITL, TL, and rankers. We cannot use the eager variants, since Y_a says that there are only finitely many a's, but this property is not Π_2-definable. Also note that $\alpha, (\infty;\infty) \models \hat{\mathsf{H}}_{\bar{a}}$ for $\hat{\mathsf{H}}_{\bar{a}} = \neg(\top \, \mathsf{L}_a^\ell \, \top)$ if and only if $a \notin \mathrm{im}(\alpha)$, i.e., if and only if a occurs at most finitely often. As before, this property is not Π_2-definable. This is the reason why we did not define $\mathsf{H}_{\bar{a}}^\ell$ simply as $\hat{\mathsf{H}}_{\bar{a}}$.

Theorem 4. *Let $L \subseteq \Gamma^\infty$. The following assertions are equivalent:*

1. *L is definable in Π_2 and* FO^2.
2. *L is definable in* $\mathrm{ITL}^+[\mathsf{F}_a^\ell, \mathsf{L}_a^\ell, \mathsf{H}_{\bar{a}}^\ell]$.
3. *L is definable in* $\mathrm{TL}^+[\mathsf{X}_a^\ell, \mathsf{Y}_a^\ell, \mathsf{H}_{\bar{a}}^\ell]$.
4. *L is a positive lazy ranker language with atomic modality* $\mathsf{H}_{\bar{a}}^\ell$.

For completeness, we give a counterpart of Theorem 1 using the lazy versions of ITL, TL, and rankers.

Theorem 5. *For $L \subseteq \Gamma^\infty$ the following assertions are equivalent:*

1. *L is definable in* FO^2.
2. *L is definable in* $\mathrm{ITL}^+[\mathsf{F}_a^\ell, \mathsf{L}_a^\ell, \mathsf{G}_{\bar{a}}^\ell, \mathsf{H}_{\bar{a}}^\ell]$.
3. *L is definable in* $\mathrm{ITL}[\mathsf{F}_a^\ell, \mathsf{L}_a^\ell]$.
4. *L is definable in* $\mathrm{TL}[\mathsf{X}_a^\ell, \mathsf{Y}_a^\ell]$.
5. *L is definable in* $\mathrm{TL}^+[\mathsf{X}_a^\ell, \mathsf{Y}_a^\ell, \mathsf{G}_{\bar{a}}^\ell, \mathsf{H}_{\bar{a}}^\ell]$.
6. *L is a positive ranker language with atomic modalities* $\mathsf{G}_{\bar{a}}^\ell$ *and* $\mathsf{H}_{\bar{a}}^\ell$.
7. *L is a lazy ranker language.*

5 Open Problems

Rankers over finite words have been introduced for characterizing quantifier alternation within FO^2. We conjecture that similar results for infinite words can be obtained using our generalizations of rankers.

Over infinite words, the class of X-ranker languages corresponds to the fragment Δ_2. Over finite words however, X-ranker languages form a strict subclass of Δ_2 (which for finite words coincides with FO^2). An algebraic counterpart of X-ranker languages over finite words is still missing. The main problem is that over finite words X-rankers do not define a variety of languages.

A well-known theorem by Schützenberger [6] implies that over finite words, arbitrary finite unions of unambiguous monomials and finite *disjoint* unions of unambiguous monomials describe the same class of languages. In the case of infinite words, it is open whether one can require that unambiguous polynomials are disjoint unions of unambiguous monomials without changing the class of languages.

Acknowledgments. We thank Volker Diekert for a suggestion which led to Theorem 2. We also thank the anonymous referees for several useful suggestions which helped to improve the presentation of this paper.

References

1. Dartois, L., Kufleitner, M., Lauser, A.: Rankers over Infinite Words. Technical report no. 2010/01, Universität Stuttgart, Informatik (2010)
2. Diekert, V., Gastin, P., Kufleitner, M.: A survey on small fragments of first-order logic over finite words. International Journal of Foundations of Computer Science 19(3), 513–548 (2008), special issue DLT 2007
3. Diekert, V., Kufleitner, M.: Fragments of first-order logic over infinite words. STACS 2009. Dagstuhl Seminar Proceedings, vol. 09001, pp. 325–336 (2009)
4. Kufleitner, M.: Polynomials, fragments of temporal logic and the variety DA over traces. Theoretical Computer Science 376, 89–100 (2007), special issue DLT 2006
5. Lodaya, K., Pandya, P.K., Shah, S.S.: Marking the chops: an unambiguous temporal logic. IFIP TCS 273, 461–476 (2008)
6. Schützenberger, M.P.: Sur le produit de concaténation non ambigu. Semigroup Forum 13, 47–75 (1976)
7. Schwentick, Th., Thérien, D., Vollmer, H.: Partially-ordered two-way automata: A new characterization of DA. Proceedings of DLT'01, LNCS vol. 2295, pp. 239–250. Springer (2001)
8. Tesson, P., Thérien, D.: Diamonds are forever: The variety DA. Semigroups, Algorithms, Automata and Languages, 475–500 (2002)
9. Thérien, D., Wilke, Th.: Over words, two variables are as powerful as one quantifier alternation. Proceedings of STOC'98, 234-240 (1998). (1998)
10. Weis, Ph., Immerman, N.: Structure theorem and strict alternation hierarchy for FO^2 on words. Logical Methods in Computer Science 5(3:3), 1–23 (2009)