INSTITUT FÜR INFORMATIK
Lehr- und Forschungseinheit für
Programmier- und Modellierungssprachen
Oettingenstraße 67, D–80538 München

Ludwig—
Maximilians—

Symmetry in XPath

Dan Olteanu, Holger Meuss, Tim Furche, François Bry

Tehnical Report, Computer Science Institute, Munich, Germany
http://www.pms.informatik.uni-muenchen.de/publikationen
Forschungsbericht/Research Report PMS-FB-2001-16, October 2001
Revised in Research Report PMS-FB-2001-17 "XPath: Looking Forward"

Symmetry in XPath

Dan Olteanu, Holger Meuss, Tim Furche, François Bry

Insitute for Computer Science and Center for Information and Language Processing
University of Munich, Germany
olteanu@informatik.uni-muenchen.de

Abstract

The location path language XPath is of particular importance for XML applications since it is a core component of many XML processing standards such as XSLT or XQuery. In this paper, based on axis symmetry of XPath, equivalences of XPath 1.0 location paths involving "reverse axes", such as `ancestor` and `preceding`, are established. These equivalences are used as rewriting rules in an algorithm for transforming location paths with reverse axes into equivalent reverse-axis-free ones. Location paths without reverse axes, as generated by the presented rewriting algorithm, enable efficient SAX-like streamed data processing of XPath.

1 Introduction

Query languages for XML and semistructured data rely on location paths for selecting nodes in data items. In particular, XQuery [21] and XSLT [18] are based on XPath [17]. XPath takes a navigational approach for specifying the nodes to be selected, hence a large number of navigational axes (e.g. `child`, `descendant`, `preceding`) have been defined in XPath. The number as well as the relevance of these navigational axes for querying XML has been challenged in [8, 21, 23].

The random access to XML data that is enabled by the various navigational axes of XPath has proven particularly difficult for an efficient stream-based processing of XPath queries. Processing of XML has seen the widespread use of the W3C document object model (DOM) [19], where an in-memory representation of the entire XML data is used. As DOM has been developed with focus on document processing in user agents (e.g. browsers), this approach has several shortcomings for other application areas:

First, a considerable amount of XML applications, in particular data-centric applications, handle documents too large to be processed in memory. Such documents are often encountered in natural language processing [11], in biological [13] and astronomical [1] projects.

Second, the need for progressive processing (also referred to as sequential processing) of XML has emerged: Stream-based processing generating partial results as soon as they are available gives rise to a more efficient evaluation in certain contexts, e.g.:

- For selective dissemination of information (SDI), documents have to be filtered according to complex requirements specified as XPath queries before being distributed to the subscribers [7, 4]. The routing of data to selected receivers is also becoming increasingly important in the context of web service architectures.
- To integrate data over the Internet, in particular from slow sources, it is desirable to progressively process the input before the full data is retrieved [14, 10].
- As a general processing scheme for XML, several solutions for pipelined processing have been suggested, where the input is sent through a chain of processors each of which takes the output of the preceding processor as input, e.g. Apache Cocoon [2] and XPipe [15].

- Progressive rendering of large documents, e.g. by means of XSL(T) (cf. Requirement 19 of [23]). There have been several attempts to solve this problem [3].

There is a great interest in the identification of a subset of XPath that allows efficient progressive or stream-based processing (cf. [8] and Requirement 19 of [23]).

For stream-based processing of XML data, the Simple API for XML (SAX) [16] has been specified. Of particular concern for progressive SAX-like processing are the reverse axes of XPath, i.e. those navigational axes (e.g. `parent`, `preceding`) that select nodes occuring before the context node in document order. A restriction to forward axes (i.e. axes selecting only nodes after the context node) in location paths is a straightforward consideration for an efficient stream-based evaluation of XPath queries [8].

There are three principal options how to evaluate reverse axes in a stream-based context:

- Storing in memory sufficient information that allows to access past events when evaluating a reverse axis. This amounts to keeping in memory a (possibly pruned) DOM representation of the data [3].

- Evaluating an XPath expression in more than one run. With this approach, it is also necessary to store additional information to be used in successive runs. This information can be considerably smaller than what is needed in the first approach.

- Replacing XPath expressions by equivalent ones without reverse axes.

In this paper it is shown that the third approach is possible. It is less time consuming than the second approach and does not require the in-memory storage of fragments of the input as the first approach does. Hence, XPath can be evaluated without restriction on the use of reverse axes.

The notion of equivalence between location paths is defined in Section 3 using a formal model and a denotational semantics for XPath based on [24, 25] introduced in Section 2. This model and semantics also provide with a convenient formal framework for proving the above mentioned equivalences. Two sets of equivalences (with rather different properties) are established in Section 3. These equivalences are used as rewriting rules in an algorithm, called "rare", for transforming absolute XPath location paths with reverse axes into equivalent reverse-axis-free ones (Section 4). Two rewritings, based on the two rule sets, are considered. Section 5 is devoted to those location paths for which the "rare" algorithms do not give reverse-axis-free equivalent paths. It is shown how these paths can be rewritten to equivalent XPath 2.0 expressions without reverse axes. In Section 6, related work is discussed. Section 7 is a conclusion. Proofs for the equivalences can be found in the appendix.

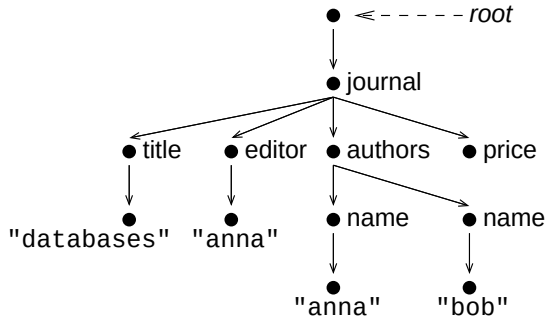
Familiarity with XPath 1.0 [17] is assumed. For section 5 some knowledge of XPath 2.0 [20] is desirable.

2 Preliminaries

In this paper, specificities of XML that are irrelevant to the issue of concern are left out. Thus, namespaces, comments, processing instructions, attributes, attribute values, document collections, schema types, references, and white space processing are not considered. The results given in this paper extend straightforwardly to unrestricted XML documents.

Section 2.1 defines the syntax of the location path language used throughout the paper. The mathematical model described in Section 2.2 is adapted from [24, 25]. It consists of a few mathematical functions that can be seen as (formal specifications of) elementary procedures. In Section 2.3 they are used in defining two so-called semantics functions that can be seen as (a formal specification of) an implementation of XPath.

Figure 1 Tree and XML data it represents according to Section 2.2




```
<journal>
  <title>databases</title>
  <editor>anna</editor>
  <authors>
    <name>anna</name>
    <name>bob</name>
  </authors>
  <price />
</journal>
```

2.1 Location Path Language

The location path language considered in the following is unabbreviated XPath without those constructs (such as those needed for processing attributes) irrelevant to the formal model introduced above. For convenience, this language will be referred to as xPath. Recall that every abbreviated XPath expression can easily be translated into an unabbreviated XPath expression. It is worth stressing that the results given below for xPath easily extend to XPath 1.0 [17]. The (abstract) syntax of xPath is as follows:

$$
\begin{aligned}
\mathit{path} &::= \mathit{path}\ \texttt{|}\ \mathit{path} \mid \texttt{/}\ \mathit{path} \mid \mathit{path}\ \texttt{/}\ \mathit{path} \mid \mathit{path}\ \texttt{[}\ \mathit{qualif}\ \texttt{]} \mid \mathit{axis}\ \texttt{::}\ \mathit{nodetest} \mid \bot\ .\\
\mathit{qualif} &::= \mathit{qualif}\ \texttt{and}\ \mathit{qualif} \mid \mathit{qualif}\ \texttt{or}\ \mathit{qualif} \mid \texttt{not(}\ \mathit{qualif}\ \texttt{)} \mid \texttt{(}\ \mathit{qualif}\ \texttt{)} \mid\\
&\quad \mathit{path}\ \texttt{=}\ \mathit{path} \mid \mathit{path}\ \texttt{==}\ \mathit{path} \mid \mathit{path}\ .\\
\mathit{axis} &::= \mathit{reverse\_axis} \mid \mathit{forward\_axis}\ .\\
\mathit{reverse\_axis} &::= \texttt{parent} \mid \texttt{ancestor} \mid \texttt{ancestor-or-self} \mid\\
&\quad \texttt{preceding} \mid \texttt{preceding-sibling}\ .\\
\mathit{forward\_axis} &::= \texttt{self} \mid \texttt{child} \mid \texttt{descendant} \mid \texttt{descendant-or-self} \mid\\
&\quad \texttt{following} \mid \texttt{following-sibling}\ .\\
\mathit{nodetest} &::= \mathit{tagname} \mid \texttt{*} \mid \texttt{text()} \mid \texttt{node()}\ .
\end{aligned}
$$

As XPath 2.0, and in contrast to XPath 1.0, xPath allows the union p_1 `|` p_2 of two paths at every level. Such paths can easily be transformed into paths with unions at top level only. Note also that while we do not consider functions in the following sections, the results almost immediately apply to location paths with functions. The only class of functions that needs special treatment are functions for accessing the context position or size of a node. If the order of the resulting node set of an expression is changed by a rewriting, the predicates containing position functions have to be readjusted accordingly. This is easy to achieve, as there are only two orders (document order and reverse document order) for node sets in XPath 1.0.

$\bot$ is convenient for simplifying proofs. It is used as a canonical equivalent path (a notion formally defined in Section 3) to the xPath expressions that select no nodes whatever the current node and document are, e.g. `/parent::*`.

p_1 `==` p_2 expresses node equality based on identity. The semantics of `==` is formally defined in the next section. Thus, if p_1 and p_2 are two paths, then p_1 `==` p_2 holds if there is a node selected by p_1 which is *identical* to a node selected by p_2. `==` corresponds to the built-in node equality operator (`==`) in XPath 2.0 and XQuery 1.0, but it can also be used for comparing node sets similar to general comparisons in XPath 2.0. As XPath 1.0 has built-in support for equality based on node values only, the XPath 1.0 expression `count(`p_1 `|` p_2`) < count(`p_1`) + count(`p_2`)` can be used for expressing `==`.

A *path* expression will be called a "location path", or "path" for short. A *qualif* expression is a "qualifier" (or pattern). Expressions *axis*`::`*nodetest* and *axis*`::`*nodetest*`[`*qualif*`]` are "steps",

also called "location steps". The length of a location path is the number of location steps it contains outside and inside qualifiers. Note that every location path is a qualifier, but the converse is false.

Absolute location paths are recursively defined as follows: A disjunctive path, i.e. a path of the form $p_1 \mid \ldots \mid p_i \mid \ldots \mid p_k$, is an absolute path if for all $i = 1, \ldots, k$, p_i is an absolute path. A non-disjunctive path is an absolute path if it is of the form `/`p, where p is a path. A location path, which is not an absolute path, is a "relative path". A step is a "forward step", if its axis is a forward axis, or a "reverse step", if its axis is a reverse axis.

The axes of the following pairs are "symmetrical" of each other: `parent` – `child`, `ancestor` – `descendant`, `descendant-or-self` – `ancestor-or-self`, `preceding` – `following`, `preceding-sibling` – `following-sibling`, and (useful in proofs) `self` – `self`.

2.2 Formal Model

There is a basic data type (i.e. a set) Node. An element of Node, a node, is one of the three kinds "root", "element", or "text". There are three functions of type Node → Boolean: *isRoot*, *isElement*, and *isText*. Applied to a same node, exactly one of them returns true.

The root node of a document corresponds to the document node of DOM and of the XQuery 1.0 and XPath 2.0 Data Model [22] – i.e. it is none of the document elements. A leaf is an empty element or a text node – cf. Figure 1.

The following three (standard) functions relate nodes in the considered document: *children* : Node → Set(Node), *parent* : Node → Set_1(Node), and *root* : Node → Node. Set(T) ($\text{Set}_1(T)$, resp.) denotes the type of sets (singleton sets, resp.) of elements of type T. *children* and *parent* are defined as their names suggest; *root*(x) is the root of the document in which x occurs. These functions can be seen as specifications of basic services that the storage system and/or document parser might provide.

A binary relation on two types T_1 and T_2, i.e. a subset of $T_1 \times T_2$, can be represented as a function of type $T_1 \to \text{Set}(T_2)$. If f is such a function, f^0 denotes the function $x \to \{x\}$, f^1 denotes f itself, and f^+ and f^* are defined as usual by:

$$\begin{aligned} f^{n+1}(x) &= \{z \mid y \in f^n(x) \wedge z \in f(y)\} \text{ for } n \in N \\ f^+(x) &= \cup_{n \in \mathbb{N} \setminus \{0\}} f^n(x) \\ f^*(x) &= \cup_{n \in \mathbb{N}} f^n(x) \end{aligned}$$

Two functions *sibling* and *branch-sibling* of type Node → Set(Node) are defined by:

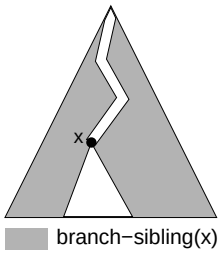
$$\begin{aligned} \textit{sibling}(x) &= \{x_1 | x_2 \in \textit{parent}(x) \wedge x_1 \in \textit{children}(x_2)\} \\ \textit{branch-sibling}(x) &= \{x_1 | x_1 \in \textit{children}^+(\textit{root}(x)) \wedge x_1 \notin \textit{children}^+(x) \wedge \\ &\qquad x_1 \notin \textit{parent}^+(x) \wedge x_1 \neq x\} \end{aligned}$$

The (non-standard) function *branch-sibling* is illustrated by Figure 2. Note the following:

$$\begin{aligned} x \in \textit{sibling}(y) \text{ iff } y \in \textit{sibling}(x) \\ x \in \textit{branch-sibling}(y) \text{ iff } y \in \textit{branch-sibling}(x) \end{aligned}$$

The (infix) function $\ll$ of type Node × Node → Boolean denotes the "document order": An element node n_1 $\ll$-precedes a node n_2 if its start tag precedes the representation of n_2 in the linear document (i.e. XML source), a text node n_1 $\ll$-precedes a node n_2, if its text precedes the representation of n_2 in the linear document.

Figure 2 Branch-Siblings



Using *branch-sibling*, the following functions of type Node → Set(Node) have simple definitions:

$$\begin{aligned} \mathit{preceding}(x) &= \{x_1 | x_1 \in \mathit{branch\text{-}sibling}(x) \wedge x_1 \ll x\} \\ \mathit{following}(x) &= \{x_1 | x_1 \in \mathit{branch\text{-}sibling}(x) \wedge x \ll x_1\} \\ \mathit{preceding\text{-}sibling}(x) &= \{x_1 | x_1 \in \mathit{sibling}(x) \wedge x_1 \ll x\} \\ \mathit{following\text{-}sibling}(x) &= \{x_1 | x_1 \in \mathit{sibling}(x) \wedge x \ll x_1\} \end{aligned}$$

Note the following (immediate) properties:

$$\begin{aligned} x \in \mathit{preceding}(y) \text{ iff } y \in \mathit{following}(x) \\ x \in \mathit{preceding\text{-}sibling}(y) \text{ iff } y \in \mathit{following\text{-}sibling}(x) \end{aligned}$$

A function *name* of type Node → String assigns to each element node its tag name.

Let NodeTest be the set (of node tests) containing all tag names and the strings (XPath constructs) `*`, `node()`, and `text()`. The function *nodetest* of type Node × NodeTest → Boolean will be used, as shown in Table 1.

Finally, the function *value* assigns as follows a string to a node x. If x is a text node, then $\mathit{value}(x)$ is the text attached to x. Otherwise (x is a root or element node) $\mathit{value}(x)$ is the concatenation of the text of the values of all children of x in document order.

Table 1 gives an overview of the mathematical functions referred to in this paper.

2.3 Denotational Semantics

The denotational semantics of xPath is specified by two functions $\mathcal{S}$ and $\mathcal{Q}$. $\mathcal{S}$ has two arguments, a location path and a node. $\mathcal{Q}$ has also two arguments, a qualifier and a node. Following a common practice, $\mathcal{S}$ and $\mathcal{Q}$ are curried: Instead of writing $\mathcal{S}(\mathit{path}, x)$ and $\mathcal{Q}(\mathit{qualif}, x)$ one writes $\mathcal{S}[\![\mathit{path}]\!]x$ and $\mathcal{Q}[\![\mathit{qualif}]\!]x$. The brackets $[\![$ and $]\!]$ are used to stress that the semantics of the expressions they enclose is defined.

$\mathcal{S}[\![\mathit{path}]\!]x$ denotes the set of all nodes selected by the location path *path* when the current node is x. Note, that in the rest of this paper the order attached to a node set selected by a location path is ignored. $\mathcal{Q}[\![\mathit{qualif}]\!]x$ is true if the qualifier *qualif* holds for (current) node x.

$\mathcal{S}$ and $\mathcal{Q}$ are defined below by recursion on the structure of xPath expressions (defined in Section 2.1) referring to the functions introduced in Section 2.2. The following definitions are adapted from [24, 25], so as to make the proofs of this paper easier. Cf. the cited papers for more information.

Table 1

$$
\begin{array}{l}
\mathit{nodetest} : \text{Node} \times \text{NodeTest} \to \text{Boolean} \\
\mathit{nodetest}(x, n) = \begin{cases} \mathit{isElement}(x) \wedge \mathit{name}(x) = n & \text{for } n \text{ tag name} \\ \mathit{isElement}(x) & \text{for } n = \texttt{*} \\ \text{true} & \text{for } n = \texttt{node()} \\ \mathit{isText}(x) & \text{for } n = \texttt{text()} \end{cases} \\
\mathit{branch\text{-}sibling} : \text{Node} \to \text{Set(Node)} \\
\mathit{branch\text{-}sibling}(x) = \{x_1 | x_1 \in \mathit{children}^+(\mathit{root}(x)) \wedge x_1 \notin \mathit{children}^+(x) \wedge \\
\qquad\qquad x_1 \notin \mathit{parent}^+(x) \wedge x_1 \neq x\} \\
\mathit{sibling} : \text{Node} \to \text{Set(Node)} \\
\mathit{sibling}(x) = \{x_1 | x_2 \in \mathit{parent}(x) \wedge x_1 \in \mathit{children}(x_2)\} \\
\\
\mathit{preceding} : \text{Node} \to \text{Set(Node)} \\
\mathit{preceding}(x) = \{x_1 | x_1 \in \mathit{branch\text{-}sibling}(x) \wedge x_1 \ll x\} \\
\mathit{preceding\text{-}sibling} : \text{Node} \to \text{Set(Node)} \\
\mathit{preceding\text{-}sibling}(x) = \{x_1 | x_1 \in \mathit{sibling}(x) \wedge x_1 \ll x\} \\
\mathit{following} : \text{Node} \to \text{Set(Node)} \\
\mathit{following}(x) = \{x_1 | x_1 \in \mathit{branch\text{-}sibling}(x) \wedge x \ll x_1\} \\
\mathit{following\text{-}sibling} : \text{Node} \to \text{Set(Node)} \\
\mathit{following\text{-}sibling}(x) = \{x_1 | x_1 \in \mathit{sibling}(x) \wedge x \ll x_1\}
\end{array}
$$

$$
\begin{array}{ll}
\mathcal{S} : \text{Pattern} \to \text{Node} \to \text{Set(Node)} & \\
\mathcal{S}[\![p_1 \mid p_2]\!]x & = \mathcal{S}[\![p_1]\!]x \cup \mathcal{S}[\![p_2]\!]x \\
\mathcal{S}[\![/p]\!]x & = \mathcal{S}[\![p]\!](\mathit{root}(x)) \\
\mathcal{S}[\![p_1/p_2]\!]x & = \{x_2 | x_1 \in \mathcal{S}[\![p_1]\!]x \wedge x_2 \in \mathcal{S}[\![p_2]\!]x_1\} \\
\mathcal{S}[\![p\texttt{[}q\texttt{]}]\!]x & = \{x_1 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathcal{Q}[\![q]\!]x_1\} \\
\mathcal{S}[\![\texttt{(}p\texttt{)}]\!]x & = \mathcal{S}[\![p]\!]x \\
\mathcal{S}[\![\texttt{self::}n]\!]x & = \{x_1 | x_1 = x \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\texttt{child::}n]\!]x & = \{x_1 | x_1 \in \mathit{children}(x) \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\texttt{parent::}n]\!]x & = \{x_1 | x_1 \in \mathit{parent}(x) \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\texttt{descendant::}n]\!]x & = \{x_1 | x_1 \in \mathit{children}^+(x) \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\texttt{descendant-or-self::}n]\!]x & = \{x_1 | x_1 \in \mathit{children}^*(x) \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\texttt{ancestor::}n]\!]x & = \{x_1 | x_1 \in \mathit{parent}^+(x) \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\texttt{ancestor-or-self::}n]\!]x & = \{x_1 | x_1 \in \mathit{parent}^*(x) \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\texttt{preceding::}n]\!]x & = \{x_1 | x_1 \in \mathit{preceding}(x) \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\texttt{preceding-sibling::}n]\!]x & = \{x_1 | x_1 \in \mathit{preceding\text{-}sibling}(x) \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\texttt{following::}n]\!]x & = \{x_1 | x_1 \in \mathit{following}(x) \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\texttt{following-sibling::}n]\!]x & = \{x_1 | x_1 \in \mathit{following\text{-}sibling}(x) \wedge \mathit{nodetest}(x_1, n)\} \\
\mathcal{S}[\![\bot]\!]x & = \emptyset
\end{array}
$$

$\mathcal{Q}$: Qualifier → Node → Boolean

$$
\begin{aligned}
\mathcal{Q}[\![q_1 \texttt{ and } q_2]\!]x &= \mathcal{Q}[\![q_1]\!]x \wedge \mathcal{Q}[\![q_2]\!]x \\
\mathcal{Q}[\![q_1 \texttt{ or } q_2]\!]x &= \mathcal{Q}[\![q_1]\!]x \vee \mathcal{Q}[\![q_2]\!]x \\
\mathcal{Q}[\![\texttt{not(}q\texttt{)}]\!]x &= \neg\mathcal{Q}[\![q]\!]x \\
\mathcal{Q}[\![\texttt{(}q\texttt{)}]\!]x &= \mathcal{Q}[\![q]\!]x \\
\mathcal{Q}[\![p]\!]x &= \mathcal{S}[\![p]\!]x \neq \emptyset \\
\mathcal{Q}[\![p_1 \texttt{ = } p_2]\!]x &= \{x_1 | x_1 \in \mathcal{S}[\![p_1]\!]x \wedge x_2 \in \mathcal{S}[\![p_2]\!]x \wedge \\
&\qquad\qquad value(x_1) = value(x_2)\} \neq \emptyset \\
\mathcal{Q}[\![p_1 \texttt{ == } p_2]\!]x &= \{x_1 | x_1 \in \mathcal{S}[\![p_1]\!]x \wedge x_1 \in \mathcal{S}[\![p_2]\!]x\} \neq \emptyset
\end{aligned}
$$

Note that p_1 `=` p_2 (p_1 `==` p_2, resp.) means that there is at least one node selected by p_1 which has the same value as (is identical to, resp.) a node selected by p_2. This "existential" interpretation of `=` is that of XPath 1.0.

The definitions of $\mathcal{S}$ and $\mathcal{Q}$ can be seen as (formal specifications of) an implementation of xPath (a functional program is quite easy to derive from these specifications).

3 Location Path Equivalences

A set of simple equivalences is first established. These are then used to prove equivalences of paths with reverse axes. We distinguish between general equivalences that can be applied to remove any reverse axes, and specific equivalences, each of them being applicable to a certain case.

Making use of the semantics of xPath given in Section 2, the equivalence of location paths can be formally defined as follows.

Definition 3.1 (Path equivalence). Two location paths p_1 and p_2 are *equivalent,* noted $p_1 \equiv p_2$, if $\mathcal{S}[\![p_1]\!] = \mathcal{S}[\![p_2]\!]$, i.e. if $\mathcal{S}[\![p_1]\!]x = \mathcal{S}[\![p_2]\!]x$ for all nodes x (from any document).

Intuitively, two location paths are equivalent if they select the same set of nodes for every document and every context node in this document.

Lemma 3.1. *Let p, p_1, and p_2 be location paths, q, q_1, and q_2 qualifiers, n a node test, and $\theta \in \{$`==`, `=`$\}$.*

1. **Right step adjunction:** *If $p_1 \equiv p_2$ and p relative, then p_1`/`$p \equiv p_2$`/`p.*
2. **Left step adjunction:** *If $p_1 \equiv p_2$ and p_1, p_2 relative, then p`/`$p_1 \equiv p$`/`p_2.*
3. **Qualifier adjunction:** *If $p_1 \equiv p_2$, then p_1`[`q`]` $\equiv p_2$`[`q`]` and p`[`p_1`]` $\equiv p$`[`p_2`]`.*
4. **Relative/absolute path conversion:** *If $p_1 \equiv p_2$, then* `/`$p_1 \equiv$ `/`p_2.
5. **Union/disjunction conversion:** p`[`q_1`]` `|` p`[`q_2`]` $\equiv p$`[`q_1 `or` q_2`]`.
6. **Qualifier flattening:** p`[`p_1`/`p_2`]` $\equiv p$`[`p_1`[`p_2`]]`.
7. **Ancestor-or-self axis decomposition:**
   `ancestor-or-self::`$n \equiv$ `ancestor::`n `|` `self::`n.
8. **Descendant-or-self axis decomposition:**
   `descendant-or-self::`$n \equiv$ `descendant::`n `|` `self::`n.
9. **Union term adjunction:** *If $p_1 \equiv p_2$, then p_1* `|` $p \equiv p_2$ `|` p.
10. **Union commutativity:** p_1 `|` $p_2 \equiv p_2$ `|` p_1.
11. **Qualifiers with joins:** p`[`p_1 θ `/`p_2`]` $\equiv p$`[`p_1`[self::node()` θ `/`p_2`]]`.

Recall that $\perp$ is a location path never selecting any node whatever the current node and document are. Since the root node has no parents and therefore no siblings, the following holds:

Lemma 3.2. *Let m and n be node tests, i.e. m and n are tag names or one of the xPath constructs* `*`, `node()`, *or* `text()`.

- *Let a be one of the axes* `parent`, `ancestor`, `preceding`, `preceding-sibling`, `self`, `following`, *or* `following-sibling`. *Then the following holds:*

$$/a::n \equiv \begin{cases} / & \text{if } a = \texttt{self} \text{ and } n = \texttt{node()} \\ \perp & \text{otherwise} \end{cases}$$

- *Let a be the* `preceding` *or* `ancestor` *axis. Then the following equivalences hold:*

$$/\texttt{child::}m/a::n \equiv \begin{cases} /\texttt{self::node()[child::}m\texttt{]} & \text{if } a = \texttt{ancestor} \text{ and } n = \texttt{node()} \\ \perp & \text{otherwise} \end{cases}$$

$$/\texttt{child::}m\texttt{[}a::n\texttt{]} \equiv \begin{cases} /\texttt{child::}m & \text{if } a = \texttt{ancestor} \text{ and } n = \texttt{node()} \\ \perp & \text{otherwise} \end{cases}$$

3.1 General equivalences

The nodes selected by a reverse step within a location path are necessarily descendants of the document root. The following equivalences show how for any reverse axis only those descendants of the root can be selected that are also matched by the original reverse step.

Proposition 3.1. *Let p and s be relative location paths, n and m node tests, a_m a reverse axis, a_n a forward axis, and b_m the symmetrical axis of a_m. Then the following holds*

$$p\texttt{[}a_m\texttt{::}m/s\texttt{]} \equiv p\texttt{[/descendant::}m\texttt{[}s\texttt{]/}b_m\texttt{::node() == self::node()]} \tag{1}$$

$$/p/a_n\texttt{::}n/a_m\texttt{::}m \equiv \texttt{/descendant::}m\texttt{[}b_m\texttt{::}n\texttt{ == /}p/a_n\texttt{::}n\texttt{]} \tag{2}$$

$$/a_n\texttt{::}n/a_m\texttt{::}m \equiv \texttt{/descendant::}m\texttt{[}b_m\texttt{::}n\texttt{ == /}a_n\texttt{::}n\texttt{]} \tag{2a}$$

Equivalence (1) shows that it is possible to remove the first step in a location path within a qualifier. With help of Lemma 3.1.5 this result is generalized to reverse steps having an arbitrary position within a qualifier.

Figure 3 illustrates the key idea of Equivalence (1) (solid arrows for the navigation in the left-hand side of Equivalence (1), dashed for the right-hand side): Instead of looking back from the context node specified by path p for matching a certain node ($a_m\texttt{::}m$) (solid arrows), one can look forward from the beginning of the document for matching the node (`/descendant::`m) and then, still forward, for reaching the initial context node (b_m`::node()`, dashed arrows). The selected node (indicated with a box) is the same for both location paths. Hence, e.g. instead of checking whether the context node specified by path p has a preceding m (p`[preceding::`m`]`), we rather look for an m node and then for a following node that is identical to the context node:

$$p\texttt{[/descendant::}m\texttt{/following::node() == self::node()]}.$$

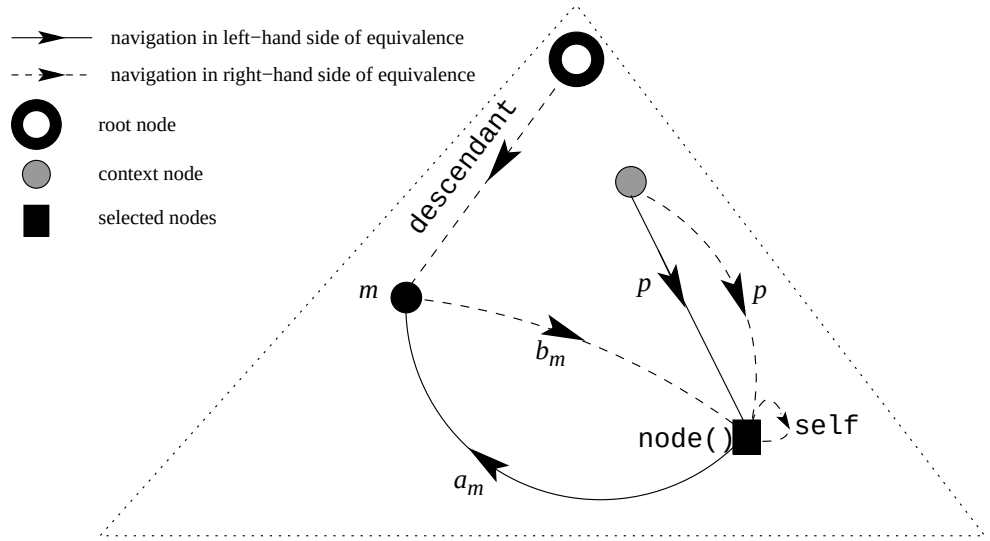
Equivalences (2) and (2a) remove the first step from an absolute location path using the same underlying idea.

Note that the equality occurring in these equivalences is based on node identity. The equivalent paths might remain expensive to evaluate, but no evaluation of the $a_m\texttt{::}m$ reverse step is needed anymore.

Example 3.1. Consider the example of Figure 1 and a query asking for all `name`s that appear before a `price`. A way to select these nodes is using the following location path:

```
/descendant::price/preceding::name
```

Figure 3 Tree navigation used in Equivalence (1)



$p[a_m::m] \equiv p$`[/descendant::`m/b_m`::node() == self::node()]`

By Equivalence (2a), the `preceding` axis can be removed yielding to the following equivalent location path:

```
/descendant::name[following::price == /descendant::price]
```

While the initial location path selects all `name` nodes preceding a `price` node, the equivalent location path selects all `name` nodes, that have a following `price` node, if that node is also a descendant of the root. It is obvious, that there is a considerably simpler equivalent location path (dropping the join), `/descendant::name[following::price]`. The need for the join arises, as the location path selecting the context nodes, relative to which the reverse step is evaluated, (in this case the `price` nodes) can be arbitrarily complex:

Consider a slightly modified case of the previous one, where only `price`s, that are inside of a `journal` with a `title`, should be considered. A possible location path for this query with reverse axis is:

```
/descendant::journal[child::title]/descendant::price/preceding::name
```

Again, by Equivalence (2) this is equivalent to

```
/descendant::name[following::price == /descendant::journal[child::title]/descendant::price].
```

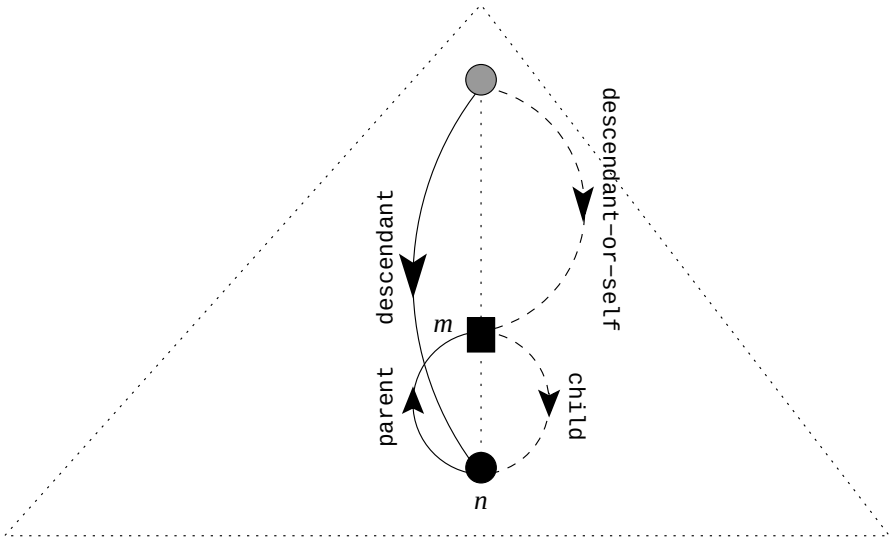
As argued above, it is impossible in this case to remove the introduced join. Note that the join in the first example can be removed by additional equivalence rules for simplifying location paths that are outside the scope of this paper.

Using the equivalences above, it is possible to replace reverse steps in xPath expressions. Nonetheless, in the following section specific equivalences for reverse axes are given, that yield to location paths without joins.

3.2 Specific Equivalences

In this section the interaction of the reverse axes (`ancestor`, `ancestor-or-self`, `parent`, `preceding`, and `preceding-sibling`) with forward axes is treated, i.e. equivalences are given, that (if read as rewriting rules from left to right), depending on the location step L_f before a reverse location step L_r, either replace the reverse location step L_r or rewrite the location path

Figure 4 Tree navigation used in Equivalence (3)



`descendant::`n`/parent::`$m \equiv$ `descendant-or-self::`m`[child::`n`]`.

into one, where the reverse step L_r is "pushed leftwise". For every reverse step, the interaction with every forward step is shown.

In general the equivalences have the following structure

$$p/L_f/L_r \equiv p' \quad \text{or} \quad p/L_f[L_r] \equiv p',$$

where p is an absolute path, L_f a forward location step, L_r a reverse location step, and p' the equivalent location path. Sometimes the equivalences can be formulated without the leading path p.

Note that interaction with reverse axes, e.g. interaction of `parent` with `preceding-sibling`, is not necessary to investigate in these equivalences due to the way our algorithm works (removing reverse steps from left to right of the location path in question). Also, equivalences involving `ancestor-or-self` or `descendant-or-self` are not necessary since we can replace these location steps using Equivalences (3.1.7) and (3.1.8).

Some of the following equivalences do still contain a reverse step on the right-hand side, but this reverse step is either more on the left of the location path, or the right hand side is of a form, where other equivalences can be applied to fully remove the reverse location steps as elaborated in Section 4.

3.2.1 Parent

The equivalences in the following proposition are divided in two sets. The first set (Equivalences (3) to (7)) covers the case of `parent` location steps outside, the second inside a qualifier. Note, that there is a strong structural similarity between the equivalences of the two sets. The basic idea of the following equivalences is that a `parent` step can be avoided if the location path is modified in a certain way, depending on the location step before the `parent` step. Figure 4 illustrates this for the simple example of the interaction between a `descendant` and a `parent` step (Equivalence (3)). The solid arrows, following the navigation in the left-hand side of Equivalence (3), end at the same node (indicated with a box) as the dashed arrows, that correspond to the right-hand side of the equivalence.

Proposition 3.2 (`parent` axis). *Let m and n be node tests and p a location path.*

$$\texttt{descendant::}n\texttt{/parent::}m \equiv \texttt{descendant-or-self::}m\texttt{[child::}n\texttt{]} \tag{3}$$

$$\texttt{child::}n\texttt{/parent::}m \equiv \texttt{self::}m\texttt{[child::}n\texttt{]} \tag{4}$$

$$p\texttt{/self::}n\texttt{/parent::}m \equiv p\texttt{[self::}n\texttt{]/parent::}m \tag{5}$$

$$p\texttt{/following-sibling::}n\texttt{/parent::}m \equiv p\texttt{[following-sibling::}n\texttt{]/parent::}m \tag{6}$$

$$\begin{aligned} p\texttt{/following::}n\texttt{/parent::}m \equiv\ & p\texttt{/following::}m\texttt{[child::}n\texttt{]} \\ & \mid p\texttt{/ancestor-or-self::*[following-sibling::}n\texttt{]} \\ & \quad\texttt{/parent::}m \end{aligned} \tag{7}$$

$$\texttt{descendant::}n\texttt{ [parent::}m\texttt{]} \equiv \texttt{descendant-or-self::}m\texttt{/child::}n \tag{8}$$

$$\texttt{child::}n\texttt{[parent::}m\texttt{]} \equiv \texttt{self::}m\texttt{/child::}n \tag{9}$$

$$p\texttt{/self::}n\texttt{[parent::}m\texttt{]} \equiv p\texttt{[parent::}m\texttt{]/self::}n \tag{10}$$

$$p\texttt{/following-sibling::}n\texttt{[parent::}m\texttt{]} \equiv p\texttt{[parent::}m\texttt{]/following-sibling::}n \tag{11}$$

$$\begin{aligned} p\texttt{/following::}n\texttt{[parent::}m\texttt{]} \equiv\ & p\texttt{/following::}m\texttt{/child::}n \\ & \mid p\texttt{/ancestor-or-self::*[parent::}m\texttt{]} \\ & \quad\texttt{/following-sibling::}n \end{aligned} \tag{12}$$

Example 3.2. Consider the data of Figure 1. The following location path selects all elements that are parents of an `editor` element, which itself is a child of a `journal` element (recall that `journal` is a child of the root node):

```
/child::journal/child::editor/parent::*.
```

According to Equivalence (4), this path is equivalent to:

```
/child::journal/self::*[child::editor].
```

It is worth noting, that the dispensable location step `self::*` can be removed (resulting in `/child::journal[child::editor]`) by additional equivalences for simplifying location paths that are beyond the scope of this paper.

3.2.2 Ancestor

The following proposition gives equivalences that either move an `ancestor` step to the left of a path or remove it completely. Equivalences (13a) and (18a) are special cases of Equivalences (13) and (18), respectively. Figure 5 illustrates as an example the interaction between a `descendant` and an `ancestor` step (Equivalence (13)): On the left-hand side of that figure the three possibilities where the m node may occur are indicated by the three boxes and the three arrows labeled `ancestor`, (1) above the context node, (2) between the context node and the node selected by path p, or (3) between that node and the n node. In the middle part of Figure 5 the first disjunct, covering the first two cases, is illustrated, whereas the right part covers case (3).

Proposition 3.3 (`ancestor` axis). *Let m and n be node tests and p a location path.*

$$\begin{aligned} p\texttt{/descendant::}n\texttt{/ancestor::}m \equiv\ & p\texttt{[descendant::}n\texttt{]/ancestor::}m \\ & \mid p\texttt{/descendant-or-self::}m\texttt{[descendant::}n\texttt{]} \end{aligned} \tag{13}$$

$$\texttt{/descendant::}n\texttt{/ancestor::}m \equiv \texttt{/descendant-or-self::}m\texttt{[descendant::}n\texttt{]} \tag{13a}$$

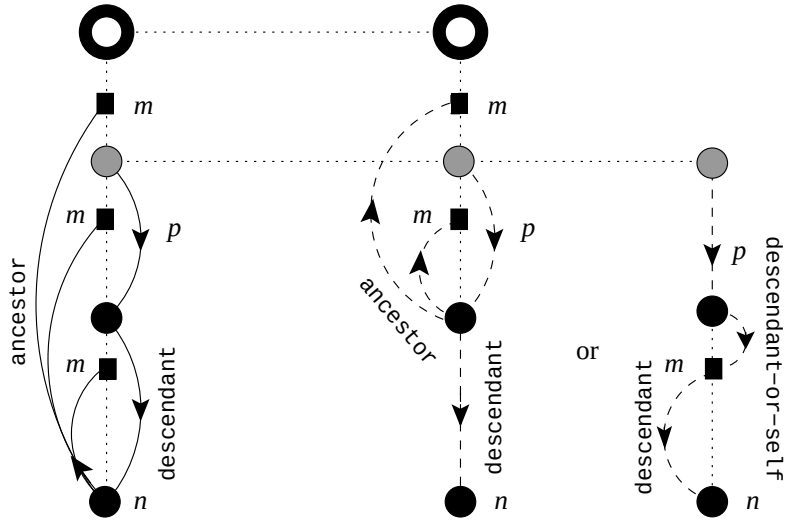
$$p\texttt{/child::}n\texttt{/ancestor::}m \equiv p\texttt{[child::}n\texttt{]/ancestor-or-self::}m \tag{14}$$

$$p\texttt{/self::}n\texttt{/ancestor::}m \equiv p\texttt{[self::}n\texttt{]/ancestor::}m \tag{15}$$

$$p\texttt{/following-sibling::}n\texttt{/ancestor::}m \equiv p\texttt{[following-sibling::}n\texttt{]/ancestor::}m \tag{16}$$

$$\begin{aligned} p\texttt{/following::}n\texttt{/ancestor::}m \equiv\ & p\texttt{/following::}m\texttt{[descendant::}n\texttt{]} \\ & \mid p\texttt{/ancestor-or-self::*} \\ & \quad\texttt{[following-sibling::*/descendant-or-self::}n\texttt{]} \\ & \quad\texttt{/ancestor::}m \end{aligned} \tag{17}$$

Figure 5 Tree navigation used in Equivalence (13)



p/descendant::n/ancestor::$m \equiv p$[descendant::n]/ancestor::m
| p/descendant-or-self::m[descendant::n]

$$p\texttt{/descendant::}n\texttt{[ancestor::}m\texttt{]} \equiv p\texttt{[ancestor::}m\texttt{]/descendant::}n \quad (18)$$
$$\texttt{| } p\texttt{/descendant-or-self::}m\texttt{/descendant::}n$$

$$\texttt{/descendant::}n\texttt{[ancestor::}m\texttt{]} \equiv \texttt{/descendant-or-self::}m\texttt{/descendant::}n \quad (18a)$$

$$p\texttt{/child::}n\texttt{[ancestor::}m\texttt{]} \equiv p\texttt{[ancestor-or-self::}m\texttt{]/child::}n \quad (19)$$

$$p\texttt{/self::}n\texttt{[ancestor::}m\texttt{]} \equiv p\texttt{[ancestor::}m\texttt{]/self::}n \quad (20)$$

$$p\texttt{/following-sibling::}n\texttt{[ancestor::}m\texttt{]} \equiv p\texttt{[ancestor::}m\texttt{]/following-sibling::}n \quad (21)$$

$$p\texttt{/following::}n\texttt{[ancestor::}m\texttt{]} \equiv p\texttt{/following::}m\texttt{/descendant::}n \quad (22)$$
$$\texttt{| } p\texttt{/ancestor-or-self::*[ancestor::}m\texttt{]}$$
$$\texttt{/following-sibling::*/descendant-or-self::}n$$

Example 3.3. The following location path selects all `name`s that have a `journal` element among their ancestors:

```
/descendant::name[ancestor::journal].
```

According to Equivalence (18a), this path is equivalent to:

```
/descendant-or-self::journal/descendant::name.
```

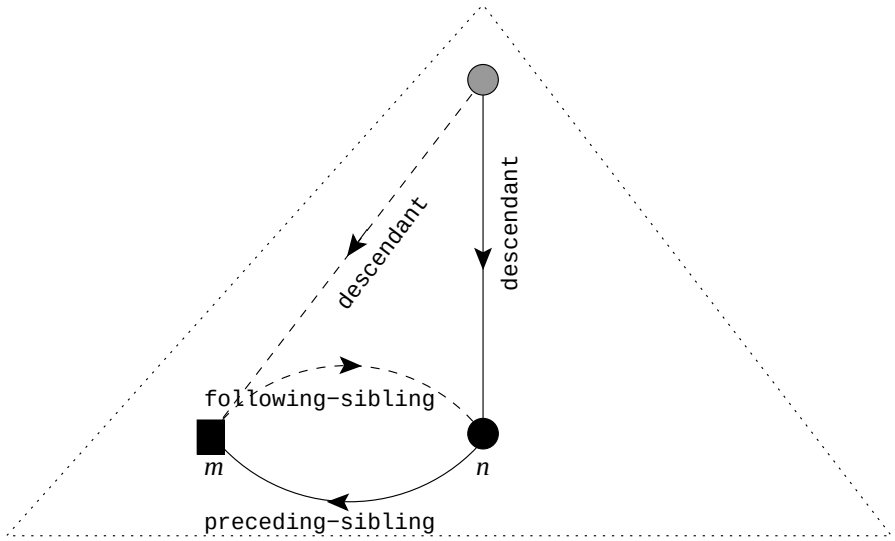
It is easy to see that the two location paths select the same set of nodes.

3.2.3 Preceding-sibling

In the following proposition the `preceding-sibling` axis is treated. Note that the right-hand sides of equivalences for `preceding-sibling` (and `preceding`) contain more union terms than the other equivalences, since there is no `-or-self` variant of these axes. Figure 6 illustrates as an example the interaction between a `descendant` and a `preceding-sibling` step (Equivalence (23)).

Proposition 3.4 (`preceding-sibling` axis). *Let m and n be node tests and p a location path. The following equivalences hold:*

Figure 6 Tree navigation used in Equivalence (23)



`descendant::`n`/preceding-sibling::`$m \equiv$ `descendant::`m`[following-sibling::`n`]`

$$\texttt{descendant::}n\texttt{/preceding-sibling::}m \equiv \texttt{descendant::}m\texttt{[following-sibling::}n\texttt{]} \quad (23)$$

$$\texttt{child::}n\texttt{/preceding-sibling::}m \equiv \texttt{child::}m\texttt{[following-sibling::}n\texttt{]} \quad (24)$$

$$p\texttt{/self::}n\texttt{/preceding-sibling::}m \equiv p\texttt{[self::}n\texttt{]/preceding-sibling::}m \quad (25)$$

$$\begin{aligned} p\texttt{/following-sibling::}n\texttt{/preceding-sibling::}m \equiv\ & p\texttt{[self::}m\texttt{/following-sibling::}n\texttt{]} \\ & \mid p\texttt{[following-sibling::}n\texttt{]/preceding-sibling::}m \\ & \mid p\texttt{/following-sibling::}m\texttt{[following-sibling::}n\texttt{]} \end{aligned} \quad (26)$$

$$\begin{aligned} p\texttt{/following::}n\texttt{/preceding-sibling::}m \equiv\ & p\texttt{/following::}m\texttt{[following-sibling::}n\texttt{]} \\ & \mid p\texttt{/ancestor-or-self::*[following-sibling::}n\texttt{]} \\ & \quad \texttt{/preceding-sibling::}m \\ & \mid p\texttt{/ancestor-or-self::}m\texttt{[following-sibling::}n\texttt{]} \end{aligned} \quad (27)$$

$$\texttt{descendant::}n\texttt{[preceding-sibling::}m\texttt{]} \equiv \texttt{descendant::}m\texttt{/following-sibling::}n \quad (28)$$

$$\texttt{child::}n\texttt{[preceding-sibling::}m\texttt{]} \equiv \texttt{child::}m\texttt{/following-sibling::}n \quad (29)$$

$$p\texttt{/self::}n\texttt{[preceding-sibling::}m\texttt{]} \equiv p\texttt{[self::}n\texttt{]/following-sibling::}m \quad (30)$$

$$\begin{aligned} p\texttt{/following-sibling::}n\texttt{[preceding-sibling::}m\texttt{]} \equiv\ & p\texttt{[self::}m\texttt{]/following-sibling::}n \\ & \mid p\texttt{/following-sibling::}m\texttt{/following-sibling::}n \\ & \mid p\texttt{[preceding-sibling::}m\texttt{]/following-sibling::}n \end{aligned} \quad (31)$$

$$\begin{aligned} p\texttt{/following::}n\texttt{[preceding-sibling::}m\texttt{]} \equiv\ & p\texttt{/following::}m\texttt{/following-sibling::}n \\ & \mid p\texttt{/ancestor-or-self::*[preceding-sibling::}m\texttt{]} \\ & \quad \texttt{/following-sibling::}n \\ & \mid p\texttt{/ancestor-or-self::/following-sibling::}n \end{aligned} \quad (32)$$

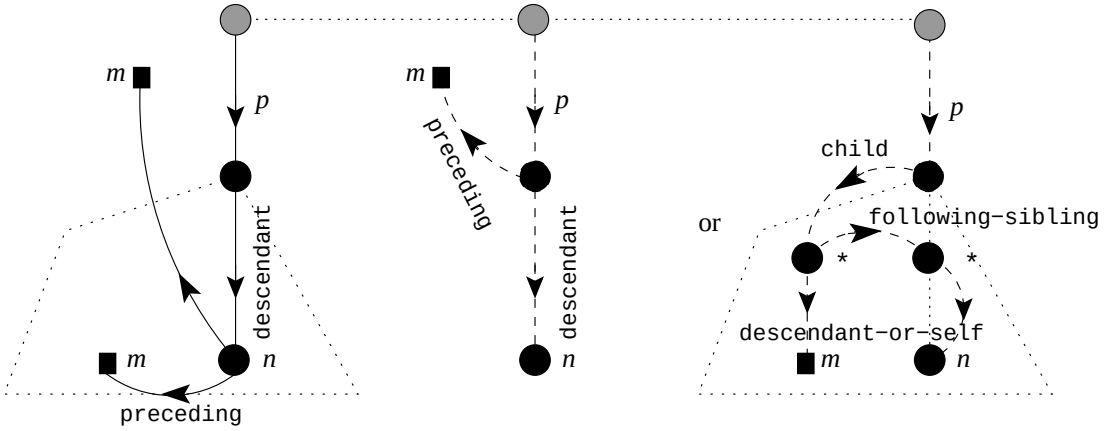
Example 3.4. Consider the example of Figure 1 and a query asking for all `editor`s that have siblings `price`s inside of a `journal` following them. This query can be expressed by the location path:

```
/child::journal/descendant::price/preceding-sibling::editor
```

By Equivalence (23), the `preceding-sibling` axis can be removed yielding the following equivalent location path:

```
/child::journal/descendant::editor[following-sibling::price]
```

Figure 7 Tree navigation used in Equivalence (33)



p/descendant::n/preceding::$m \equiv p$[descendant::n]/preceding::m
| p/child::*[following-sibling::*/descendant-or-self::n]/descendant-or-self::m

3.2.4 Preceding

The following proposition contains the equivalences that treat the interaction of `preceding` with other axes. Figure 4 illustrates this for the example of the interaction between a `descendant` and a `preceding` step (Equivalence (33)). In the left part it can be seen that there are two cases where the m node can appear, either as a descendant of the node selected by p or not. These two cases directly correspond to the two disjuncts in the right-hand side of Equivalence (33). The right part of the figure covers the first case, whereas the middle part covers the second case.

Proposition 3.5 (`preceding` axis). *Let m and n be node tests and p a location path.*

$$p\text{/descendant::}n\text{/preceding::}m \equiv p\text{[descendant::}n\text{]/preceding::}m \quad (33)$$
$$\quad | \; p\text{/child::*[following-sibling::*/descendant-or-self::}n\text{]/descendant-or-self::}m$$
$$\text{/descendant::}n\text{/preceding::}m \equiv \text{/descendant::}m\text{[following::}n\text{]} \quad (33a)$$
$$p\text{/child::}n\text{/preceding::}m \equiv p\text{[child::}n\text{]/preceding::}m \quad (34)$$
$$\quad | \; p\text{/child::*[following-sibling::}n\text{]/descendant-or-self::}m$$
$$p\text{/self::}n\text{/preceding::}m \equiv p\text{[self::}n\text{]/preceding::}m \quad (35)$$
$$p\text{/following-sibling::}n\text{/preceding::}m \equiv p\text{[following-sibling::}n\text{]/preceding::}m \quad (36)$$
$$\quad | \; p\text{/following-sibling::*[following-sibling::}n\text{]/descendant-or-self::}m$$
$$\quad | \; p\text{[following-sibling::}n\text{]/descendant-or-self::}m$$
$$p\text{/following::}n\text{/preceding::}m \equiv p\text{[following::}n\text{]/preceding::}m \quad (37)$$
$$\quad | \; p\text{/following::}m\text{[following::}n\text{]}$$
$$\quad | \; p\text{[following::}n\text{]/descendant-or-self::}m$$

$$p\text{/descendant::}n\text{[preceding::}m\text{]} \equiv p\text{[preceding::}m\text{]/descendant::}n \quad (38)$$
$$\quad | \; p\text{/child::*[descendant-or-self::}m\text{]/following-sibling::*/descendant-or-self::}n$$
$$\text{/descendant::}n\text{[preceding::}m\text{]} \equiv \text{/descendant::}m\text{/following::}n \quad (38a)$$

$$p/\texttt{child::}n\texttt{[preceding::}m\texttt{]} \equiv p\texttt{[preceding::}m\texttt{]/child::}n \quad (39)$$
$$\mid p/\texttt{child::*[descendant-or-self::}m\texttt{]}$$
$$\texttt{/following-sibling::}n$$
$$p/\texttt{self::}n\texttt{[preceding::}m\texttt{]} \equiv p\texttt{[preceding::}m\texttt{]/self::}n \quad (40)$$
$$p/\texttt{following-sibling::}n\texttt{[preceding::}m\texttt{]} \equiv p\texttt{[preceding::}m\texttt{]/following-sibling::}n \quad (41)$$
$$\mid p/\texttt{following-sibling::*[descendant-or-self::}m\texttt{]}$$
$$\texttt{/following-sibling::}n$$
$$\mid p\texttt{[descendant-or-self::}m\texttt{]/following-sibling::}n$$
$$p/\texttt{following::}n\texttt{[preceding::}m\texttt{]} \equiv p\texttt{[preceding::}m\texttt{]/following::}n \quad (42)$$
$$\mid p/\texttt{following::}m\texttt{/following::}n$$
$$\mid p\texttt{[descendant-or-self::}m\texttt{]/following::}n$$

Example 3.5. Consider the location path

```
/descendant::price/preceding::name
```

of Example 3.1. With Rule (33a) it can be rewritten to

```
/descendant::name[following::price].
```

This result is more compact and closer to the original than the result of Example 3.1 using Equivalence (2a).

4 Location Path Rewriting

Each Equivalence (i) $p_1 \equiv p_2$ of Section 3 gives rise to a rewriting rule: A path matching with the left hand-side p_1 of Equivalence (i) can be rewritten into a path corresponding to the right-hand side p_2 of Equivalence (i). In the following, Rule (i) denotes the rewriting rule $p_1 \rightarrow p_2$ induced by Equivalence (i) $p_1 \equiv p_2$.

The equivalences of Lemma 3.1 and Lemma 3.2 induce rewriting rules, denoted Rules (3.1.1) to (3.1.11) and (3.2).

The equivalences of Section 3 are splitted in two sets of rules for use in a rewriting algorithm:

1. RuleSet_1, containing the general Rules (1), (2), (2a) and (3.2).
2. RuleSet_2, containing the specific Rules (3) to (42) and (3.2).

A rule can be applied to a location path in the following straightforward manner:

Definition 4.1 (Rule application). Let p be a path expression without unions (|), and let $p_l \rightarrow p_r$ be a rule either from RuleSet_1 or RuleSet_2. If p is of the form p_l/p', then let q denote the path p_r/p'. If p_l is a relative path and if p is of the form $p_1/p_l/p_2$, then let q denote the path $p_1/p_r/p_2$. In both cases q is called the *result* of the application of rule $p_l \rightarrow p_r$ to p.

An algorithm, called "rare" (sketched in Figure 8) for computing a reverse-axis-free path equivalent to an absolute path is defined below. The input for the algorithm is restricted to paths with no qualifiers containing so-called "RR joins":

Definition 4.2 (RR join). An RR join is an expression of the form $p_1\ \theta\ p_2$ where $\theta \in \{\texttt{==}, \texttt{=}\}$, and both p_1 and p_2 are **R**elative paths such that at least one of them contains a **R**everse step.

For the consideration of termination and correctness of the algorithm, some important properties of the application of the rewriting rules to a location path are required:

Lemma 4.1 (Properties of rule application). *Let p be an absolute location path with no qualifier containing RR joins.*

1. *If p contains a reverse step, then a rule from RuleSet$_1$ and a rule from RuleSet$_2$ is applicable to p. Possibly, Rules (3.1.1) to (3.1.11) have to be applied first.*

2. *The result of a rule application to the first reverse step in p is an absolute path with no qualifiers containing RR joins.*

3. *If q is the result of a rule application to p, then $p \equiv q$.*

Proof. **(1):** Let L be the first reverse location step.

First consider RuleSet$_1$: If L occurs as a location step outside a qualifier, Rules (2), (2a) or (3.2) can be applied, since p is an absolute location path. If L occurs as the first location step inside a qualifier we can apply Rule (1). If L appears at any other position inside a qualifier, we can apply Lemma 3.1.6 in order to construct a qualifier with L as first location step. We can now apply Rule (1).

RuleSet$_2$ provides rules for interaction between each reverse step and an arbitrary forward step, so there is always a rule, that can be applied to the first reverse step in p.
(2) Only Rules (1), (2) and (2a) introduce a binary relation (namely `==`), if they are applied to a location path. But always one of the two paths related by `==` is absolute. So in any case the result of the rule application contains no RR join. Furthermore since p is an absolute path the result of applying a rule to p is also an absolute path.
(3) This holds due to Lemma 3.1.1−4. □

"rare" Algorithm. The "rare" algorithm, outlined in Figure 8, can be used for RuleSet$_1$ as well as for RuleSet$_2$. The algorithm takes as input a location path which is absolute, since some rules from RuleSet$_1$ and RuleSet$_2$ are applicable to absolute location paths only.

Theorem 4.1 (Removal of reverse location steps using RuleSet$_1$). *Let p be an absolute path with no qualifiers in which RR joins occur. There exists an absolute path p' with no reverse steps such that $p \equiv p'$. Using "rare" and RuleSet$_1$, this path p' has a length and can be computed in a time linear in the length of p.*

Proof. We construct a path equivalent to p as sketched in Figure 8. All reverse location steps are rewritten, one by one, as follows: Lemma 4.1 guarantees that a rule of RuleSet$_1$ can be applied to any path containing a reverse location step. The resulting path p' contains no reverse location steps and is equivalent to p.

The location path p' is of linear size and constructed in linear time, since each rule application removes one reverse step, adds at most two forward location steps and no reverse ones. □

Theorem 4.2 (Removal of reverse location steps using RuleSet$_2$). *Let p be an absolute path with no qualifiers in which RR joins occur. There exists an absolute path p' with no reverse steps such that $p \equiv p'$. Using "rare" and RuleSet$_2$, this path p' has a length and can be computed in a time exponential in the length of p in the worst case.*

Proof. The rule application for RuleSet$_2$ can have three different result types:

1. removes completely a reverse step, e.g. Rules (3.2) or (3);

2. pushes the reverse step from right to left in the path, e.g. Rule (5);

3. for the interaction between a `following` and a reverse step L, i.e. `following::`n`/`L or `following::`n`[`L`]`, a union of several other paths is obtained, e.g. Rule (7); the resulting union terms have reverse steps at positions less or equal than in the original path and they do not contain anymore the interaction between the initial `following` step and a reverse step.

Figure 8 Algorithm ***rare*** (**r**everse **a**xis **re**moval)

Let ξ = RuleSet$_1$ or RuleSet$_2$.

Auxiliary functions:

match(p): returns the result of a rule application from ξ to the first reverse location step in p.

apply-lemmas(p): returns p if Rules (3.1.1-11) are not applicable to p. Otherwise, it returns the result of the repeated application of Rules (3.1.1-11) to p.

union-flattening(p): returns a path equivalent to p with unions at top level only.

***rare*(p)**
Input: p {absolute location path with no qualifiers containing RR joins}.

$p \leftarrow$ *apply-lemmas*(p).
$p \leftarrow$ *union-flattening*(p) $= U_1 \mid \ldots \mid U_n$ ($n \geq 1$).
$S \leftarrow$ empty stack.
for $i \leftarrow 1$ to n **do**
  push(U_i, S).
end for
$p' \leftarrow \perp$ {initialization}
while *not*(*empty*(S)) **do**
  $U \leftarrow$ *pop*(S).
  while U contains a reverse step **do**
    $U \leftarrow$ *match*(U).
    $U \leftarrow$ *apply-lemmas*(U).
    $U \leftarrow$ *union-flattening*(U) $= V_1 \mid \ldots \mid V_n$ ($n \geq 1$).
    for $i \leftarrow 2$ to n **do**
      push(V_i, S).
    end for
    $U \leftarrow V_1$.
  end while
  $p' \leftarrow p' \mid U$
end while

Output: p' {location path without reverse axes equivalent to p }.

Figure 9 Example run of ***rare*** algorithm with RuleSet_1

Consider the example of Figure 1 and a query asking for all `titles` that appear before a `name` and are inside `journals`. This query can be expressed as the following location path:

```
/descendant::name/preceding::title[ancestor::journal]
```

Note that p is an absolute path without qualifiers containing RR-joins.

Step 1: $p \leftarrow$ *apply-lemmas*$(p) = p$.
Step 2: $U_1 \leftarrow$ `/descendant::name/preceding::title[ancestor::journal]`.
Step 3: *push*(U_1, S).
Step 4: $p' \leftarrow \perp$.
Step 5: $U \leftarrow$ *pop*(S).
Step 6: U contains a reverse step (`preceding::title`).
Step 7: $U \leftarrow$ *match*$(U) =$ `/descendant::title[ancestor::journal]`
`[following::name == /descendant::name]`. {Rule (2)}
Step 8: $U \leftarrow$ *apply-lemmas*$(U) = U$.
Step 9: U contains a reverse step (`ancestor::journal`).
Step 10: $U \leftarrow$ *match*$(U) =$ `/descendant::title`
`[/descendant::journal/descendant::node() == self::node()]`
`[following::name == /descendant::name]`. {Rule (1)}
Step 11: $U \leftarrow$ *apply-lemmas*$(U) = U$.
Step 12: U does not contain reverse steps.
Step 13: $p' \leftarrow U$.
Step 14: S is empty.

Output: $p =$ `/descendant::title`
`[/descendant::journal/descendant::node() == self::node()]`
`[following::name == /descendant::name]`.

Since the path has a finite length, the procedure of pushing reverse steps closer to the root terminates. Also, the number of interactions between `following` and reverse steps is finite. Hence, the algorithm terminates.

Each rule application having the first result type removes a reverse step and does not change the number of union terms. Hence, in the best case, i.e. using only rules with the first result type, the algorithm has a linear complexity in the length of the input path p.

The last two result types are significant for the worst-case complexity of the algorithm, since each rule application can produce intermediate rewritten paths with more than one union term (up to three union terms). Hence, each rule application can increase the order of the input for the next rule application. This yields to an exponential complexity in the length of the input path p. □

Example runs of the algorithm for both set of rules are presented in Figure 9 and Figure 10.

Comparison. Both RuleSet_1 and RuleSet_2 have advantages and it is an open issue which one is preferable. The path rewriting using RuleSet_2 has an exponential time complexity and output size in the length of the input location path. As location paths are in practice small (less than ten steps), the exponential worst-case complexity of RuleSet_2 does not necessarily generate longer paths than RuleSet_1. In addition, since they do not contain joins, the location paths generated using RuleSet_2 are simpler (as can be seen in the examples), hence more convenient to evaluate, than those generated using RuleSet_1, which contain the same number of joins as there are reverse steps in the input location path. It remains to decide in practical tests, up to which size of an input path RuleSet_2 generates simpler paths than RuleSet_1.

Figure 10 Example run of ***rare*** algorithm with RuleSet$_2$

Consider the example query in Figure 9. For conciseness, the union terms in Step 8 are not pushed to the stack.

Step 1: $p \leftarrow$ *apply-lemmas*$(p) = p$.
Step 2: $U_1 \leftarrow$ `/descendant::name/preceding::title[ancestor::journal]`.
Step 3: *push*(U_1, S).
Step 4: $p' \leftarrow \perp$.
Step 5: $U \leftarrow$ *pop*(S).
Step 6: U contains a reverse step (`preceding::title`).
Step 7: $U \leftarrow$ *match*$(U) =$
`/descendant-or-self::title[following::name][ancestor::journal]` {Rule (33a)}
Step 8: $U \leftarrow$ *apply-lemmas*$(U) =$ `/descendant::title[following::name][ancestor::journal]`
`| /self::title[following::name][ancestor::journal]`.
Step 9: U contains a reverse step (`ancestor::journal`).
Step 10: $U \leftarrow$ *match*$(U) =$
`/descendant::title[following::name][ancestor::journal]` | $\perp$. {Rule (3.2)}
Step 11: U contains a reverse step (`ancestor::journal`).
Step 12: $U \leftarrow$ *match*$(U) =$
`/descendant::journal/descendant::title[following::name]`. {Rule (18a)}
Step 13: $U \leftarrow$ *apply-lemmas*$(U) = U$.
Step 14: U does not contain reverse steps.
Step 15: $p' \leftarrow U$.
Step 16: S is empty.

Output: $p' =$ `/descendant::journal/descendant::title[following::name]`.

5 Rewriting Location Paths Using Variables

There are two classes of location paths that are not covered by the rules given so far: relative location paths and location paths with RR joins (cf. Definition 4.2), e.g. p`[self::* = preceding::*]`. Any attempt to remove the reverse location steps in these cases results in losing the context given by p.

The approach proposed in this section solves this problem by remembering the context in a variable. XPath 2.0 uses the `for` clause (originating from XQuery 1.0) to set the value of variables dynamically. In the following paragraphs the XPath 2.0 syntax is used in the examples, as this syntax is also available in XQuery 1.0 and XSLT 2.0. Some familiarity with XPath 2.0 is assumed.

Rewriting relative location paths. Every relative location path can be rewritten by reselecting the context node in an absolute location path: The context node is bound to a variable, and all descendants of the root (i.e. all nodes) that are identical to this variable are selected. Obviously this selects only the context node. For this technique the use of variables is necessary, as there is no other way to refer to the context node inside of an absolute location path. This approach is very similar to the idea of the general equivalences.

The resulting XPath 2.0 location path is a (`for`) expression binding a variable to the context node and returning nodes selected by the path r from the reselected context node (`/descendant::*[self::node() == $ctx]`). Note that there is only one binding for `$ctx`, since the node set created by `self` is a singleton.

Example 5.1. Consider an relative location path r containing a reverse step. For saving the context node we introduce a new variable `$ctx` that will be bound to the context node.

```
for $ctx in self::node()
  return /descendant::*[self::node() == $ctx]/r
```

The path r, where a reverse occurs, is now part of an absolute path and can be rewritten by "rare".

The reselection of the context node is a considerable overhead and should be avoided. To this end a rewriting using information from the host language is preferable such that relative location paths can be avoided.

Rewriting location paths with RR joins. For the second class a rewriting can also be achieved using variables.

Example 5.2. Consider the following location path with an RR join:

```
/descendant::name[preceding::editor = self::*].
```

This location path selects from the data of Figure 1 the `name` element with value "`anna`". In XPath 2.0 it can be rewritten in a reverse-axis-free location path using variables:

```
for $m in /descendant::editor
 return $m/following::name[self::node() == /descendant::name and $m  = self::*]
```

Now consider the general case of a location path with an RR join: Let p, p_1 be relative location paths, a_n and a_m axes with a_m a reverse axis, b_m the symmetrical axis of a_m, n and m node tests and $\theta \in \{=, ==\}$:

$$/p/a_n::n[a_m::m\ \theta\ p_1]$$

Following the idea of the general equivalences this will be rewritten to a location path first selecting any m node in the document, and then n nodes in the b_m axis that are also selected by $/p/a_n::n[\$m\ \theta\ p_1]$, where $\$m$ stands for a reference to the currently selected m node, so again variables are necessary for the rewriting:

```
for $m in /descendant::m
   return $m/b_m::n[self::node() == /p/a_n::n][$m θ p_1]
```

This technique is similar to the general equivalences and only adds the use of variables in the handling of the original join.

6 Related Work

Several methods have been proposed for rewriting XPath expressions taking integrity constraints or schemas into account [6, 26], and the equivalence and containment problems for XPath expressions have been investigated [9, 27]. Furthermore, a growing interest in query optimization for XML databases, including optimization of XPath expressions, recently emerged. To the best of our knowledge, however, no other approach has been proposed for removing reverse steps from XPath expressions relying upon XPath symmetry. Note that using equivalence preserving rewriting rules for removing reverse steps from XPath expressions, as it is proposed in the present paper, is not closely related to the general equivalence problem for XPath expressions.

In [5] redundancies in XPath expressions based on a "model-oriented" approach are investigated. Such an approach relies on an abstract model of XPath that views XPath expressions as tree patterns. [5] shows that redundant branches of a tree pattern can be eliminated in polynomial time. Tree patterns are more abstract than XPath expressions in a way which is relevant to the work described in the present paper: A same tree pattern represents multiple equivalent XPath expressions. In particular, the symmetries in XPath exploited in the present paper are absent from tree patterns. Tree patterns do not consider the document order and therefore the concept of forward and reverse steps. In some sense the present work shows in which cases this simplified view upon an XPath expression can be justified.

Stream-based query processing has gained considerable interest in the past few years, e.g. due to its application in data integration [10, 14] and in publish-subscribe architectures [4, 7]. They all consider a navigational approach (XML-QL or XPath) consisting of a restricted subset of forward axes from XPath. This fact contrasts with the present work, which enables the use of the unrestricted set of XPath axes in a stream-based context.

7 Conclusion

The main result of this paper consists in two rule sets, RuleSet_1 and RuleSet_2, used in an algorithm for transforming XPath 1.0 expressions containing reverse axes into reverse-axis-free equivalents. Both RuleSet_1 and RuleSet_2 have advantages and it is an open issue which one is preferable. The location paths generated using RuleSet_2 do not contain joins, in contrast with those generated using RuleSet_1, which contain the same number of joins as there are reverse steps in the input location path. However, the path rewriting using RuleSet_2 has an exponential complexity in the length of the input location path, in contrast with rewriting using RuleSet_1 which has only a linear complexity.

Closely related to the comparison of RuleSet_1 and RuleSet_2 we plan to investigate the notion of "minimality" or "simplicity" of XPath expressions. We are focusing on defining a notion of a minimal XPath expression that can be evaluated more efficiently in a stream-based context than its equivalents. A notion of minimality will allow for well-founded optimization techniques for XPath expressions.

We believe that the equivalences proposed in this paper are an important step towards an efficient progressive evaluation of full XPath. In particular, we show that it is not necessary to restrict the use of axes in XPath for progressive processing, as suggested in [8]. Based on the results of this paper we are designing and implementing a progressive XPath processor, that supports unrestricted XPath [12].

References

[1] *Astronomical Data Center*, Home page, `http://adc.gsfc.nasa.gov/`.

[2] *Cocoon 2.0: XML publishing framework*, `http://xml.apache.org/cocoon/index.html`.

[3] *Xalan-Java Version 2.2*, Apache Project, `http://xml.apache.org/xalan-j/index.html`.

[4] M. Altinel and M. Franklin, *Efficient Filtering of XML Documents for Selective Dissemination of Information*, Proc. of 26th Conference on Very Large Databases (VLDB), 2000.

[5] Sihem Amer-Yahia, SungRan Cho, Laks V. S. Lakshmanan, and Divesh Srivastava, *Minimization of tree pattern queries*, SIGMOD, 2001.

[6] K. Boehm, K. Gayer, T. Oezsu, and K. Aberer, *Query optimization for structured documents based on knowledge on the document type definition*, Proc. of the Advances in Digital Libraries Conference, 1998.

[7] C. Chan, P. Felber, M. Garofalakis, and R. Rastogi, *Efficient Filtering of XML Documents with XPath Expressions*, Proc. of International Conference on Data Engineering (ICDE), 2002.

[8] A. Desai, *Introduction to Sequential XPath*, Proc. of IDEAlliance XML Conference, 2001, `http://www.idealliance.org/papers/xml2001/papers/html/05-01-01.html`.

[9] A. Deutsch and V. Tannen, *Containment for classes of XPath expressions under integrity constraints*, Knowledge Representation meets Databases (KRDB), 2001.

[10] T. J. Green, M. Onizuka, and D. Suciu, *Processing XML Streams with Deterministic Automata and Stream Indexes*, Tech. report, University of Washington, 2001.

[11] N. Ide, P. Bonhomme, and L. Romary, *XCES: An XML-based standard for linguistic corpora*, Proc. of the Second Annual Conference on Language Resources and Evaluation, 2000.

[12] T. Kiesling, *Towards a streamed XPath evaluation*, 2002, `http://www.pms.informatik.uni-muenchen.de/lehre/projekt-diplom-arbeit/streamedxpath.html`.

[13] P. Kroeger, *Modeling of Biological Data*, Tech. report, University of Munich, 2001.

[14] A. Levy, Z. Ives, and D. Weld, *Efficient Evaluation of Regular Path Expressions on Streaming XML Data*, Tech. report, University of Washington, 2000.

[15] S. McGrath, *XPipe*, `http://xpipe.sourceforge.net/`.

[16] D. Megginson, *SAX: The Simple API for XML*, `http://www.saxproject.org/`.

[17] W3C, *XML Path Language (XPath) Version 1.0*, W3C Recommendation, 1999, `http://www.w3.org/TR/xpath`.

[18] W3C, *XSL Transformations (XSLT) Version 1.0*, W3C Recommendation, 1999.

[19] W3C, *Document Object Model (DOM) Level 2 Core Specification*, W3C Recommendation, 2000.

[20] W3C, *XML path language (XPath) version 2.0*, W3C Working Draft, 2001, `http://www.w3.org/TR/xpath20`.

[21] W3C, *XQuery 1.0: An XML query language*, W3C Working Draft, 2001, `http://www.w3.org/TR/xquery/`.

[22] W3C, *XQuery 1.0 and XPath 2.0 data model*, W3C Working Draft, 2001, `http://www.w3.org/TR/query-datamodel/`.

[23] W3C, *XSL Transformations (XSLT) Version 2.0*, W3C Working Draft, 2001, `http://www.w3.org/TR/xslt20`.

[24] P. Wadler, *A formal semantics of patterns in XSLT*, Proc. of Conference on Markup Technologies, 1999.

[25] P. Wadler, *Two semantics of XPath*, Tech. report, 2000.

[26] P. T. Wood, *Optimising web queries using document type definitions*, 2nd ACM Workshop on Web Information and Data Management (WIDM'99), 1999.

[27] P. T. Wood, *On the equivalence of XML patterns*, Proc. 6th Int. Conf. on Rules and Objects in Databases (DOOD), 2000.

A Proofs

In the following proofs some obvious, yet important properties are often used:

Remark A.1. Let x_0 and x_1 be nodes. If $x_0 \in \mathit{sibling}(x_1)$, then $\mathit{parent}(x_0) = \mathit{parent}(x_1)$. Hence also $\mathit{parent}^+(x_0) = \mathit{parent}^+(x_1)$.

Remark A.2. Let x_0, x_1 and x_2 be nodes. Then

$$x_0 \in \mathit{parent}(x_1) \wedge x_1 \in \mathit{children}^+(x_2) \leftrightarrow x_0 \in \mathit{parent}(x_1) \wedge x_0 \in \mathit{children}^*(x_2)$$

Remark A.3. Let x_0, x_1 and x_2 be nodes. If $x_0 \in \mathit{parent}^+(x_1)$ and $x_1 \in \mathit{children}(x_2)$, then $x_0 \in \mathit{parent}^*(x_2)$.

Remark A.4. Let x_0, x_1, and x_2 be nodes. If x_2 is a descendant of x_0 and x_1, either x_0 is an descendant of x_1 or x_1 is an descendant of x_0 or $x_1 = x_0$ (treated in the first disjunctive term):

$$x_2 \in \mathit{children}^+(x_0) \wedge x_2 \in \mathit{children}^+(x_1) \text{ iff}$$
$$(x_2 \in \mathit{children}^+(x_0) \wedge x_0 \in \mathit{children}^*(x_1)) \vee (x_2 \in \mathit{children}^+(x_1) \wedge x_1 \in \mathit{children}^+(x_0))$$

Remark A.5. Let x_0, x_1, and x_2 be nodes.

$$x_0 \in \mathit{sibling}(x_1) \wedge x_0\, \mathit{children}(x_2) \rightarrow x_1 \in \mathit{children}(x_2)$$

The analogous propositions do also hold for *following-sibling*, *preceding-sibling*, $\mathit{children}^+$, and $\mathit{children}^*$.

Remark A.6. Let n be a node test and p a non-empty path. Then

$$p/\texttt{self::}n \equiv p\texttt{[self::}n\texttt{]}$$

Remark A.7. Let q_1 and q_2 be qualifiers and p a non-empty path. Then

$$p\texttt{[}q_1\texttt{]}\texttt{[}q_2\texttt{]} \equiv p\texttt{[}q_2\texttt{]}\texttt{[}q_1\texttt{]}$$

Lemma 3.1 Let x be the context node.

1. Right step adjunction:

$$\mathcal{S}[\![p/q]\!]x = \{x_2 | x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathcal{S}[\![q]\!]x_1\} = \{x_2 | x_1 \in \mathcal{S}[\![p']\!]x \wedge x_2 \in \mathcal{S}[\![q]\!]x_1\} = \mathcal{S}[\![p'/q]\!]x$$

2. Left step adjunction:

$$\mathcal{S}[\![q/p]\!]x = \{x_2 | x_1 \in \mathcal{S}[\![q]\!]x \wedge x_2 \in \mathcal{S}[\![p]\!]x_1\} = \{x_2 | x_1 \in \mathcal{S}[\![q]\!]x \wedge x_2 \in \mathcal{S}[\![p']\!]x_1\} = \mathcal{S}[\![q/p']\!]x$$

3. Qualifier adjunction:

$$\mathcal{S}[\![p\texttt{[}a\texttt{]}]\!]x = \{x_1 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathcal{Q}[\![a]\!]x_1\} = \{x_1 | x_1 \in \mathcal{S}[\![q]\!]x \wedge \mathcal{Q}[\![a]\!]x_1\} = \mathcal{S}[\![q\texttt{[}a\texttt{]}]\!]x.$$

4. Relative/absolute path conversion:

$$\mathcal{S}[\![/p]\!]x = \mathcal{S}[\![p]\!](\mathit{root}(x)) = \mathcal{S}[\![p']\!](\mathit{root}(x)) = \mathcal{S}[\![/p']\!]x.$$

5. Union/disjunction conversion:

$$\begin{aligned}
\mathcal{S}[\![p\texttt{[}q_1\texttt{]} \mid p\texttt{[}q_2\texttt{]}]\!]x &= \mathcal{S}[\![p\texttt{[}q_1\texttt{]}]\!]x \cup \mathcal{S}[\![p\texttt{[}q_2\texttt{]}]\!]x \\
&= \{x_1 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathcal{Q}[\![q_1]\!]x_1\} \cup \{x_2 | x_2 \in \mathcal{S}[\![p]\!]x \wedge \mathcal{Q}[\![q_2]\!]x_2\} \\
&= \{x_3 | x_3 \in (\mathcal{S}[\![p]\!]x \cup \mathcal{S}[\![p]\!]x) \wedge \mathcal{Q}[\![q_1]\!]x_3 \vee \mathcal{Q}[\![q_2]\!]x_3\} = \{x_3 \mid x_3 \in \mathcal{S}[\![p]\!]x, \mathcal{Q}[\![q_1 \texttt{ or } q_2]\!]x_3\} \\
&= \mathcal{S}[\![p\texttt{[}q_1 \texttt{ or } q_2\texttt{]}]\!]x.
\end{aligned}$$

6. Qualifier flattening:

$$\begin{aligned}
\mathcal{S}[\![p\texttt{[}p_1\texttt{[}p_2\texttt{]]}]\!]x &= \{x_1 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathcal{Q}[\![p_1\texttt{[}p_2\texttt{]}]\!]x_1\} = \{x_1 | x_1 \in \mathcal{S}[\![p]\!]x \wedge (\mathcal{S}[\![p_1\texttt{[}p_2\texttt{]}]\!]x_1 \neq \emptyset)\} \\
&= \{x_1 | x_1 \in \mathcal{S}[\![p]\!]x \wedge (\{x_2 | x_2 \in \mathcal{S}[\![p_1]\!]x_1 \wedge (\mathcal{S}[\![p_2]\!]x_2 \neq \emptyset)\} \neq \emptyset)\} \\
&= \{x_1 | x_1 \in \mathcal{S}[\![p]\!]x \wedge (\{x_2 | x_2 \in \mathcal{S}[\![p_1]\!]x_1 \wedge x_3 \in \mathcal{S}[\![p_2]\!]x_2\} \neq \emptyset)\} \\
&= \{x_1 | x_1 \in \mathcal{S}[\![p]\!]x \wedge (\mathcal{S}[\![p_1/p_2]\!]x_1 \neq \emptyset)\} \\
&= \mathcal{S}[\![p\texttt{[}p_1/p_2\texttt{]}]\!]x.
\end{aligned}$$

7. Decomposition of `ancestor-or-self` axis:

$$\begin{aligned}\mathcal{S}[\![\texttt{ancestor-or-self::}n]\!]x &= \{x_1 | x_1 \in \mathit{parent}^*(x) \wedge \mathit{nodetest}(x_1, n)\} \\ &= \{x_1 | (x_1 \in \mathit{parent}^+(x) \vee x_1 = x) \wedge \mathit{nodetest}(x_1, n)\} \\ &= \{x_1 | x_1 \in \mathit{parent}^+(x) \wedge \mathit{nodetest}(x_1, n)\} \cup \{x_1 | x_1 = x, \mathit{nodetest}(x_1, n)\} \\ &= \mathcal{S}[\![\texttt{self::}n]\!]x \cup \mathcal{S}[\![\texttt{ancestor::}n]\!]x \\ &= \mathcal{S}[\![\texttt{self::}n \texttt{ | ancestor::}n]\!]x.\end{aligned}$$

8. Decomposition of `descendant-or-self` axis:

$$\begin{aligned}\mathcal{S}[\![\texttt{descendant-or-self::}n]\!]x &= \{x_1 | x_1 \in \mathit{children}^*(x) \wedge \mathit{nodetest}(x_1, n)\} \\ &= \{x_1 | (x_1 \in \mathit{children}^+(x) \vee x_1 = x) \wedge \mathit{nodetest}(x_1, n)\} \\ &= \{x_1 | x_1 \in \mathit{children}^+(x) \wedge \mathit{nodetest}(x_1, n)\} \cup \{x_1 | x_1 = x, \mathit{nodetest}(x_1, n)\} \\ &= \mathcal{S}[\![\texttt{self::}n]\!]x \cup \mathcal{S}[\![\texttt{descendant::}n]\!]x \\ &= \mathcal{S}[\![\texttt{self::}n \texttt{ | descendant::}n]\!]x.\end{aligned}$$

9. Union term adjunction:

$$\mathcal{S}[\![p_1 \texttt{ | } p_3]\!]x = \mathcal{S}[\![p_1]\!]x \cup \mathcal{S}[\![p_3]\!]x = \mathcal{S}[\![p_2]\!]x \cup \mathcal{S}[\![p_3]\!]x = \mathcal{S}[\![p_2 \texttt{ | } p_3]\!]x.$$

10. Union commutativity:

$$\mathcal{S}[\![p_1 \texttt{ | } p_2]\!]x = \mathcal{S}[\![p_1]\!]x \cup \mathcal{S}[\![p_2]\!]x = \mathcal{S}[\![p_2]\!]x \cup \mathcal{S}[\![p_1]\!]x = \mathcal{S}[\![p_2 \texttt{ | } p_1]\!]x.$$

□

Lemma 3.2

- First consider the `self` axis: `/self::node()` selects the document root, therefore `/self::node()` ≡ `/`. As `node()` is the only node test matching the document root (it is neither an element nor a text node), $\mathcal{S}[\![\texttt{/self::}n]\!] = \emptyset$ for $n \neq \texttt{node()}$. If $a \neq \texttt{self}$ the result of `/`a`::`n is the empty set, since the document root is the root of the document tree (cf. Figure 1), i.e. has neither a parent nor any siblings.
- Let $a = \texttt{ancestor}$, $n = \texttt{node()}$, and x be the context node.

$$\begin{aligned}&\mathcal{S}[\![\texttt{/child::}m\texttt{/ancestor::node()}]\!]x \\ &= \mathcal{S}[\![\texttt{child::}m\texttt{/ancestor::node()}]\!]\mathit{root}(x) \\ &= \{x_2 | x_2 \in \mathcal{S}[\![\texttt{ancestor::node()}]\!]x_1 \wedge x_1 \in \mathcal{S}[\![\texttt{child::}m]\!]\mathit{root}(x)\} \\ &= \{x_2 | x_2 \in \mathit{parent}^+(x_1) \wedge \mathit{nodetest}(x_2, \texttt{node()}) \wedge x_1 \in \mathit{children}(\mathit{root}(x)) \wedge \mathit{nodetest}(x_1, m)\} \\ &= \{x_2 | x_2 = \mathit{root}(x) \wedge \mathit{nodetest}(x_2, \texttt{node()}) \wedge x_1 \in \mathit{children}(\mathit{root}(x)) \wedge \mathit{nodetest}(x_1, m)\} \\ &= \{x_2 | x_2 \in \mathcal{S}[\![\texttt{self::node()}]\!]x_1 \wedge x_1 \in \mathcal{S}[\![\texttt{child::}m]\!]\mathit{root}(x)\} \\ &= \mathcal{S}[\![\texttt{/self::node()[child::}m\texttt{]}]\!]x\end{aligned}$$

  In the other cases the result is the empty set since the document root has no ancestors nor siblings. The proof for `/child::`m`[a::`n`]` is analogous. □

Equivalence 1: Let x be the context node.

$$
\begin{aligned}
&\mathcal{S}[\![p\texttt{[/descendant::}m\texttt{[}s\texttt{]/}b_m\texttt{::node() == self::node()]}]\!]x = \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![p]\!]x \wedge \mathcal{Q}[\![\texttt{/descendant::}m\texttt{[}s\texttt{]/}b_m\texttt{::node() == self::node()}]\!]x_0\} \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![p]\!]x \wedge \{x_1 | x_1 \in \mathcal{S}[\![\texttt{/descendant::}m\texttt{[}s\texttt{]/}b_m\texttt{::node()}]\!]x_0 \wedge \\
&\qquad\qquad x_1 \in \mathcal{S}[\![\texttt{self::node()}]\!]x_0\} \neq \emptyset\} \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![p]\!]x \wedge \{x_1 | x_1 = x_0 \wedge x_1 \in \mathcal{S}[\![b_m\texttt{::node()}]\!]x_2 \wedge \\
&\qquad\qquad x_2 \in \mathcal{S}[\![\texttt{descendant::}m\texttt{[}s\texttt{]}]\!]root(x_0)\} \neq \emptyset\} \\
&\quad \stackrel{*}{=} \{x_0 | x_0 \in \mathcal{S}[\![p]\!]x \wedge \{x_1 | x_1 = x_0 \wedge x_2 \in \mathcal{S}[\![a_m\texttt{::node()}]\!]x_1 \wedge x_2 \in children^+(root(x_0)) \wedge \\
&\qquad\qquad nodetest(x_2, m) \wedge \mathcal{Q}[\![s]\!]x_2\} \neq \emptyset\} \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![p]\!]x \wedge \{x_1 | x_1 = x_0 \wedge x_2 \in \mathcal{S}[\![a_m\texttt{::}m\texttt{[}s\texttt{]}]\!]x_1\} \neq \emptyset\} \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![p]\!]x \wedge \mathcal{Q}[\![a_m\texttt{::}m\texttt{[}s\texttt{]}]\!]x_0\} \\
&\quad = \mathcal{S}[\![p\texttt{[}a_m\texttt{::}m\texttt{/}s\texttt{]}]\!]x
\end{aligned}
$$

Equality $(*)$ holds as b_m is the symmetrical axis to a_m. □

Equivalence 2: Let x be the context node.

$$
\begin{aligned}
&\mathcal{S}[\![\texttt{/descendant} :: m[b_m :: n == /p/a_n :: n]]\!]x = \\
&\quad = \{x_1 | x_1 \in \mathcal{S}[\![\texttt{descendant} :: m]\!]root(x) \wedge \mathcal{Q}[\![b_m :: n == /p/a_n :: n]\!]x_1\} \\
&\quad = \{x_1 | x_1 \in children^+(root(x)) \wedge nodetest(x_1, m) \wedge x_2 \in \mathcal{S}[\![b_m :: n]\!]x_1 \wedge x_2 \in \mathcal{S}[\![/p/a_n :: n]\!]x_1\} \\
&\quad = \{x_1 | x_1 \in children^+(root(x)) \wedge nodetest(x_1, m) \wedge x_2 \in \mathcal{S}[\![b_m :: node()]\!]x_1 \wedge \\
&\qquad\quad nodetest(x_2, n) \wedge x_2 \in \mathcal{S}[\![/p/a_n :: n]\!]x_1\} \\
&\quad \stackrel{*}{=} \{x_1 | x_1 \in children^+(root(x)) \wedge nodetest(x_1, m) \wedge x_1 \in \mathcal{S}[\![a_m :: node()]\!]x_2 \wedge \\
&\qquad\quad nodetest(x_2, n) \wedge x_2 \in \mathcal{S}[\![/p/a_n :: n]\!]x_1\} \\
&\quad = \{x_1 | x_1 \in children^+(root(x)) \wedge x_1 \in \mathcal{S}[\![a_m :: m]\!]x_2 \wedge x_2 \in \mathcal{S}[\![/p/a_n :: n]\!]x_1\} \\
&\quad = \{x_1 | x_1 \in \mathcal{S}[\![/p/a_n :: n/a_m :: m]\!]x_2\} \\
&\quad \stackrel{\dagger}{=} \mathcal{S}[\![/p/a_n :: n/a_m :: m]\!]x
\end{aligned}
$$

Equality $(*)$ holds as a_m is the symmetrical axis to b_m. Equality $(\dagger)$ holds as the semantics of an absolute path does not depend on the context node ($root(x) = root(x_2)$). □

Equivalence 2a: The proof is analogous to the proof for Equivalence 2.

Equivalence 3: Let x be the context node.

$$
\begin{aligned}
&\mathcal{S}[\![\texttt{descendant::}n\texttt{/parent::}m]\!]x \\
&\quad = \{x_2 | x_1 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x \wedge x_2 \in \mathcal{S}[\![\texttt{parent::}m]\!]x_1\} \\
&\quad = \{x_2 | x_2 \in parent(x_1) \wedge nodetest(x_2, m) \wedge x_1 \in children^+(x) \ \wedge nodetest(x_1, n)\} \\
&\stackrel{\dagger}{=} \{x_2 | x_2 \in children^*(x) \wedge nodetest(x_2, m) \wedge x_2 \in parent(x_1) \ \wedge nodetest(x_1, n)\} \\
&\quad = \{x_2 | x_2 \in children^*(x) \wedge nodetest(x_2, m) \wedge x_1 \in children(x_2) \wedge nodetest(x_1, n)\} \\
&\quad = \mathcal{S}[\![\texttt{descendant-or-self::}m\texttt{[child::}n\texttt{]}]\!]x.
\end{aligned}
$$

Equality $(\dagger)$ holds by Remark A.2. □

Equivalence 4: Let x be the context node.

$$\begin{aligned}
&\mathcal{S}[\![\texttt{child::}n\texttt{/parent::}m]\!]x \\
&\quad = \{x_2 | x_1 \in \mathcal{S}[\![\texttt{child::}n]\!]x \wedge x_2 \in \mathcal{S}[\![\texttt{parent::}m]\!]x_1\} \\
&\quad = \{x_2 | x_2 \in \mathit{parent}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}(x) \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \{x_2 | x_2 = x \wedge \mathit{nodetest}(x_2, m) \wedge x_2 \in \mathit{parent}(x_1) \ \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \{x_2 | x_2 = x \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}(x_2) \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \mathcal{S}[\![\texttt{self::}m\texttt{[child::}n\texttt{]}]\!]x.
\end{aligned}$$

□

Equivalence 5 holds with Lemma A.6. □

Equivalence 6: Let x be the context node.

$$\begin{aligned}
&\mathcal{S}[\![p\texttt{/following-sibling::}n\texttt{/parent::}m]\!]x \\
&\quad = \{x_2 | x_3 \in \mathcal{S}[\![p]\!]x \wedge x_1 \in \mathcal{S}[\![\texttt{following-sibling::}n]\!]x_3 \wedge x_2 \in \mathcal{S}[\![\texttt{parent::}m]\!]x_1\} \\
&\quad = \{x_2 | x_2 \in \mathit{parent}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{following\text{-}sibling}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n)\} \\
&\overset{*}{=} \{x_2 | x_2 \in \mathit{parent}(x_3) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{following\text{-}sibling}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/parent::}m]\!]x
\end{aligned}$$

Equality $*$ holds by Remark A.5. □

Equivalence 7: Let x be the context node.

$$\begin{aligned}
&\mathcal{S}[\![p\texttt{/following::}n\texttt{/parent::}m]\!]x \\
&\quad = \{x_2 | x_3 \in \mathcal{S}[\![p]\!]x \wedge x_1 \in \mathcal{S}[\![\texttt{following::}n]\!]x_3 \wedge x_2 \in \mathcal{S}[\![\texttt{parent::}m]\!]x_1\} \\
&\quad = \{x_2 | x_2 \in \mathit{parent}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{following}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \{x_2 | x_2 \in \mathit{parent}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge \\
&\qquad\qquad x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \{x_2 | (x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge \\
&\qquad\qquad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x)) \wedge (x_2 \notin \mathit{parent}^*(x_3) \vee x_2 \in \mathit{parent}^*(x_3))\} \\
&\quad = \{x_2 | ((x_2 \notin \mathit{parent}^*(x_3) \wedge x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3)) \vee \\
&\qquad\qquad (x_2 \in \mathit{parent}^*(x_3) \wedge x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3))) \wedge \\
&\qquad\qquad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
&\qquad\qquad \overset{*}{=} \{x_2 | ((x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_2 \ \wedge x_2 \in \mathit{branch\text{-}sibling}(x_3)) \vee \\
&\qquad\qquad (x_2 \in \mathit{parent}^*(x_3) \wedge x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3))) \wedge \\
&\qquad\qquad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
&\quad = \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[child::}n\texttt{]}]\!]x) \vee \\
&\qquad\qquad ((x_2 \in \mathit{parent}^+(x_3) \wedge x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3)) \vee \\
&\qquad\qquad (x_2 = x_3 \wedge x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3))) \wedge \\
&\qquad\qquad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
&\quad = \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[child::}n\texttt{]}]\!]x) \vee \\
&\qquad\qquad ((x_2 \in \mathit{parent}(x_5) \wedge x_5 \in \mathit{parent}^*(x_3) \wedge x_1 \in \mathit{sibling}(x_5) \wedge x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_1 \\
&\qquad\qquad\qquad\qquad \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3)) \vee \\
&\qquad\qquad \bot) \wedge \\
&\qquad\qquad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
&\overset{\dagger}{=} \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[child::}n\texttt{]}]\!]x) \vee
\end{aligned}$$

$$
\begin{aligned}
&((x_2 \in \mathit{parent}(x_5) \wedge x_5 \in \mathit{parent}^*(x_3) \wedge x_1 \in \mathit{sibling}(x_5) \wedge x_5 \ll x_1\ )) \wedge \\
&(\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
= \{x_2 | &(x_2 \in \mathcal{S}[\![p/\texttt{following::}m\texttt{[child::}n\texttt{]}]\!]x) \vee \\
&(x_2 \in \mathcal{S}[\![p/\texttt{ancestor-or-self::*[following-sibling::}n\texttt{]/parent::}m]\!]x) \\
= \mathcal{S}[\![&p/\texttt{following::}m\texttt{[child::}n\texttt{]}\ \texttt{|} \\
&p/\texttt{ancestor-or-self::*[following-sibling::}n\texttt{]/parent::}m]\!]x
\end{aligned}
$$

Equality $*$ holds due to the following:

$$
\begin{aligned}
&x_2 \notin \mathit{parent}^*(x_3) \wedge x_2 \in \mathit{parent}(x_1) \rightarrow (x_3 \ll x_1 \leftrightarrow x_3 \ll x_2) \\
&x_2 \in \mathit{parent}(x_1) \rightarrow (x_2 \notin \mathit{parent}^*(x_3) \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \leftrightarrow x_2 \in \mathit{branch\text{-}sibling}(x_3))
\end{aligned}
$$

Equality † holds due to the following:

$$
x_5 \in \mathit{parent}^*(x_3) \wedge x_5 \in \mathit{sibling}(x_1) \quad \rightarrow \quad (x_3 \ll x_1 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \leftrightarrow x_5 \ll x_1)
$$

□

Equivalence 8: Let x be the context node.

$$
\begin{aligned}
&\mathcal{S}[\![\texttt{descendant::}n\texttt{[parent::}m\texttt{]}]\!]x \\
&\quad = \{x_1 | x_1 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x \wedge \mathcal{Q}[\![\texttt{parent::}m]\!]x_1\} \\
&\quad = \{x_1 | x_2 \in \mathit{parent}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}^+(x) \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad \stackrel{\dagger}{=} \{x_1 | x_2 \in \mathit{children}^*(x) \wedge \mathit{nodetest}(x_2, m) \wedge x_2 \in \mathit{parent}(x_1) \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \{x_1 | x_2 \in \mathit{children}^*(x) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}(x_2) \wedge \mathit{nodetest}(x_1, n) \\
&\quad = \mathcal{S}[\![\texttt{descendant-or-self::}m\texttt{/child::}n]\!]x.
\end{aligned}
$$

Equality (†) holds by Remark A.2. □

Equivalence 9: Let x be the context node.

$$
\begin{aligned}
&\mathcal{S}[\![\texttt{child::}n\texttt{[parent::}m\texttt{]}]\!]x \\
&\quad = \{x_1 | x_1 \in \mathcal{S}[\![\texttt{child::}n]\!]x \wedge \mathcal{Q}[\![\texttt{parent::}m]\!]x_1\} \\
&\quad = \{x_1 | x_2 \in \mathit{parent}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}(x_) \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \{x_1 | x_1 \in \mathit{children}(x_2) \wedge x_2 = x \wedge \mathit{nodetest}(x_2, m) \wedge x_2 \in \mathit{parent}(x_1) \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \{x_1 | x_1 \in \mathit{children}(x) \wedge \mathit{nodetest}(x, m) \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \mathcal{S}[\![\texttt{self::}m\texttt{/child::}n]\!]x.
\end{aligned}
$$

Equivalence 10 holds with Lemmas A.6 and A.7. □

Equivalences 11 and 12: The proof is analogous to the proofs for Equivalences 6 and 7. □

Equivalence 13:

$$
\begin{aligned}
&\mathcal{S}[\![p\texttt{/descendant::}n\texttt{/ancestor::}m]\!]x \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_2 \wedge x_2 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x\} \\
&\quad = \{x_0 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_0, m) \wedge \mathit{nodetest}(x_2, n) \wedge x_0 \in \mathit{parent}^+(x_2) \wedge \\
&\qquad\quad x_2 \in \mathit{children}^+(x_1)\} \\
&\quad \stackrel{*}{=} \{x_0 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_0, m) \wedge \mathit{nodetest}(x_2, n) \wedge \big[(x_0 \in \mathit{parent}^+(x_1) \wedge \\
&\qquad\quad x_2 \in \mathit{children}^+(x_1)) \vee (x_0 \in \mathit{children}^*(x_1) \wedge x_2 \in \mathit{children}^+(x_0))\big]\} \\
&\quad = \{x_0 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \big[(x_0 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_1 \wedge x_2 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x_1) \vee \\
&\qquad\quad (x_0 \in \mathcal{S}[\![\texttt{descendant-or-self::}m]\!]x_1 \wedge x_2 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x_0)\big]\} \\
&\quad = \mathcal{S}[\![p\texttt{[descendant::}n\texttt{]/ancestor::}m\ \texttt{|}\ p\texttt{/descendant-or-self::}m\texttt{[descendant::}n\texttt{]}]\!]x
\end{aligned}
$$

The equality $(*)$ holds due to Remark A.4. □

Equivalence 13a:

$$
\begin{aligned}
&\mathcal{S}[\![\texttt{/descendant::}n\texttt{/ancestor::}m]\!]x \\
&\quad = \mathcal{S}[\![\texttt{descendant::}n\texttt{/ancestor::}m]\!]\mathit{root}(x) \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_2 \wedge x_2 \in \mathcal{S}[\![\texttt{descendant::}n]\!]\mathit{root}(x)\} \\
&\quad = \{x_0 | \mathit{nodetest}(x_0, m) \wedge \mathit{nodetest}(x_2, n) \wedge x_0 \in \mathit{parent}^+(x_2) \wedge \\
&\qquad\qquad x_2 \in \mathit{children}^+(\mathit{root}(x))\} \\
&\quad \stackrel{*}{=} \{x_0 | \mathit{nodetest}(x_0, m) \wedge \mathit{nodetest}(x_2, n) \wedge \big[(x_0 \in \mathit{parent}^+(\mathit{root}(x)) \wedge \\
&\qquad\qquad x_2 \in \mathit{children}^+(\mathit{root}(x))) \vee (x_0 \in \mathit{children}^*(\mathit{root}(x)) \wedge x_2 \in \mathit{children}^+(x_0))\big]\} \\
&\quad = \{x_0 | \bot \vee (x_0 \in \mathcal{S}[\![\texttt{descendant-or-self::}m]\!]\mathit{root}(x) \wedge x_2 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x_0)\} \\
&\quad = \mathcal{S}[\![\texttt{descendant::}n\texttt{/descendant-or-self::}m]\!]\mathit{root}(x) \\
&\quad = \mathcal{S}[\![\texttt{/descendant::}n\texttt{/descendant-or-self::}m]\!]x
\end{aligned}
$$

The equality $(*)$ holds due to Remark A.4. □

Equivalence 14:

$$
\begin{aligned}
&\mathcal{S}[\![p\texttt{/child::}n\texttt{/ancestor::}m]\!]x \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_2 \wedge x_2 \in \mathcal{S}[\![\texttt{child::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x\} \\
&\quad = \{x_0 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_0, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_0 \in \mathit{parent}^+(x_2) \wedge \\
&\qquad\qquad x_2 \in \mathit{children}(x_1)\} \\
&\quad \stackrel{*}{=} \{x_0 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_0, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_0 \in \mathit{parent}^*(x_1) \wedge \\
&\qquad\qquad x_2 \in \mathit{children}(x_1)\} \\
&\quad = \{x_0 | x_1 \in \mathcal{S}[\![p]\!]x \wedge x_0 \in \mathcal{S}[\![\texttt{ancestor-or-self::}m]\!]x_1 \wedge \mathcal{Q}[\![\texttt{child::}n]\!]x_1\} \\
&\quad = \mathcal{S}[\![p\texttt{[child::}n\texttt{]/ancestor-or-self::}m]\!]x
\end{aligned}
$$

The equality $(*)$ holds due to Remark A.3. □

Equivalence 15 holds with Lemma A.6. □

Equivalence 16:

$$
\begin{aligned}
&\mathcal{S}[\![p\texttt{/following-sibling::}n\texttt{/ancestor::}m]\!]x \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_2 \wedge x_2 \in \mathcal{S}[\![\texttt{following-sibling::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x\} \\
&\stackrel{*}{=} \{x_0 | x_1 \in \mathcal{S}[\![p]\!]x \wedge x_0 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_1 \wedge \mathcal{Q}[\![\texttt{following-sibling::}n]\!]x_1\} \\
&\quad = \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/ancestor::}m]\!]x
\end{aligned}
$$

The equality $(*)$ holds due to Remark A.1. □

Equivalence 17: Let x be the context node.

$$
\begin{aligned}
&\mathcal{S}[\![p/\texttt{following::}n/\texttt{ancestor::}m]\!]x \\
&\quad = \{x_2 | x_3 \in \mathcal{S}[\![p]\!]x \wedge x_1 \in \mathcal{S}[\![\texttt{following::}n]\!]x_3 \wedge x_2 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_1\} \\
&\quad = \{x_2 | x_2 \in \mathit{parent}^+(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{following}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \{x_2 | x_2 \in \mathit{parent}^+(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge \\
&\qquad\quad x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad = \{x_2 | (x_2 \in \mathit{parent}^+(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge \\
&\qquad\quad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x)) \wedge (x_2 \notin \mathit{parent}^*(x_3) \vee x_2 \in \mathit{parent}^*(x_3))\} \\
&\quad = \{x_2 | ((x_2 \notin \mathit{parent}^*(x_3) \wedge x_2 \in \mathit{parent}^+(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3)) \vee \\
&\qquad\quad (x_2 \in \mathit{parent}^*(x_3) \wedge x_2 \in \mathit{parent}^+(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3))) \wedge \\
&\qquad\quad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
&\quad \stackrel{*}{=} \{x_2 | ((x_2 \in \mathit{parent}^+(x_1) \wedge x_3 \ll x_2 \ \wedge x_2 \in \mathit{branch\text{-}sibling}(x_3)) \vee \\
&\qquad\quad (x_2 \in \mathit{parent}^*(x_3) \wedge x_1 \in \mathit{children}^+(x_2) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3))) \wedge \\
&\qquad\quad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
&\quad = \{x_2 | (x_2 \in \mathcal{S}[\![p/\texttt{following::}m\texttt{[descendant::}n\texttt{]}]\!]x) \vee \\
&\qquad\quad ((x_2 \in \mathit{parent}^+(x_3) \wedge x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3)) \vee \\
&\qquad\quad (x_2 = x_3 \wedge x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3))) \wedge \\
&\qquad\quad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
&\quad = \{x_2 | (x_2 \in \mathcal{S}[\![p/\texttt{following::}m\texttt{[descendant::}n\texttt{]}]\!]x) \vee \\
&\qquad\quad ((x_2 \in \mathit{parent}^+(x_3) \wedge x_2 \in \mathit{parent}(x_1) \wedge x_3 \ll x_1 \ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3)) \vee \\
&\qquad\quad \bot) \wedge \\
&\qquad\quad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
&\quad \stackrel{\dagger}{=} \{x_2 | (x_2 \in \mathcal{S}[\![p/\texttt{following::}m\texttt{[descendant::}n\texttt{]}]\!]x) \vee \\
&\qquad\quad ((x_2 \in \mathit{parent}^+(x_3) \wedge x_2 \in \mathit{parent}(x_1) \wedge x_4 \ll x_5 \ \wedge x_5 \in \mathit{sibling}(x_4) \wedge x_4 \in \mathit{parent}^*(x_3) \wedge \\
&\qquad\qquad x_1 \in \mathit{children}^*(x_5))) \\
&\qquad\quad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
&\quad = \{x_2 | (x_2 \in \mathcal{S}[\![p/\texttt{following::}m\texttt{[descendant::}n\texttt{]}]\!]x) \vee \\
&\qquad\quad ((x_4 \ll x_5 \ \wedge x_5 \in \mathit{sibling}(x_4) \wedge x_4 \in \mathit{parent}^*(x_3) \wedge x_1 \in \mathit{children}^*(x_5))) \wedge \\
&\qquad\quad (\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \in \mathcal{S}[\![p]\!]x))\} \\
&\quad = \{x_2 | (x_2 \in \mathcal{S}[\![p/\texttt{following::}m\texttt{[descendant::}n\texttt{]}]\!]x) \vee \\
&\qquad\quad (x_2 \in \mathcal{S}[\![p/\texttt{ancestor-or-self::*[following-sibling::*descendant-or-self::}n\texttt{]} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \texttt{/ancestor::}m]\!]x) \\
&\quad = \mathcal{S}[\![p/\texttt{following::}m\texttt{[descendant::}n\texttt{]} \ \texttt{|} \\
&\qquad\quad p\texttt{/ancestor-or-self::*[following-sibling::}n\texttt{]/parent::}m]\!]x
\end{aligned}
$$

Equality $*$ holds due to the following:

$$
\begin{aligned}
&x_2 \notin \mathit{parent}^*(x_3) \wedge x_2 \in \mathit{parent}(x_1) \rightarrow (x_3 \ll x_1 \leftrightarrow x_3 \ll x_2) \\
&x_2 \in \mathit{parent}(x_1) \rightarrow (x_2 \notin \mathit{parent}^*(x_3) \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \leftrightarrow x_2 \in \mathit{branch\text{-}sibling}(x_3))
\end{aligned}
$$

Equality † holds due to the following:

$$
\begin{aligned}
&x_3 \ll x_1 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \leftrightarrow \\
&\exists x_4, x_5 (x_4 \ll x_5 \wedge x_5 \in \mathit{sibling}(x_4) \wedge x_4 \in \mathit{parent}^*(x_3) \wedge x_1 \in \mathit{children}^*(x_5))
\end{aligned}
$$

□

Equivalence 18:

$$
\begin{aligned}
&\mathcal{S}[\![p/\texttt{descendant::}n\texttt{[ancestor::}m\texttt{]}]\!]x \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathcal{Q}[\![\texttt{ancestor::}m]\!]x_0\} \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_0\} \\
&\quad = \{x_0 | x_0 \in \mathit{children}^+(x_1) \wedge \mathit{nodetest}(x_1, n) \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathit{parent}^+(x_0) \wedge \\
&\qquad\quad \mathit{nodetest}(x_2, m)\} \\
&\quad \overset{*}{=} \{x_0 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_0, m) \wedge \mathit{nodetest}(x_1, n) \wedge \big[(x_2 \in \mathit{parent}^+(x_1) \wedge \\
&\qquad\quad x_0 \in \mathit{children}^+(x_1)) \vee (x_2 \in \mathit{children}^*(x_1) \wedge x_0 \in \mathit{children}^+(x_1))\big]\} \\
&\quad = \{x_0 | x_1 \in \mathcal{S}[\![p]\!]x \wedge \big[(x_2 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_1 \wedge x_0 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x_1) \\
&\qquad\quad \vee (x_2 \in \mathcal{S}[\![\texttt{descendant-or-self::}m]\!]x_1 \wedge x_0 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x_2)\big]\} \\
&\quad = \mathcal{S}[\![p\texttt{[ancestor::}m\texttt{]/descendant::}n \mid p\texttt{/descendant-or-self::}m\texttt{/descendant::}n]\!]x
\end{aligned}
$$

The equality $(*)$ holds due to Remark A.4. □

Equivalence 18a: The proof is analogous to the proof for Equivalence 13a. □

Equivalence 19:

$$
\begin{aligned}
&\mathcal{S}[\![p/\texttt{child::}n\texttt{[ancestor::}m\texttt{]}]\!]x \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{child::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathcal{Q}[\![\texttt{ancestor::}m]\!]x_0\} \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{child::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_0\} \\
&\quad = \{x_0 | x_0 \in \mathit{children}(x_1) \wedge \mathit{nodetest}(x_1, n) \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathit{parent}^+(x_0) \wedge \\
&\qquad\quad \mathit{nodetest}(x_2, m)\} \\
&\quad \overset{*}{=} \{x_0 | x_0 \in \mathit{children}(x_1) \wedge \mathit{nodetest}(x_1, n) \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathit{parent}^*(x_1) \wedge \\
&\qquad\quad \mathit{nodetest}(x_2, m)\} \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{child::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathcal{S}[\![\texttt{ancestor-or-self::}m]\!]x_1\} \\
&\quad = \mathcal{S}[\![p\texttt{[ancestor-or-self::}m\texttt{]/child::}n]\!]x
\end{aligned}
$$

The equality $(*)$ holds due to Remark A.3. □

Equivalence 20 holds with Lemmas A.6 and A.7. □

Equivalence 21:

$$
\begin{aligned}
&\mathcal{S}[\![p/\texttt{following-sibling::}n\texttt{[ancestor::}m\texttt{]}]\!]x \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{following-sibling::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge \mathcal{Q}[\![\texttt{ancestor::}m]\!]x_0\} \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{following-sibling::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_0\} \\
&\quad = \{x_0 | x_0 \in \mathit{following\text{-}sibling}(x_1) \wedge \mathit{nodetest}(x_1, n) \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathit{parent}^+(x_0) \wedge \\
&\qquad\quad \mathit{nodetest}(x_2, m)\} \\
&\quad \overset{*}{=} \{x_0 | x_0 \in \mathit{following\text{-}sibling}(x_1) \wedge \mathit{nodetest}(x_1, n) \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathit{parent}^+(x_1) \wedge \\
&\qquad\quad \mathit{nodetest}(x_2, m)\} \\
&\quad = \{x_0 | x_0 \in \mathcal{S}[\![\texttt{following-sibling::}n]\!]x_1 \wedge x_1 \in \mathcal{S}[\![p]\!]x \wedge x_2 \in \mathcal{S}[\![\texttt{ancestor::}m]\!]x_1\} \\
&\quad = \mathcal{S}[\![p\texttt{[ancestor::}m\texttt{]/following-sibling::}n]\!]x
\end{aligned}
$$

The equality $(*)$ holds due to Remark A.1. □

Equivalence 22: The proof is analogous to the proof for Equivalence 17. □

Equivalence 23: Let x be the context node.

$$
\begin{aligned}
&\mathcal{S}[\![\texttt{descendant::}n\texttt{/preceding-sibling::}m]\!]x\\
&\quad= \{x_2 | x_1 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x \wedge x_2 \in \mathcal{S}[\![\texttt{preceding-sibling::}m]\!]x_1\}\\
&\quad= \{x_2 | x_2 \in \mathit{preceding\text{-}sibling}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}^+(x) \ \wedge \mathit{nodetest}(x_1, n)\}\\
&\quad= \{x_2 | x_1 \in \mathit{following\text{-}sibling}(x_2) \wedge \mathit{nodetest}(x_2, m) \wedge x_2 \in \mathit{children}^+(x) \ \wedge \mathit{nodetest}(x_1, n)\}\\
&\quad= \mathcal{S}[\![\texttt{descendant::}m\texttt{[following-sibling::}n\texttt{]}]\!]x
\end{aligned}
$$

□

Equivalence 24: Let x be the context node.

$$
\begin{aligned}
&\mathcal{S}[\![\texttt{child::}n\texttt{/preceding-sibling::}m]\!]x\\
&\quad= \{x_2 | x_1 \in \mathcal{S}[\![\texttt{child::}n]\!]x \wedge x_2 \in \mathcal{S}[\![\texttt{preceding-sibling::}m]\!]x_1\}\\
&\quad= \{x_2 | x_2 \in \mathit{preceding\text{-}sibling}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}(x) \wedge \mathit{nodetest}(x_1, n)\}\\
&\quad= \{x_2 | x_1 \in \mathit{following\text{-}sibling}(x_2) \wedge \mathit{nodetest}(x_2, m) \wedge x_2 \in \mathit{children}(x) \ \wedge \mathit{nodetest}(x_1, n)\}\\
&\quad= \mathcal{S}[\![\texttt{child::}m\texttt{[following-sibling::}n\texttt{]}]\!]x
\end{aligned}
$$

□

Equivalence 25 holds with Lemma A.6. □

Equivalence 26: Let x be the context node.

$$
\begin{aligned}
&\mathcal{S}[\![p\texttt{/following-sibling::}n\texttt{/preceding-sibling::}m]\!]x\\
&\quad= \{x_2 | x_3 \in \mathcal{S}[\![p]\!]x \wedge x_1 \in \mathcal{S}[\![\texttt{following-sibling::}n]\!]x_3 \wedge x_2 \in \mathcal{S}[\![\texttt{preceding-sibling::}m]\!]x_1\}\\
&\quad= \{x_2 | x_2 \in \mathit{preceding\text{-}sibling}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{following\text{-}sibling}(x_3) \wedge\\
&\qquad\quad x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n)\}\\
&\quad= \{x_2 | x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_3) \wedge\\
&\qquad\quad x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n)\}\\
&\quad= \{x_2 | x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge\\
&\qquad\quad \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge (x_2 = x_3 \vee x_2 \ll x_3 \vee x_3 \ll x_2)\}\\
&\quad= \{x_2 | ((x_2 = x_3 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x) \vee\\
&\qquad\quad (x_2 \ll x_3 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x) \vee\\
&\qquad\quad (x_3 \ll x_2 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x)) \wedge\\
&\qquad\quad \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m)\}\\
&\quad= \{x_2 | ((x_2 = x_3 \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_2) \wedge x_3 \in \mathcal{S}[\![p]\!]x) \vee\\
&\qquad\quad (x_2 \ll x_3 \wedge x_2 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x) \vee\\
&\qquad\quad (x_3 \ll x_2 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x)) \wedge\\
&\qquad\quad \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m)\}\\
&\quad= \{x_2 | (x_2 = x_3 \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_2) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m)) \vee\\
&\qquad\quad (x_2 \ll x_3 \wedge x_2 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge\\
&\qquad\quad \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\\
&\qquad\quad \vee\\
&\qquad\quad (x_3 \ll x_2 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge\\
&\qquad\quad \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}\\
&\quad= \{x_2 | x_2 \in \mathcal{S}[\![p\texttt{[self::}m\texttt{/following-sibling::}n\texttt{]}]\!]x \vee\\
&\qquad\quad x_2 \in \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/preceding-sibling::}m]\!]x \vee\\
&\qquad\quad x_2 \in \mathcal{S}[\![p\texttt{/following-sibling::}m\texttt{[following-sibling::}n\texttt{]}]\!]x
\end{aligned}
$$

$= \mathcal{S}[\![p\texttt{[self::}m\texttt{/following-sibling::}n\texttt{]} \mid p\texttt{[following-sibling::}n\texttt{]/preceding-sibling::}m$
$\quad \mid p\texttt{/following-sibling::}m\texttt{[following-sibling::}n\texttt{]}]\!]x\}$

□

Equivalence 27: Let x be the context node.

$\mathcal{S}[\![p\texttt{/following::}n\texttt{/preceding-sibling::}m]\!]x$
$= \{x_2 | x_2 \in \mathcal{S}[\![\texttt{preceding-sibling::}m]\!]x_1 \wedge x_1 \in \mathcal{S}[\![\texttt{following::}n]\!]x_3 \wedge x_3 \in \mathcal{S}[\![p]\!]x\}$
$= \{x_2 | x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge$
$\quad (x_3 \ll x_2 \vee x_2 \ll x_3 \vee x_2 = x_3) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m)\}$
$= \{x_2 | ((x_3 \ll x_2 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3)) \vee$
$\quad ((x_2 \ll x_3 \vee x_2 = x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3))) \wedge$
$\quad (x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$
$\stackrel{*}{=} \{x_2 | ((x_3 \ll x_2 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_2 \in \mathit{branch\text{-}sibling}(x_3)) \vee$
$\quad ((x_2 \ll x_3 \vee x_2 = x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3))) \wedge$
$\quad (x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$
$= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[following-sibling::}n\texttt{]}]\!]x) \vee$
$\quad ((x_2 \ll x_3 \vee x_2 = x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3)) \wedge$
$\quad (x_2 \notin \mathit{parent}^*(x_3) \vee x_2 \in \mathit{parent}^*(x_3)) \wedge (x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$
$= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[following-sibling::}n\texttt{]}]\!]x) \vee$
$\quad ((x_2 \notin \mathit{parent}^*(x_3) \wedge (x_2 \ll x_3 \vee x_2 = x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1$
$\qquad \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3)) \vee$
$\quad (x_2 \in \mathit{parent}^*(x_3) \wedge (x_2 \ll x_3 \vee x_2 = x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1$
$\qquad \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3))) \wedge$
$\quad \wedge (x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$
$\stackrel{\dagger}{=} \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[following-sibling::}n\texttt{]}]\!]x) \vee$
$\quad ((x_4 \in \mathit{parent}^*(x_3) \wedge x_2 \ll x_4 \wedge x_4 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_4) \wedge x_1 \in \mathit{sibling}(x_4)) \vee$
$\quad (x_2 \in \mathit{parent}^*(x_3) \wedge (x_2 \ll x_3 \vee x_2 = x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1$
$\qquad \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3))) \wedge$
$\quad \wedge (x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$
$= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[following-sibling::}n\texttt{]}]\!]x \vee$
$\quad x_2 \in \mathcal{S}[\![p\texttt{/ancestor-or-self::*[following-sibling::}n\texttt{]/preceding-sibling::}m]\!]x) \vee$
$\quad (x_2 \in \mathit{parent}^*(x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge$
$\quad \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$
$= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[following-sibling::}n\texttt{]}]\!]x \vee$
$\quad x_2 \in \mathcal{S}[\![p\texttt{/ancestor-or-self::*[following-sibling::}n\texttt{]/preceding-sibling::}m]\!]x) \vee$
$\quad (x_2 \in \mathit{parent}^*(x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge$
$\qquad \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$
$= \{x_2 | x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[following-sibling::}n\texttt{]}]\!]x \vee$
$\quad x_2 \in \mathcal{S}[\![p\texttt{/ancestor-or-self::*[following-sibling::}n\texttt{]/preceding-sibling::}m]\!]x \vee$
$\quad x_2 \in \mathcal{S}[\![p\texttt{/ancestor-or-self::}m\texttt{[following-sibling::}n\texttt{]}]\!]x\}$
$= \mathcal{S}[\![p\texttt{/following::}m\texttt{[following-sibling::}n\texttt{]} \mid$
$\quad p\texttt{/ancestor-or-self::*[following-sibling::}n\texttt{]/preceding-sibling::}m \mid$
$\quad p\texttt{/ancestor-or-self::}m\texttt{[following-sibling::}n\texttt{]}]\!]x\}$

The equality $*$ holds due to the following:

$$x_3 \ll x_2 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \rightarrow \\ (x_3 \ll x_1 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \leftrightarrow x_2 \in \mathit{branch\text{-}sibling}(x_3))$$

Equality † holds because of

$$(x_2 \notin \mathit{parent}^*(x_3) \wedge (x_2 \ll x_3 \vee x_2 = x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_1) \wedge x_3 \ll x_1 \\ \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3)) \\ \leftrightarrow (x_4 \in \mathit{parent}^*(x_3) \wedge x_2 \ll x_4 \wedge x_4 \ll x_1 \wedge x_2 \in \mathit{sibling}(x_4) \wedge x_1 \in \mathit{sibling}(x_4))$$

□

Equivalences 28 and 29: The proof is analogous to the proofs for Equivalences 23 and 24. □

Equivalence 30 holds with Lemmas A.6 and A.7. □

Equivalences 31 and 32: The proof is analogous to the proofs for Equivalences 26 and 27. □

Equivalence 33: Let x be the context node.

$$\begin{aligned}
&\mathcal{S}[\![p/\texttt{descendant::}n/\texttt{preceding::}m]\!]x \\
&\quad = \{x_2 | x_3 \in \mathcal{S}[\![p]\!]x \wedge x_1 \in \mathcal{S}[\![\texttt{descendant::}n]\!]x_3 \wedge x_2 \in \mathcal{S}[\![\texttt{preceding::}m]\!]x_1\} \\
&\quad = \{x_2 | x_2 \in \mathit{preceding}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}^+(x_3) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}[\![p]\!]x\} \\
&\quad = \{x_2 | x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}^+(x_3) \wedge \mathit{nodetest}(x_1, n) \wedge \\
&\qquad\qquad x_3 \in \mathcal{S}[\![p]\!]x\} \\
&\quad = \{x_2 | x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}^+(x_3) \wedge \\
&\qquad (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}[\![p]\!]x) \wedge (x_2 \ll x_3 \vee x_3 \ll x_2 \vee x_2 = x_3)\} \\
&\quad = \{x_2 | ((x_2 \ll x_3 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}^+(x_3)) \vee \\
&\qquad (x_3 \ll x_2 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}^+(x_3)) \vee \\
&\qquad (x_2 = x_3 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}^+(x_3))) \wedge \\
&\qquad (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}[\![p]\!]x)\} \\
&\quad = \{x_2 | ((x_2 \ll x_3 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_1 \in \mathit{children}^+(x_3)) \vee \\
&\qquad (x_3 \ll x_2 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}^*(x_4) \wedge x_4 \in \mathit{children}(x_3)) \vee \\
&\qquad \bot) \wedge \\
&\qquad (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}[\![p]\!]x)\} \\
&\quad = \{x_2 | x_2 \in \mathcal{S}[\![p\texttt{[descendant::}n\texttt{]/preceding::}m]\!]x \vee \\
&\qquad (x_3 \ll x_2 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}^*(x_4) \wedge x_4 \in \mathit{children}(x_3) \wedge \\
&\qquad x_5 \in \mathit{children}(x_3) \wedge x_2 \in \mathit{children}^*(x_5) \wedge (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}[\![p]\!]x)\} \\
&\quad = \{x_2 | x_2 \in \mathcal{S}[\![p\texttt{[descendant::}n\texttt{]/preceding::}m]\!]x \vee \\
&\qquad (x_3 \ll x_2 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}^*(x_4) \wedge x_4 \in \mathit{children}(x_3) \wedge \\
&\qquad x_5 \in \mathit{children}(x_3) \wedge x_2 \in \mathit{children}^*(x_5) \wedge x_5 \ll x_4 \wedge (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge \\
&\qquad\qquad x_3 \in \mathcal{S}[\![p]\!]x)\} \\
&\quad = \{x_2 | x_2 \in \mathcal{S}[\![p\texttt{[descendant::}n\texttt{]/preceding::}m]\!]x \vee \\
&\qquad (x_1 \in \mathit{children}^*(x_4) \wedge x_4 \in \mathit{children}(x_3) \wedge x_5 \in \mathit{children}(x_3) \wedge x_2 \in \mathit{children}^*(x_5) \wedge x_5 \ll x_4 \wedge \\
&\qquad (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}[\![p]\!]x)\} \\
&\quad = \{x_2 | x_2 \in \mathcal{S}[\![p\texttt{[descendant::}n\texttt{]/preceding::}m]\!]x \vee \\
&\qquad x_2 \in \mathcal{S}[\![p\texttt{/child::*[following-sibling::*/descendant-or-self::}n\texttt{]}]\!]
\end{aligned}$$

$$\texttt{/descendant-or-self::}m\llbracket x\}$$
$$= \mathcal{S}\llbracket p\texttt{[descendant::}n\texttt{]/preceding::}m \mid$$
$$p\texttt{/child::*[following-sibling::*/descendant-or-self::}n\texttt{]/descendant-or-self::}m\rrbracket x$$

□

Equivalence 33a: Let x be the context node.

$$\begin{aligned}
&\mathcal{S}\llbracket\texttt{/descendant::}n\texttt{/preceding::}m\rrbracket x \\
&\quad= \mathcal{S}\llbracket\texttt{descendant::}n\texttt{/preceding::}m\rrbracket \mathit{root}(x) \\
&\quad= \{x_2 | x_1 \in \mathcal{S}\llbracket\texttt{descendant::}n\rrbracket \mathit{root}(x) \wedge x_2 \in \mathcal{S}\llbracket\texttt{preceding::}m\rrbracket x_1\} \\
&\quad= \{x_2 | x_2 \in \mathit{preceding}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}^+(\mathit{root}(x)) \wedge \mathit{nodetest}(x_1, n)\} \\
&\quad= \{x_2 | x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge \mathit{nodetest}(x_2, m) \wedge x_2 \in \mathit{children}^+(\mathit{root}(x)) \wedge \\
&\qquad\qquad \mathit{nodetest}(x_1, n)\} \\
&\quad= \mathcal{S}\llbracket\texttt{descendant::}m\texttt{[following::}n\texttt{]}\rrbracket \mathit{root}(x) \\
&\quad= \mathcal{S}\llbracket\texttt{/descendant::}m\texttt{[following::}n\texttt{]}\rrbracket x
\end{aligned}$$

□

Equivalence 34: Let x be the context node.

$$\begin{aligned}
&\mathcal{S}\llbracket p\texttt{/child::}n\texttt{/preceding::}m\rrbracket x \\
&\quad= \{x_2 | x_3 \in \mathcal{S}\llbracket p\rrbracket x \wedge x_1 \in \mathcal{S}\llbracket\texttt{child::}n\rrbracket x_3 \wedge x_2 \in \mathcal{S}\llbracket\texttt{preceding::}m\rrbracket x_1\} \\
&\quad= \{x_2 | x_2 \in \mathit{preceding}(x_1) \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}(x_3) \ \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}\llbracket p\rrbracket x\} \\
&\quad= \{x_2 | x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge \mathit{nodetest}(x_2, m) \wedge x_1 \in \mathit{children}(x_3) \ \wedge \mathit{nodetest}(x_1, n) \wedge \\
&\qquad\qquad x_3 \in \mathcal{S}\llbracket p\rrbracket x\} \\
&\quad= \{x_2 | x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}(x_3) \wedge \\
&\qquad (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}\llbracket p\rrbracket x) \wedge (x_2 \ll x_3 \vee x_3 \ll x_2 \vee x_2 = x_3)\} \\
&\quad= \{x_2 | ((x_2 \ll x_3 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}(x_3)) \vee \\
&\qquad (x_3 \ll x_2 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}(x_3)) \vee \\
&\qquad (x_2 = x_3 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}(x_3))) \wedge \\
&\qquad (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}\llbracket p\rrbracket x)\} \\
&\quad= \{x_2 | (x_2 \in \mathcal{S}\llbracket p\texttt{[child::}n\texttt{]/preceding::}m\rrbracket x) \vee \\
&\qquad (((x_3 \ll x_2 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}(x_3)) \vee \\
&\qquad \bot) \wedge \\
&\qquad (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}\llbracket p\rrbracket x))\} \\
&\quad\stackrel{*}{=} \{x_2 | (x_2 \in \mathcal{S}\llbracket p\texttt{[child::}n\texttt{]/preceding::}m\rrbracket x) \vee \\
&\qquad ((x_3 \ll x_2 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}(x_3) \wedge \\
&\qquad\qquad x_4 \in \mathit{children}(x_3) \wedge x_4 \ll x_1 \wedge x_2 \in \mathit{children}^*(x_4)) \wedge \\
&\qquad (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}\llbracket p\rrbracket x))\} \\
&\quad= \{x_2 | (x_2 \in \mathcal{S}\llbracket p\texttt{[child::}n\texttt{]/preceding::}m\rrbracket x) \vee \\
&\qquad ((x_1 \in \mathit{sibling}(x_4) \wedge x_4 \in \mathit{children}(x_3) \wedge x_4 \ll x_1 \wedge x_2 \in \mathit{children}^*(x_4)) \wedge \\
&\qquad (\mathit{nodetest}(x_2, m) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \in \mathcal{S}\llbracket p\rrbracket x))\} \\
&\quad= \{x_2 | x_2 \in \mathcal{S}\llbracket p\texttt{[child::}n\texttt{]/preceding::}m\rrbracket x \vee \\
&\qquad x_2 \in \mathcal{S}\llbracket p\texttt{/child::*[following-sibling::}n\texttt{]/descendant-or-self::}m\rrbracket x\} \\
&\quad= \mathcal{S}\llbracket p\texttt{[child::}n\texttt{]/preceding::}m \\
&\qquad \mid p\texttt{/child::*[following-sibling::}n\texttt{]/descendant-or-self::}m\rrbracket x\}
\end{aligned}$$

Equality $*$ holds due to

$$x_3 \ll x_2 \wedge x_2 \ll x_1 \wedge x_1 \in \mathit{children}^*(x_3) \rightarrow \exists x_4(x_4 \in \mathit{children}(x_3) \wedge x_4 \ll x_1 \wedge x_2 \in \mathit{children}^*(x_4))$$

□

Equivalence 35 follows from Lemma A.6.

Equivalence 36: Let x be the context node.

$$\begin{aligned}
&\mathcal{S}[\![p/\texttt{following-sibling::}n/\texttt{preceding::}m]\!]x\\
&\quad= \{x_2 | x_3 \in \mathcal{S}[\![p]\!]x \wedge x_1 \in \mathcal{S}[\![\texttt{following-sibling::}n]\!]x_3 \wedge x_2 \in \mathcal{S}[\![\texttt{preceding::}m]\!]x_1\}\\
&\quad= \{x_2 | x_3 \in \mathcal{S}[\![p]\!]x \wedge x_1 \in \mathit{sibling}(x_3) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge\\
&\qquad\qquad x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge \mathit{nodetest}(x_2, m)\}\\
&\quad= \{x_2 | x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge\\
&\qquad\qquad x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge (x_2 \ll x_3 \vee x_3 \ll x_2 \vee x_2 = x_3)\}\\
&\quad= \{x_2 | ((x_2 \ll x_3 \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1)) \vee\\
&\qquad\qquad ((x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1))) \wedge\\
&\qquad\qquad (x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}\\
&\quad= \{x_2 | ((x_2 \ll x_3 \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_3)) \vee\\
&\qquad\qquad ((x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1))) \wedge\\
&\qquad\qquad (x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}\\
&\quad= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/preceding::}m]\!]x) \vee\\
&\qquad\qquad ((x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge\\
&\qquad\qquad x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}\\
&\quad= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/preceding::}m]\!]x) \vee\\
&\qquad\qquad (((x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1)) \wedge\\
&\qquad\qquad (x_2 \notin \texttt{child}^*(x_3) \vee x_2 \in \texttt{child}^*(x_3)) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}\\
&\quad= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/preceding::}m]\!]x) \vee\\
&\qquad\qquad ((x_2 \notin \texttt{child}^*(x_3) \wedge (x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge\\
&\qquad\qquad\qquad\qquad x_2 \in \mathit{branch\text{-}sibling}(x_1)) \vee\\
&\qquad\qquad (x_2 \in \texttt{child}^*(x_3) \wedge (x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge\\
&\qquad\qquad\qquad\qquad x_2 \in \mathit{branch\text{-}sibling}(x_1)) \wedge\\
&\qquad\qquad \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}\\
&\quad= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/preceding::}m]\!]x) \vee\\
&\qquad\qquad ((x_2 \notin \texttt{child}^*(x_3) \wedge x_3 \ll x_2 \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1)) \vee\\
&\qquad\qquad (x_2 \in \texttt{child}^*(x_3) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1)) \wedge\\
&\qquad\qquad \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}\\
&\quad\overset{*}{=} \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/preceding::}m]\!]x) \vee\\
&\qquad\qquad ((x_2 \in \texttt{child}^*(x_4) \wedge x_3 \ll x_4 \wedge x_4 \in \mathit{sibling}(x_3) \wedge x_4 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_4)) \vee\\
&\qquad\qquad (x_2 \in \texttt{child}^*(x_3) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1)) \wedge\\
&\qquad\qquad \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}\\
&\quad= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/preceding::}m]\!]x \vee\\
&\qquad\qquad x_2 \in \mathcal{S}[\![p\texttt{/following-sibling::*[following-sibling::}n\texttt{]/descendant-or-self::}m]\!]x) \vee\\
&\qquad\qquad (x_2 \in \texttt{child}^*(x_3) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge\\
&\qquad\qquad \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}\\
&\quad= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/preceding::}m]\!]x \vee\\
&\qquad\qquad x_2 \in \mathcal{S}[\![p\texttt{/following-sibling::*[following-sibling::}n\texttt{]/descendant-or-self::}m]\!]x) \vee
\end{aligned}$$

$$(x_2 \in \texttt{child}^*(x_3) \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_3 \ll x_1 \wedge \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge$$
$$\mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$$
$$= \{x_2 | x_2 \in \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/preceding::}m]\!]x \vee$$
$$x_2 \in \mathcal{S}[\![p\texttt{/following-sibling::*[following-sibling::}n\texttt{]/descendant-or-self::}m]\!]x \vee$$
$$x_2 \in \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/descendant-or-self::}m]\!]x\}$$
$$= \mathcal{S}[\![p\texttt{[following-sibling::}n\texttt{]/preceding::}m \texttt{ |}$$
$$p\texttt{/following-sibling::*[following-sibling::}n\texttt{]/descendant-or-self::}m \texttt{ |}$$
$$p\texttt{[following-sibling::}n\texttt{]/descendant-or-self::}m]\!]x\}$$

Equality $*$ holds due to

$$(x_2 \notin \texttt{child}^*(x_3) \wedge x_3 \ll x_2 \wedge x_1 \in \mathit{sibling}(x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1))$$
$$\leftrightarrow \exists x_4(x_2 \in \texttt{child}^*(x_4) \wedge x_3 \ll x_4 \wedge x_4 \in \mathit{sibling}(x_3) \wedge x_4 \ll x_1 \wedge x_1 \in \mathit{sibling}(x_4))$$

□

Equivalence 37: Let x be the context node.

$\mathcal{S}[\![p\texttt{/following::}n\texttt{/preceding::}m]\!]x$
$$= \{x_2 | x_3 \in \mathcal{S}[\![p]\!]x \wedge x_1 \in \mathcal{S}[\![\texttt{following::}n]\!]x_3 \wedge x_2 \in \mathcal{S}[\![\texttt{preceding::}m]\!]x_1\}$$
$$= \{x_2 | x_3 \in \mathcal{S}[\![p]\!]x \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge \mathit{nodetest}(x_1, n) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge$$
$$x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge \mathit{nodetest}(x_2, m)\}$$
$$= \{x_2 | x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge$$
$$x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m) \wedge (x_2 \ll x_3 \vee x_3 \ll x_2 \vee x_2 = x_3)\}$$
$$= \{x_2 | ((x_2 \ll x_3 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_1)) \vee$$
$$((x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge$$
$$x_2 \in \mathit{branch\text{-}sibling}(x_1))) \wedge$$
$$(x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$$
$$= \{x_2 | ((x_2 \ll x_3 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \in \mathit{branch\text{-}sibling}(x_3)) \vee$$
$$((x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge$$
$$x_2 \in \mathit{branch\text{-}sibling}(x_1))) \wedge$$
$$(x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$$
$$= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following::}n\texttt{]/preceding::}m]\!]x) \vee$$
$$((x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge$$
$$x_2 \in \mathit{branch\text{-}sibling}(x_1) \wedge$$
$$x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$$
$$= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following::}n\texttt{]/preceding::}m]\!]x) \vee$$
$$(((x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge$$
$$x_2 \in \mathit{branch\text{-}sibling}(x_1)) \wedge$$
$$(x_2 \notin \texttt{child}^*(x_3) \vee x_2 \in \texttt{child}^*(x_3)) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$$
$$= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following::}n\texttt{]/preceding::}m]\!]x) \vee$$
$$((x_2 \notin \texttt{child}^*(x_3) \wedge (x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge$$
$$x_2 \in \mathit{branch\text{-}sibling}(x_1)) \vee$$
$$(x_2 \in \texttt{child}^*(x_3) \wedge (x_3 \ll x_2 \vee x_2 = x_3) \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge$$
$$x_2 \in \mathit{branch\text{-}sibling}(x_1)) \wedge$$
$$\wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \mathit{nodetest}(x_1, n) \wedge \mathit{nodetest}(x_2, m))\}$$
$$= \{x_2 | (x_2 \in \mathcal{S}[\![p\texttt{[following::}n\texttt{]/preceding::}m]\!]x) \vee$$
$$((x_2 \notin \texttt{child}^*(x_3) \wedge x_3 \ll x_2 \wedge x_1 \in \mathit{branch\text{-}sibling}(x_3) \wedge x_2 \ll x_1 \wedge$$

$$
\begin{aligned}
&\qquad\qquad\qquad\qquad\qquad\qquad x_2 \in \textit{branch-sibling}(x_1))\vee\\
&\qquad (x_2 \in \texttt{child}^*(x_3) \wedge x_1 \in \textit{branch-sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \textit{branch-sibling}(x_1))\wedge\\
&\qquad \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \textit{nodetest}(x_1, n) \wedge \textit{nodetest}(x_2, m))\}\\
&= \{x_2|(x_2 \in \mathcal{S}[\![p\texttt{[following::}n\texttt{]/preceding::}m]\!]x)\vee\\
&\qquad ((x_3 \ll x_2 \wedge x_2 \in \textit{branch-sibling}(x_3) \wedge x_2 \ll x_1 \wedge x_2 \in \textit{branch-sibling}(x_1))\vee\\
&\qquad (x_2 \in \texttt{child}^*(x_3) \wedge x_1 \in \textit{branch-sibling}(x_3) \wedge x_3 \ll x_1 \wedge x_2 \ll x_1 \wedge x_2 \in \textit{branch-sibling}(x_1))\wedge\\
&\qquad \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \textit{nodetest}(x_1, n) \wedge \textit{nodetest}(x_2, m))\}\\
&= \{x_2|(x_2 \in \mathcal{S}[\![p\texttt{[following::}n\texttt{]/preceding::}m]\!]x\vee\\
&\qquad x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[following::}n\texttt{]}]\!]x) \vee ((x_2 \in \texttt{child}^*(x_3) \wedge\\
&\qquad x_1 \in \textit{branch-sibling}(x_3) \wedge x_3 \ll x_1) \wedge x_3 \in \mathcal{S}[\![p]\!]x \wedge \textit{nodetest}(x_1, n) \wedge \textit{nodetest}(x_2, m))\}\\
&= \{x_2|x_2 \in \mathcal{S}[\![p\texttt{[following::}n\texttt{]/preceding::}m]\!]x\vee\\
&\qquad x_2 \in \mathcal{S}[\![p\texttt{/following::}m\texttt{[following::}n\texttt{]}]\!]x \vee\\
&\qquad x_2 \in \mathcal{S}[\![p\texttt{[following::}n\texttt{]/descendant-or-self::}m]\!]x\}\\
&= \mathcal{S}[\![p\texttt{[following::}n\texttt{]/preceding::}m \texttt{ | } p\texttt{/following::}m\texttt{[following::}n\texttt{]} \texttt{ |}\\
&\qquad p\texttt{[following::}n\texttt{]/descendant-or-self::}m]\!]x\}
\end{aligned}
$$

□

Equivalences 38, 38a and39: The proof is analogous to the proofs for Equivalences 33, 33a and 34. □

Equivalence 40 holds with Lemmas A.6 and A.7. □

Equivalences 41 and 42: The proof is analogous to the proofs for Equivalences 36 and 37. □