



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN



Nathan P. Gibson • University of Munich (LMU) • [usaybia.net](http://usaybia.net)

# Thinking in **RTL**



Reorienting the Directional Assumptions of  
Global Digital Scholarship



<https://tinyurl.com/gibson190608>

Right2Left Workshop  
#DHSI19RTL  
Victoria, 8 June 2019



# Outline



**LMU**

LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

# A Dizzying Tour of Directionality (The Past)

**LMU**

LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

# #sinistrodextrification

## Modern Writing Systems



Image: Adapted from  
<https://commons.wikimedia.org/wiki/File:WritingSystemsOfTheWorld.svg>. Creator of the vector version:  
Pmx. Original work: Maximilian  
Dörrbecker, [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/).

# Selected Languages\* with Primarily Right-to-Left Writing Systems by Number of First-Language Speakers

|     |                                |                |
|-----|--------------------------------|----------------|
| #5  | Arabic                         | 319 (millions) |
| #10 | Lahnda (Western Punjabi, etc.) | 119            |
| #20 | Urdu                           | 69             |
| #23 | Persian                        | 62             |
|     | Pushto                         | 38**           |
|     | Sindhi                         | 25             |
|     | Kurdish                        | 22**           |
|     | Uyghur                         | 10**           |
|     | Hebrew                         | 5              |
|     | Rohingya                       | 3**            |

\* Or macrolanguages.

\*\* May include non-primary speakers

Source: Eberhard, David M., Gary F. Simons, and Charles D. Fennig, eds. 2019. "Summary by Language Size" [and individual language profile pages]. In Ethnologue: Languages of the World, 22nd ed. Dallas: SIL International.  
<https://www.ethnologue.com/statistics/size>.

> **half a billion people**



almost 10% of the world population

**LMU**

LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

# The State of Standards

(Past Meets Present)



# Plain Text

## Logical vs. Visual Ordering



Image: Richard Ishida, "Visual vs. logical ordering of text," W3C, 10 June 2016.

<https://www.w3.org/International/questions/qa-visual-vs-logical>.

Copyright © 2011-2016 W3C® (MIT, ERCIM, Keio, Beihang)

<http://www.w3.org/Consortium/Legal/2015/doc-license>.



# Plain Text

Unicode bidi controls

1991 1.0

2013 6.3.0 additional controls

# Plain Text: Unicode Bidi Algorithm

- Base direction determines order of directional runs
- Base direction default (HTML) = LTR
- Neutral chars (e.g., spaces and punctuation) between same strongly typed inherit direction.
- Numbers are weakly typed--LTR but don't break directional runs.
- "Mirror characters" (parentheses, etc.) reorient automatically within directional runs.
- Markup needed for:
  - Punctuation at bidi boundary that belongs to opposite-direction text.
  - Bidi text inside bidi.
  - Isolating adjacent ranges

See Richard Ishida, "Unicode Bidirectional Algorithm Basics," W3C, August 9, 2016, <https://www.w3.org/International/articles/inline-bidi-markup/uba-basics>,

Nathan Gibson, "Thinking in JTЯ," Right2Left DHSI, 8 June 2019, <https://tinyurl.com/gibson190608>

## HTML & CSS

- When you know (or can find out) the base direction, set it explicitly. When you don't know, use `<bdi>` or `@dir="auto"` to isolate text portions that could go in opposite directions.
- Whenever possible use markup (e.g., HTML) rather than Unicode characters to specify the base direction of text.
- Use HTML markup rather than CSS to control direction. (But you might need to use CSS for XML.)
- Don't combine Unicode and CSS formatting directions!
- CSS "direction" usually does nothing unless you combine it with a "unicode-bidi" instruction.
- Use Unicode bidi formatting for plain text, HTML for HTML, and CSS (if needed) for XML.

# HTML & CSS

Test out these examples! <https://jsfiddle.net/x4fy71r6/1/>



| Unicode                  | HTML                                    | CSS   | Incorrect  | Correct  |
|--------------------------|---|---|--|--|
| RLI ... PDI              | <code>@dir="rtl"</code>                 | <code>direction:rtl; unicode-bidi:isolate</code>          | This text is LTR with some عربية, فارسی, and اردو in the middle. | This text is LTR with some عربية, فارسی, and اردو in the middle. |
| LRI ... PDI              | <code>@dir = "ltr"</code>               | <code>direction:ltr; unicode-bidi:isolate</code>          | English. إن في هذا الخط  | إن في هذا الخط English.  |
| FSI ... PDI              | <code>&lt;bdi&gt; or @dir="auto"</code> | <code>unicode-bidi:plaintext</code>                       | Handle 4 كلمة automatically.                                     | Handle 4 كلمة automatically.                                     |
| FSI LRO . .<br>. PDF PDI | <code>&lt;bdo dir="ltr"&gt;</code>      | <code>direction:ltr; unicode-bidi:isolate-override</code> | Force this Hebrew text left-to-right: אבג.                       | Force this Hebrew text left-to-right: גבא.                       |
| FSI RLO . .<br>. PDF PDI | <code>&lt;bdo dir="rtl"&gt;</code>      | <code>direction:rtl; unicode-bidi:isolate-override</code> | Force this Latin text right-to-left: abc.                        | Force this Latin text right-to-left: cba.                        |

Adapted from "Unicode Standard Annex #9, [https://www.unicode.org/reports/tr9/#Markup\\_And\\_Formatting](https://www.unicode.org/reports/tr9/#Markup_And_Formatting).

# TEI-XML

- XML supports the Unicode Bidi Algorithm.
- Wrap different languages/scripts in elements with BCP-47 language and script codes. Make sure your stylesheets look for script, not just language. (E.g., “ar” = Arabic in default script “Arab”, but “ar-Latn” would be Romanized Arabic, in the opposite direction!)
- When you need to mark direction explicitly beyond this, use CSS instructions for direction, unicode-bidi, writing mode, and text-orientation with @tei:style on individual elements or using the <tei:rendition> mechanism.
- You can do crazy stuff (like boustrophedon) with CSS “transform”! Try it out.

TEI P5 Guidelines 5.6 “Writing Modes”

**LMU**

LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

# **Forward from Failures** (Present Meets Future)

# #redextrosinistrify

= #bringbackrtl, #Right2Left,  
#DHSI19RTL



**Image of Holocaust  
remembrance sign at  
Königsplatz, Munich with  
erroneously rendered Hebrew  
and Arabic text removed for  
copyright reasons.**

Image: (c) Andreas Kaplony, 11 November 2018, Königsplatz, Munich.

Nathan Gibson, "Thinking in JTR," Right2Left DHSI, 8 June 2019, <https://tinyurl.com/gibson190608>

# Example: Syriaca.org



## The Syriac Biographical Dictionary

---

Gregorius bar Hebraeus - ܩܪܝܘܪܝܘܫ ܒܪ ܗܒܪܝܐ (1225/6 - 1286)

---

URI <http://syriaca.org/person/239>

---

**Identity**  
 "Maph. (since 1264) and polymath." <sup>1</sup>

**Names:**

ܩܪܝܘܪܝܘܫ ܒܪ ܗܒܪܝܐ    Gregorius bar Hebraeus    ܩܪܝܘܪܝܘܫ ܒܪ ܗܒܪܝܐ <sup>6</sup>

ܩܪܝܘܪܝܘܫ ܒܪ ܗܒܪܝܐ ܗܘܐ ܩܪܝܘܪܝܘܫ ܒܪ ܗܒܪܝܐ <sup>4</sup>    ܩܪܝܘܪܝܘܫ ܒܪ ܗܒܪܝܐ <sup>4</sup>

ܩܪܝܘܪܝܘܫ ܒܪ ܗܒܪܝܐ ܗܘܐ ܩܪܝܘܪܝܘܫ ܒܪ ܗܒܪܝܐ <sup>3</sup>    Abū al-Faraj <sup>1</sup>    Barhebraeus <sup>1</sup>    Barhebraeus <sup>1</sup>



LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

## Example: **Syriaca.org**

Huge shoutout to the work of

[@wsalesky](#)

on

[srophe.app](#)

# Implementation: TEI

```

- <persName xml:id="name239-2" corresp="#name239-3 #name239-4 #name239-5" xml:lang="en" source="#bib239-2">
  <forename sort="2">Gregorius</forename>
  <addName type="family" sort="1">Bar Hebraeus</addName>
</persName>
- <persName xml:id="name239-3" corresp="#name239-2 #name239-4 #name239-5" xml:lang="ar" source="#bib239-3">
  <addName type="untagged-title" sort="2">عمار</addName>
  <forename sort="2">غريغور يوس</forename>
  <addName type="untagged-title" sort="2">ابو الفرج</addName>
  <addName type="untagged-title" sort="2">الملطي</addName>
  <addName type="untagged-title" sort="2">مفران</addName>
  <addName type="untagged-title" sort="2">المشهور بابن العبري</addName>
</persName>
- <persName xml:id="name239-4" corresp="#name239-2 #name239-3 #name239-5" xml:lang="syr" source="#bib239-4">
  <addName type="untagged-title" sort="2">ܡܚܕܡ</addName>
  <forename sort="2">ܠܕܘܢܗܘܘܣ</forename>
  <addName type="untagged-title" sort="2">ܡܚܦܘܢ</addName>
  <addName type="family" sort="1">ܡܚܕܘܡ</addName>
</persName>
- <persName xml:id="name239-5" corresp="#name239-2 #name239-3 #name239-4" xml:lang="syr-Syrj" source="#bib239-4">
  <addName type="untagged-title" sort="2">ܡܚܕܡ</addName>
  <forename sort="2">ܠܕܘܢܗܘܘܣ</forename>
  <addName type="untagged-title" sort="2">ܡܚܦܘܢ</addName>
  <addName type="family" sort="1">ܡܚܕܘܡ</addName>
</persName>

```

# Implementation: HTML & CSS

The screenshot displays a web browser's developer tools interface. The top section shows a list of names in Arabic script, with one name highlighted. The bottom section shows the corresponding HTML code and CSS styles.

**Names:**

- Gregorius bar Hebraeus
- Abū al-Faraj<sup>1</sup>
- Barhebraeus<sup>1</sup>
- Barhebraeus<sup>1</sup>

**HTML Code:**

```
<span id="name239-5" class="tei-persName" xml:lang="syr-Syrj" dir="rtl" lang="syr-Syrj">
  كَذِبٌ كَذُّهُنَّهٖ مَعْفُؤُنَا كَذُّ حُنَّارِ
</span>
```

**CSS Styles:**

```
:lang(syr-Syrj), :lang(syc-Syrj) {
  font-family: "SertoBatnan", "Serto
  Verdana, Arial, Helvetica, sans-
  font-size: 1.25em;
  line-height: 1em;
}
```

## Example: Code Editing in OxygenXML (LTR base direction)

```
<p>This text is LTR with some عربية , فارسی , and اردو in the middle.</p>  
<p dir="ltr">This text is LTR with some <span dir="rtl">فارسی</span>, <span dir="rtl">  
>عربية</span>, and <span dir="rtl">اردو</span> in the middle.</p>  
<p>English. إن في هذا الخط</p>  
<p dir="rtl">إن في هذا الخط <span dir="ltr">English</span>.</p>  
<p>Handle 4 كلمة automatically.</p>  
<p dir="auto">Handle <bdi>كلمة</bdi> 4 automatically.</p>  
<p>Force this Hebrew text left-to-right: אבג.</p>  
<p>Force this Hebrew text left-to-right: <bdo dir="ltr">אבג</bdo>.</p>  
<p>Force this Latin text right-to-left: abc.</p>  
<p>Force this Latin text right-to-left: <bdo dir="rtl">abc</bdo>.</p>
```

See Richard Ishida, "Authoring HTML: Handling Right-to-left Scripts," W3C, <https://www.w3.org/TR/i18n-html-tech-bidi>.

## Example: Code Editing in OxygenXML (RTL base direction)

```

<in the middle.</p> اردو and , فارسی, عربية , <p>This text is LTR with some
"span>, <span dir="rtl"/>فارسی"<p dir="ltr">This text is LTR with some <span dir="rtl"
<span> in the middle.</p/> اردو"&span>, and <span dir="rtl"/>عربية<
<English.</p> إن في هذا الخط
<span dir="ltr">English</span>.</p> إن في هذا الخط
<automatically.</p> 4 كلمة <p>Handle
<bdi> 4 automatically.</p/> كلمة<p dir="auto">Handle <bdi>
<p/>. אבג :<p>Force this Hebrew text left-to-right
<bdo>.</p/>אבג"<p>Force this Hebrew text left-to-right: <bdo dir="ltr"
<p>Force this Latin text right-to-left: abc.</p>
<p>Force this Latin text right-to-left: <bdo dir="rtl">abc</bdo>.</p>

```

## Example: Syriac Connecting Letters in OxygenXML

(Syriac script unsupported in Java 8 but supported in OpenJDK)

```
<persName xml:id="name239-4" corresp="#name239-2 #name239-3 :  
  <addName type="untagged-title" sort="2">ܩܕܝܫܐܢܐ</addName>  
  <forename sort="2">ܫܘܒܗܐܢܐܩܕܝܫܐܢܐ</forename>  
  <addName type="untagged-title" sort="2">ܩܕܝܫܐܢܐ</addName>  
  <addName type="family" sort="1">ܩܕܝܫܐܢܐ</addName>  
</persName>
```



# Humanist Principles for Text Input

- **Use arrow keys or buttons for direction (left, right, up, down) rather than sequence (forward, back, next, previous).** When used for sequence, arrows should be clearly labeled and, when possible, match the reading direction of the text.
- **Allow the user to set the base direction.**
- **Visually indicate the direction of text entry in bidi contexts.**
- **Provide mechanisms for users to override default writing modes (character/word/line order) and character orientation (rotated or flipped).**

# Humanist Principles for Text Processing

- **Explicitly set the base direction of the text whenever known.** When unknown, at least isolate (wrap) text that could potentially go in the opposite direction.
- **Take into account the additional features of right-to-left writing systems, such as connecting and combining letters.**

# Humanist Principles for Text Display

- **Pay attention to script codes.** Script or writing system indicates text direction more reliably than the language does. (E.g., Turkish may be written in Latin or Arabic script, Hebrew may be Romanized, etc.) Additionally, character-level defaults are not always enough to reconstruct the text's direction (e.g., boustrophedon).

# Humanist Principles for Code Authoring

- **Work toward wide-scale standards and support for bidi code .**
- **Consider implementing convenience features with bidi/RTL code in mind.**  
E.g., direction-aware auto-indentation and line breaks, code autocompletion, and intelligent text selection.

# What do we need to advance humanist principles?

- **A list of principles, with**
  - Links to tech specs
  - Example implementations
- Sandbox area for fiddling with RTL?
- Community conversations?
- Compatibility reviews?
- An RTL badge?

## Humanist Principles

Because almost 1 in 10 people in the world primarily use a language that primarily uses a RTL writing system.

Because an enormous part of the world's cultural heritage is written in RTL.

# #redextrosinistrify