

langsci-avm

Felix Kopecky*

Version 0.3.0-rc3 – 12 October 2020

1 Introduction

`langsci-avm` is a \LaTeX 3 package aimed at typesetting beautiful feature structures, also known as *attribute-value matrices*, for use in linguistics. The package provides a minimal and easy to read syntax. It depends only on the `array` package and can be placed almost everywhere, in particular in footnotes or graphs and tree structures. The package is meant as an update to, and serves the same purpose as, Christopher Manning’s `avm` package, but shares no code base with that package. When you come from `avm`, please see Section 4.6 for a quick conversion guide.

To start using `langsci-avm`, place `\usepackage{langsci-avm}` in your preamble.

This document is structured as follows: Section 2 describes the input syntax for AVMs and their parts. Ways to customise your AVM’s layout follow in Section 3, and selected usage cases are presented in Section 4. There’s also an administrative and \TeX nicol appendix at the end of this document, in case you are interested.

1.1 Example

```
\avm{
  [ ctxt & [ max-qud \[
    sal-utt & \{ [ cat \[
      cont <ind & i> ] \}
    ]
  ]
}
```

$$\left[\begin{array}{c} \text{CTXT} \\ \left[\begin{array}{c} \text{MAX-QUD} \\ \text{SAL-UTT} \end{array} \left\{ \left[\begin{array}{c} \text{CAT} \\ \text{CONT} \end{array} \langle \text{IND } i \rangle \right] \right\} \right] \end{array} \right]$$

1.2 Acknowledgements

Thanks to Phelype Oleinik for help on recursion and expansion with \LaTeX 3. Thanks to Ahmet Bilal Özdemir and Stefan Müller for their contributions in planning and testing this package.

*<mailto:felix.kopecky@langsci-press.org>. Please submit bug reports and feature requests to <https://github.com/langsci/langsci-avm/issues>.

2 Structuring AVMs

`\avm` `\avm [options] {structure}`

The heart of this package and its root document comand is `\avm`. In the scope of the command, delimiter characters are processed to open and close (sub-)structures, as described in Section 2.1. Special elements are described in Section 2.2. For a description of the layout *options*, see Section 3.

A *structure* is basically the content of a stylised `tabular`: The columns are separated by `&` and a new line is entered with `\\`.

2.1 Entering (sub-)structures within `\avm`

<code>[...]</code>	<code>[<i>structure</i>]</code>
<code><...></code>	<code>< <i>structure</i> ></code>
<code>(...)</code>	<code>(<i>structure</i>)</code>
<code>\{...\}</code>	<code>\{ <i>structure</i> \}</code>
<code>\[...\]</code>	<code>\[<i>structure</i> \]</code>

Updated: 2020-10-02

Within the scope of `\avm`, these delimiters create (sub-)structures that are enclosed by the respective delimiter. Due to the special meaning that curly braces have in \LaTeX , these are the only ones that need to be run with an escape token (`\`). It is currently possible to mix delimiters, e.g. with `<(structure)>`, but this may change in future versions.

`langsci-avm` expects your (sub-)structures to have *at most two columns*, so that for every line in each (sub-)structure, there should be no more than one `&`. It is recommended to have at least some lines with a `&` in your *structure*. Currently, display issues may appear in some structures if none are given.

<code>\avm{</code> <code>\[[< (\{ ... \}) >] \]</code> <code>}</code>	$\llbracket \llbracket \langle \{ \dots \} \rangle \rrbracket \rrbracket$
<code>\avm{</code> <code>\[[\{ ... \} \\</code> <code>< (...), (...) >] \]</code> <code>}</code>	$\llbracket \left[\left\{ \dots \right\} \right. \\ \left. \left\langle (\dots), (\dots) \right\rangle \right] \rrbracket$

Warning: The semantic bracket $\llbracket \dots \rrbracket$ is only available when the package option `[lfg]` is loaded, i.e. `\usepackage[lfg]{langsci-avm}`. This is because the semantic delimiters are not available in every font, and are currently not provided in standard \LaTeX documents. If you load the `[lfg]` option but do not provide the symbol (e.g. by using a font such as `libertinus`), the package `unicode-math` will automatically be loaded to provide the symbol. If the `[lfg]` option is not present, `\[{structure} \]` will result in none delimiter output, but the `{structure}` will be printed nonetheless.

`!...!` `! text !`

Escapes the `avm` mode so that all delimiters can be used as usual characters. If you need `!` as a regular character, see Section 3 for how to change the `switch`.

2.2 Commands for tags, types, unusual lines, and relations

<hr/> <code>\tag</code> <hr/>	<code>\tag {⟨<i>identifier</i>⟩}</code>	
<code>\0</code>	<code>\0, \1, \2, \3, \4, \5, \6, \7, \8, \9</code>	
<code>\1</code>	<code>\tag</code> puts its <code>{⟨<i>identifier</i>⟩}</code> in a box, more precisely an <code>\fbox</code> . Within the box, the <code>tags</code> font is applied. <code>\0, \1, ..., \9</code> are shortcuts to <code>\tag</code> and place the respective number in the box. For example, <code>\4</code> is equivalent to <code>\tag{4}</code> . The shortcuts do not take any arguments.	
<code>...</code>		
<code>\9</code>		
<hr/> Updated: 2020-04-29 <hr/>	<p>If you want to use this command outside an AVM, you can obtain, for example, <code>\4</code>, by using <code>\avm{\4}</code>, or the equivalent <code>{\fboxsep.25ex\fbox{\footnotesize 4}}</code>.</p>	
	$\begin{array}{l} \backslash\text{avm}\{[\text{ attr1 \& \4} \backslash \\ \text{ attr2 \& \4} [\text{attr3 \& val3} \backslash \\ \text{ attr4 \& val4}] \text{]}\} \end{array}$	$\left[\begin{array}{l} \text{ATTR1 } \boxed{4} \\ \text{ATTR2 } \boxed{4} \left[\begin{array}{l} \text{ATTR3 } val3 \\ \text{ATTR4 } val4 \end{array} \right] \end{array} \right]$
<hr/> <code>\type</code> <hr/>	<code>\type⟨*⟩ {⟨<i>type</i>⟩}</code>	
<code>\type*</code>	Will output the <code>⟨<i>type</i>⟩</code> in the <code>types</code> font (roman italics by default). The starred variant <code>\type*</code> will span the complete (sub-)structure and <i>can only be placed in the first column</i> of this structure. After the starred <code>\type*</code> , a <code>\</code> is recommended, but can usually be omitted.	
<hr/> Updated: 2020-03-30 <hr/>		
	$\begin{array}{l} \backslash\text{avm}\{[\text{ \type*}\{A \text{ type spanning a line}\} \\ \text{ attr \& } [\text{ \type}\{type\}] \text{]}\} \end{array}$	$\left[\begin{array}{l} A \text{ type spanning a line} \\ \text{ATTR } [type] \end{array} \right]$
<hr/> <code>\id</code> <hr/>	<code>\id {⟨<i>id</i>⟩} {⟨<i>structure</i>⟩}</code>	
New: 2020-10-02	A variant of <code>\substack</code> from <code>amsmath</code> , this command adds an identifier to the <code>{⟨<i>structure</i>⟩}</code> . The contents of <code>{⟨<i>id</i>⟩}</code> will be set in math mode by default, which is convenient given that they often contain variables with subscript indices. Multiple IDs should be separated by a new line, <code>\</code> . The alignment of the id column can be changed with the key <code>id align</code> , see Section 3.	
Updated: 2020-12-11 <hr/>		
	$\backslash\text{avm}\{ \text{ \id}\{n_1 \backslash n_2\} \{ [\text{subj} \backslash \text{pred} \& \text{swim}] \} \}$	$\begin{array}{c} \left[\begin{array}{l} \text{SUBJ} \\ \text{PRED } swim \end{array} \right] \\ \begin{array}{l} n_1 \\ n_2 \end{array} \end{array}$
<hr/> <code>\punk</code> <hr/>	<code>\punk {⟨<i>attribute</i>⟩} {⟨<i>type</i>⟩}</code>	
Some <code>⟨<i>attributes</i>⟩</code> think that the layout of the other attributes in their community leaves no space for them to express their individuality. They desire a life outside the confines of the alignment defined by the others, while still remaining a member of the matrix.		
Technically, this is a line with no snapping to the column layout, but with spacing between the <code>⟨<i>attribute</i>⟩</code> and <code>⟨<i>type</i>⟩</code> . After <code>\punk</code> , a <code>\</code> is recommended, but can be omitted in “normal” cases.		
	$\begin{array}{l} \backslash\text{avm}\{[\text{ attr1 \& val1} \backslash \\ \text{ \punk}\{a \text{ quite long attr2}\} \{val2\} \text{]}\} \\ \text{ attr3 \& val3} \backslash \\ \text{ attr4 \& val4} \\ \text{]}\} \end{array}$	$\left[\begin{array}{l} \text{ATTR1 } val1 \\ A \text{ QUITE LONG ATTR2 } val2 \\ \text{ATTR3 } val3 \\ \text{ATTR4 } val4 \end{array} \right]$

<code>\+</code>	In the scope of <code>\avm</code> , <code>\+</code> comes out as “ \oplus ”. “+” can be obtained normally. <i>In the earlier Version 0.1.0-beta, + produced “\oplus”.</i>
Updated: 2020-03-16	

<code>\-</code>	In the scope of <code>\avm</code> , <code>\-</code> comes out as “ \ominus ”. To use the “optional hyphenation” meaning of <code>\-</code> , please write <code>!\-</code> !
New: 2020-03-17	

<code>\shuffle</code>	In the scope of <code>\avm</code> , <code>\shuffle</code> is a shortcut for “ \circ ” to mark the shuffle relation.
New: 2020-03-17	

3 AVM layout

3.1 Defining styles

You can customise many aspects of how an AVM is printed, including the fonts or spacing between delimiters and content. You can apply them locally via the `[<options>]` of `\avm` or by using `\avmsetup`. And you can also define your own styles and use them via the `[<style = >]` option in `\avm`.

<code>\avmsetup</code>	<code>\avmsetup {<options>}</code> <code>{<options>}</code> is a comma-separated list of key = value settings. See the list below for all user-configurable options. The <code>{<options>}</code> are the same as in <code>\avm[<options>]</code> . When inserted in <code>\avm[<options>]</code> , they apply locally, and globally if given to <code>\avmsetup</code> . Local settings always override global ones, and you can have any feasible number of <code>\avmsetups</code> in your document.
------------------------	---

<code>\avmdefinestyle</code>	<code>\avmdefinestyle {<name>} {<settings>}</code> Instead of applying settings globally or per AVM, you can also define styles and assign them to AVMs, as in <code>\avm[style=<name>]{...}</code> . The <code><settings></code> are a comma-separated list of key = value settings, and should be a subset of the settings from <code>\avmsetup</code> . For example, the following <code>plain</code> style highlights neither attributes, values, nor types:
New: 2020-05-11	

```
\avmdefinestyle{plain}{attributes=\normalfont,
                        values=\normalfont,
                        types=\normalfont}
```

The style is applied with `\avm[style=plain]{...}`.

Now to the list of settings you can actually apply:

style = `<name>` (initially empty)
 In addition to any style that you possibly define yourself, a style **narrow** is pre-defined in the package (see Section 4.1).

stretch = `<factor>` (initially 0.9)
 Define `\arraystretch`, i.e. a factor in the determination of line height.

`columnsep` = $\langle length \rangle$ (initially `0.5ex`)
 Define the `\tabcolsep`, i.e. horizontal space between columns. The first and second column will have `0\columnsep` to the left and right, respectively. Between the two the distance is `2\columnsep`. Using relative units (like `ex` or `em`) may be a good idea so that `columnsep` scales well with changes in font size.

`delimfactor` = $\langle factor \rangle$ (initially `1000`)
 Sets `\delimeterfactor`. The calculation for the minimum height of a delimiter is $y \cdot f / 1000$, where y is the height of the content and f the value of `delimfactor`. The default `1000` ensure that the delimiters' height is at least that of the structure.

`delimfall` = $\langle length \rangle$ (initially `0pt`)
 Controls `\delimitershortfall`, i.e. the maximum height that the delimiters can be shorter than the enclosed structure. The default `0pt` ensure that the delimiters are not shorter than the contents.

`extraskip` = $\langle length \rangle$ (initially `\smallskipamount`)
 If a substructure is immediately followed by a `\\`, an extra amount of vertical skip is added so that the content of the next line, possibly another delimiter, does not clash with the delimiter in that line. This automatic skip insertion can be circumvented with placing a `\relax` before the linebreak, i.e. `\relax\\`.

`attributes` = $\langle font settings \rangle$ (initially `\scshape`)
 The font for attributes, i.e. the first column of each structure.

`values` = $\langle font settings \rangle$ (initially `\itshape`)
 The font for values, i.e. the second column of each structure.

`types` = $\langle font settings \rangle$ (initially `\itshape`)
 The font used in `\type` and `\type*`.

`tags` = $\langle format settings \rangle$ (initially `\footnotesize`)
 The font (size) used in `\tag` and the shortcuts `\1...\9`.

`switch` = $\langle token \rangle$ (initially `!`)
 Define the escape token. Change this if you need to use “!” as a text glyph.

`id align` = $\langle token \rangle$ (initially `l`)
 Change the alignment of the column inserted by `\id`. Has to be a column specification. The most probable choices are `l` and `r`.

`customise` = $\langle settings \rangle$ (initially empty)
 An interface to input custom commands to be run at the beginning of every `\avm`.

3.2 Drawing edges between AVM contents

It is possible to make AVM contents available to `tikz`, so that they can be referenced in a `tikzpicture`. To enable this feature, `langsci-avm` has to be loaded with the option `[tikz]`:

```
\usepackage[tikz]{langsci-avm}
```

Additionally, `avm` environments on which `tikz` is to be used need to have the `[pic]` option present:

`\avm[pic] {...}`

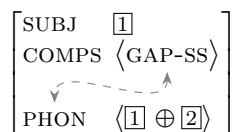
Only the parts of an AVM that are specifically marked will be known to `tikz`. To mark a part of an AVM to be used by `TikZ`, use `\node`:

<code>\node</code>	<code>\node {<id>} {<contents>}</code>
New: 2020-09-23	<p><code>{<id>}</code> serves as part of the node’s identifier in a <code>tikzpicture</code>. It will be prefixed, and it’s complete name will be <code>avm-n-\marg{id}</code>, where <code>n</code> is the counter of <code>\avm</code> in your document that have the <code>[pic]</code> option enabled and that don’t have a <code>picname</code> (see below). <code>n</code> starts at 1. For example, a <code>\node</code> named “pretty-node” in the fourth <code>[pic]</code>-enabled <code>avm</code> in your document will be <code>avm-4-pretty-node</code>. Note that <code>\node</code> will register the complete name globally in your document, and so can’t be declared by other <code>tikz</code> nodes.</p>

This behaviour can be adjusted by passing a `[picname = <avm’s name>]` to `\avm`. E.g., `\nodes` within `\avm[pic, picname = example1]` will have a full name pattern of `example1-<id>`. Named `\avms` do *not* raise the `n` mentioned in the last paragraph.

Any proper part of an AVM can be referenced in `{<contents>}`. It could be just a value, an attribute’s name, or part of either, but whole (sub-)structures can be part of `{<contents>}` as well.

A `tikzpicture` with options `[remember picture, overlay]` enabled can reference `langsci-avm`’s `\nodes`. This way, `TikZ`’ extensive drawing abilities are available for the decoration of AVMs. Here’s a very simple example document:



```

\documentclass{article}
\usepackage{tikz}{langsci-avm}
\usepackage{tikz} % optional, since langsci-avm will load tikz if option
                  % tikz is present
\usetikzlibrary{arrows,arrows.meta}

\avm[pic]{[ subj & \1\
          comps & <\node{gap}{gap-ss}> \bigskip\
          \node{phon}{phon} & <\1 \+ \2>
        ]}

\begin{tikzpicture}[remember picture,overlay]
  \path[{Stealth[]}--{Stealth[]},gray,dashed,in=90,out=270]
    (avm-1-gap.south) edge (avm-1-phon.north);
\end{tikzpicture}

```

3.3 Defining input patterns

`\avmdefinecommand`

New: 2020-06-29

`\avmdefinecommand {⟨name⟩} [⟨label⟩] {⟨settings⟩}`

Sub-structures often come in patterns. For example, AVMs often have a PHON attribute, which is mapped to a list, the entries of which are in italics. `\avmdefinecommand` can account for this and other input patterns. For example,

`\avmdefinecommand{custom}{...}`

will create a command `\custom` available only in the scope of `\avm` (this means that you can have a different meaning in the rest of your document). The `⟨settings⟩` will then be applied to the scope in which `\custom` is called. If an optional `⟨label⟩` is given, the label will be printed, in the current font, before the `⟨settings⟩` are applied.

`\custom` generated in this way automatically advances to the value column after the `⟨label⟩` is printed. This means that commands generated with `\avmdefinecommand` should be called in the attribute column of an existing structure. This behaviour can be circumvented with the starred variant `\name*`, which is automatically generated by `\avmdefinecommand` as well. However, it seems advisable to use the starred variants sparingly.

Here's an example for the aforementioned phon pattern:

```
\avmdefinecommand{phon}[phon]
{
  attributes = \itshape,
  delimfactor = 900,
  delimfall = 10pt
}
```

This creates a command `\phon` (and the variant `\phon*`) within the scope of any `\avm`. It will print the label `phon` in the current font and then apply three settings locally: italics for the attribute (first) column, and two settings for very narrow delimiter fitting.

This results in: (The font of this documentation has little support for IPA.)

```
\avm{
  [\type*{word}
  \phon <lin'gwistiks>\[
  synsem & [ ... ]
  ]
}
```

$$\left[\begin{array}{ll} \textit{word} & \\ \text{PHON} & \langle \textit{lin'gwistiks} \rangle \\ \text{SYNSEM} & [\dots] \end{array} \right]$$

Note that any other structure type would have worked instead of `⟨⟩`. But `⟨⟩` and any other markers for sub-structures are left unchanged by `\phon` and other custom commands. This is why the *attribute* font is changed by `\phon`, although *lin'gwistiks* is technically a value. Remember that `<` creates a new list sub-substructure, and the first content is printed in its attribute font.

4 Applications

4.1 Spacing and size of delimiters

`langsci-avm` automatically detects if the end of a sub-structure is followed by a line break. This is useful to find cases in which two sub-structures are printed immediately below each other, and to add extra spacing (the `extraskip` from the options). This automatic detection can be suppressed with `\relax`. See below for the effect of that detection:

<pre>\avm{[[attr1 & val1 \\ attr2 & val2] \\ [attr1 & val1 \\ attr2 & val2]]}</pre>	<pre>\avm{[[attr1 & val1 \\ attr2 & val2] \relax\\ [attr1 & val1 \\ attr2 & val2]]}</pre>
$\left[\begin{array}{l} \left[\begin{array}{l} \text{ATTR1 } val1 \\ \text{ATTR2 } val2 \end{array} \right] \\ \left[\begin{array}{l} \text{ATTR1 } val1 \\ \text{ATTR2 } val2 \end{array} \right] \end{array} \right]$	$\left[\begin{array}{l} \left[\begin{array}{l} \text{ATTR1 } val1 \\ \text{ATTR2 } val2 \end{array} \right] \\ \left[\begin{array}{l} \text{ATTR1 } val1 \\ \text{ATTR2 } val2 \end{array} \right] \end{array} \right]$

If many delimiters are nested, this occasionally results in larger delimiter sizes. There is a pre-defined `narrow` style that resets `delimfall` (to 5pt) and `delimfactor` (to 997), which are the values recommended in the *TEXbook*. This results in a more compact appearance:

<pre>\avm{[attr \{<\1>\}]]}</pre>	<pre>\avm[style=narrow]{[attr \{<\1>\}]]}</pre>
$[\text{ATTR} \langle \{\boxed{1}\} \rangle]$	$[\text{ATTR} \langle \{\boxed{1}\} \rangle]$

4.2 Disjunctions and other relations

Sometimes AMVs are placed beside other content to express disjunctions or other relations. In `langsci-avm` this is done naturally:

<pre>\avm{ [attr1 & val1\\ attr2 & val2\\ attr3 & val3] } \$\lor\$ \avm{ [attr1' & val1'\\ attr2' & val2'\\ attr3' & val3'\\] }</pre>	$\left[\begin{array}{l} \text{ATTR1 } val1 \\ \text{ATTR2 } val2 \\ \text{ATTR3 } val3 \end{array} \right] \vee \left[\begin{array}{l} \text{ATTR1'} val1' \\ \text{ATTR2'} val2' \\ \text{ATTR3'} val3' \end{array} \right]$
<pre>\textit{sign} \$\to\$ \avm{ [attribute1 & value1\\ attribute2 & value2\\ attribute3 & value3] }</pre>	$sign \rightarrow \left[\begin{array}{l} \text{ATTRIBUTE1 } value1 \\ \text{ATTRIBUTE2 } value2 \\ \text{ATTRIBUTE3 } value3 \end{array} \right]$

4.3 Use as a vector

It's possible to use `langsci-avm` for feature vectors rather than matrices, as may be useful in generative grammar.

$$\backslash\text{avm}[\text{attributes}=\backslash\text{normalfont}]\{[v1\backslash\backslash v2\backslash\backslash v3]\}\$\varphi$$

$$\begin{bmatrix} v1 \\ v2 \\ v3 \end{bmatrix} \varphi$$

4.4 Combinations with `gb4e`, `expex`, and `linguex`

This package works fine with `gb4e` and its fork `langsci-gb4e`. To align the example number at the top of your structure, please use `\attop` from `gb4e`:

```
\begin{exe}
  \ex\attop{
    \avm{[ attr1 & val1\backslash
          attr2 & val2\backslash
          attr3 & val3]}
  }
\end{exe}
```

(1) $\begin{bmatrix} \text{ATTR1} & \text{val1} \\ \text{ATTR2} & \text{val2} \\ \text{ATTR3} & \text{val3} \end{bmatrix}$

The same can be achieved with `expex` using `\envup` from `lingmacros` (see below) or using this *experimental* syntax:

```
\ex \vtop{\strut\vskip-\baselineskip{
  \avm{[ attr1 & val1\backslash
        attr2 & val2\backslash
        attr3 & val3]}
}}
\xe
```

Examples typed with `linguex` can be combined with `\envup` from `lingmacros` to align AVMs (many thanks to Jamie Findlay for pointing this out):

```
\ex. \envup{\avm{[ attr1 & val1\backslash
  attr2 & val2\backslash
  attr3 & val3]}
}
```

4.5 Combinations with `forest`

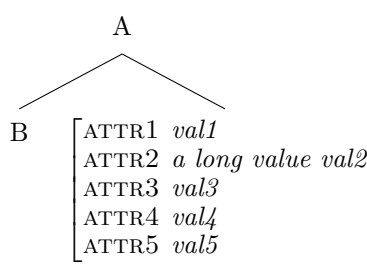
This package also works fine with `forest`. As per the `forest` documentation, it is recommended to protect any `\avm`-statements with `{}` in nodes:

```
\begin{forest}
  [A [B] [{\avm{[attr1 & val1\backslash
    attr2 & val2\backslash
    attr3 & val3]}} ] ]
\end{forest}
```

It may happen that extensive AVMs protrude into the space reserved for other `forest` nodes or edges. In this case, the `forest` setting for `children = {anchor=north}` may be

useful: (If you like, try this tree without that setting.)

```
\begin{forest}
  [A, for children = {anchor=north}
    [B] [{\avm{[attr1 & val1\\
      attr2 & a long value val2\\
      attr3 & val3\\
      attr4 & val4\\
      attr5 & val5]}} ]
  ]
\end{forest}
```



4.6 Switching from Christopher Manning’s `avm` package

Switching from `avm` to `langsci-avm` will require some, though hopefully minimal, changes to the code. In particular, `langsci-avm` doesn’t distinguish between “active” and “passive” modes, there is now a single way of sorting (see `\type`, which replaces `\asort` and `\osort`), and tags are now produced without `@` (`\4` instead of `@4`, etc.).

Paths can be printed with a normal `|`, and \oplus and other relation symbols can be input more easily (see Section 2.1), though the package will also work with `$|` and `$\oplus`.

`langsci-avm` is not yet able to draw lines in elements of AVMs. This feature is planned for Version 0.3.

4.7 Spanning both columns

You can use the `multicol` package to span both columns in a (sub-)structure. Please remember that every structure has two columns, so the only sensible usage is

```
\multicolumn{2}{1}{...}
```

but only in the first column of a (sub-)structure. For a special usage case, see `\type` and `\type*` (which do not depend on `multicol`).

5 Caveats and planned features

1. There are currently no error messages. If you do not receive the intended output, please make sure that your code fits the syntax described in this documentation. If your code is fine but the output is not, please submit a bug report or feature request at <https://github.com/langsci/langsci-avm/issues>.

These features are planned for the future:

2. A check whether the delimiters are balanced, i.e. whether all (sub-)structures are closed by a `]`, `}`, etc.
3. Introduce the ability to draw (curved) lines between structures and elements.
4. Improve the appearance of (very) large angle brackets so that they vertically span the complete structure they enclose, maybe using `scalere1`.

6 Feedback and bug reports

Comments, usage reports, and feature requests are welcome! Please open an issue for any of these at <https://github.com/langsci/langsci-avm/issues>, or write to me at <mailto:felix.kopecky@langsci-press.org> if you feel the need for a feature not listed here, big or small.

7 Implementation

```
1 <*package>
2 <@@=avm>
3 \RequirePackage{xparse,array}
4 \ProvidesExplPackage {langsci-avm}
5   {2020-10-12} {0.3.0-rc3}
6   {AVMs and feature structures in LaTeX3}
7
8 \msg_new:nnnn {avm} {lfgoptionmissing}
9   { Missing~package~option~lfg~at~line~\msg_line_number: }
10  {
11    You~issued~a~command~in~line~\msg_line_number:~that~is~only~available~when~
12    the~lfg~package~option~is~enabled.
13  }
```

Let's first check for package options.

```
14 \bool_new:N \l__avm_lfg_bool
15 \bool_new:N \l__avm_tikz_bool
16 \DeclareOption{tikz}{ \bool_set_true:N \l__avm_tikz_bool }
17 \DeclareOption{lfg}{ \bool_set_true:N \l__avm_lfg_bool }
18 \ProcessOptions\relax
```

Handling for the TikZ package option.

```
19 \bool_if:NT \l__avm_tikz_bool
20 {
21   \RequirePackage{tikz}
22   \newcounter{l__avm_picture_counter}
23   \tl_new:N \l__avm_picture_name_prefix_tl
24 }
```

Handling for the LFG package option: If the semantic bracket is not available at the end of the preamble (i.e.) it was not loaded by another package, load `unicode-math` to provide the symbol.

```
25 \bool_if:NT \l__avm_lfg_bool
26 {
27   \cs_if_exist:NF \lBrack
28   {
29     \RequirePackage{etoolbox}
30     \AtEndPreamble { \RequirePackage{unicode-math} }
31   }
32 }
```

\avm This document command initialises an AVM. The first, optional argument is a key-value list of settings (see `\keys_define:nn` below) and the second is the AVM itself, given in the syntax described in this documentation.

`\avm` enters a group so that keys- and macro-assignments remain local. It then initialises the commands and shortcuts and any user customisation, sets its mode to `true` and assigns the keys as given in the optional argument (if any). After the wrapper `\avm_wrap:n` is called, the group is closed.

```

33 \NewDocumentCommand{\avm}{ O{} +m }
34 {
35   \c_group_begin_token
36   \keys_set:nn { avm } { #1 }
37   \__avm_initialise_document_commands:
38   \__avm_initialise_custom_commands:
39   \tl_use:N \l__avm_defined_commands_tl
40   \bool_set_true:N \l__avm_mode_bool
41   \__avm_wrap:n { #2 }
42   \c_group_end_token
43 }

```

(End definition for `\avm`. This function is documented on page 2.)

`\avmsetup` Forward the key-value settings given as the optional argument to `\avm` to the keys defined in `\keys_define:nn { avm }`. For the meaning of these keys and initial values, see Section 2.

```

44 \NewDocumentCommand{\avmsetup}{ m }
45 { \keys_set:nn { avm } { #1 } }
46
47 \keys_define:nn { avm }
48 {
49   stretch .tl_set:N      = \l__avm_arraystretch_tl,
50   stretch .initial:n    = {0.9},
51   columnsep .dim_set:N   = \l__avm_tabcolsep_dim,
52   columnsep .initial:n   = {.5ex},
53   delimfactor .int_set:N = \l__avm_delimfactor_int,
54   delimfactor .initial:n = {1000},
55   delimfall .dim_set:N   = \l__avm_delimshortfall_dim,
56   delimfall .initial:n   = {0pt},
57   attributes .code:n     = {\cs_set:Nn \__avm_font_attribute: {#1}},
58   attributes .initial:n  = {\scshape},
59   types .code:n          = {\cs_set:Nn \__avm_font_type: {#1}},
60   types .initial:n       = {\itshape},
61   values .code:n         = {\cs_set:Nn \__avm_font_value: {#1}},
62   values .initial:n      = {\itshape},
63   tags .code:n           = {\cs_set:Nn \__avm_font_tag: {#1}},
64   tags .initial:n        = {\footnotesize},
65   singleton .code:n      = {\cs_set:Nn \__avm_font_singleton: {#1}},
66   singleton .initial:n   = {\normalfont},
67   switch .code:n         = {\tl_set:Nn \__avm_mode_switch_character {#1}},
68   switch .initial:n      = { ! },
69   extraskip .dim_set:N   = \l__avm_extra_skip_dim,
70   extraskip .initial:n   = {\smallskipamount},
71   extraskip~in~every~row .bool_set:N = \l__avm_extraskip_bool,
72   customise .code:n      = {\cs_set:Nn \__avm_initialise_custom_commands: {#1}},
73   customise .initial:n   = { },
74   pic .bool_set:N        = \l__avm_picture_bool,
75   pic .default:n         = { true },

```

```

76   picname .tl_set:N      = \l__avm_picture_name_tl,
77   picname .initial:n    = {automatic},
78   id~align .code:n      = { \newcolumnntype{i}{#1} },
79   id~align .initial:n   = {1},
80   style .choice:,
81   style / narrow .code:n = {\int_set:Nn \l__avm_delimfactor_int {997}
82                               \dim_set:Nn \l__avm_delimshortfall_dim {5pt}},
83   }

```

(End definition for `\avmsetup`. This function is documented on page 4.)

`\avmdefinestyle` Define a style to be used together with the `style` key.

```

84 \NewDocumentCommand{\avmdefinestyle}{m m }
85 {
86   \keys_define:nn { avm }
87   {
88     style / #1 .code:n = { \keys_set:nn { avm } { #2 } }
89   }
90 }

```

(End definition for `\avmdefinestyle`. This function is documented on page 4.)

`\avmdefinecommand` A factory function that creates commands for the layout of sub-structures and saves them to `\l__avm_defined_commands_tl`. The first argument describes the command's name, the second any (optional) label. The manufactured definitions are activated in the AVM group so that they remain local.

```

91 \NewDocumentCommand{\avmdefinecommand}{m O{ } m }
92 {
93   \tl_put_right:Nn \l__avm_defined_commands_tl
94   {
95     \exp_args:Nc \DeclareDocumentCommand { #1 } { s }
96     {
97       #2 \IfBooleanF { ##1 } { & } \avmsetup{ #3 }
98     }
99   }
100 }

```

(End definition for `\avmdefinecommand`. This function is documented on page 7.)

`\l__avm_mode_bool` We need an auxiliary variable to store the current mode. `\l__avm_parens_tracker` is a stack for a future check whether the delimiters given to `\avm` are balanced. `\l__avm_defined_commands_tl` is a token list that stores any commands provided by the user via `\avmdefinecommand`

```

101 \bool_new:N \l__avm_mode_bool
102 \seq_new:N \l__avm_parens_tracker
103 \tl_new:N \l__avm_defined_commands_tl

```

(End definition for `\l__avm_mode_bool`, `\l__avm_parens_tracker`, and `\l__avm_defined_commands_tl`.)

`\seq_set_split:Nvn` In preparation for `\avm_wrap:n`, we need to split the user input at each occurrence of the escape character. Since the character is given in a variable, we need a variant of the sequence splitter that takes the *evaluation* of the variable, rather than the variable itself, as its second argument.

```

\tl_if_eq:VnTF

```

The second variant is useful for comparing the values of token list variables with token lists.

```
104 \cs_generate_variant:Nn \seq_set_split:Nnn { NVn }
105 \cs_generate_variant:Nn \tl_if_eq:nnTF { VnTF }
```

(End definition for `\seq_set_split:Nnn` and `\tl_if_eq:VnTF`.)

`\l__avm_in_first_column` A boolean to check whether we are in the first column (value `true`) or in the second (value `false`).

```
106 \bool_new:N \l__avm_in_first_column
```

(End definition for `\l__avm_in_first_column`.)

`__avm_init_first_column:` These macros apply the settings for the columns in a (sub-)structure. They take care of font selection and report the currently active column back to the system. Knowing which column is active is important when closing the (sub-)structure. If the structure is closed without a second column present, we need to skip back `2\tabcolsep`.

```
107 \cs_new:Nn \__avm_init_first_column:
108 {
109   \bool_set_true:N \l__avm_in_first_column
110   \normalfont\__avm_font_attribute:
111 }
112
113 \cs_new:Nn \__avm_init_second_column:
114 {
115   \bool_set_false:N \l__avm_in_first_column
116   \normalfont\__avm_font_value:
117 }
```

(End definition for `__avm_init_first_column:` and `__avm_init_second_column:.`)

`__avm_deinit_first_column:` These commands control settings that are applied after each column is exited. The single check here is whether italics is currently in use. If it is, the the italic correction is automatically applied. This replaces the user-configurable setting `apptovalues` from previous versions.

```
118
119 \tl_const:Nn \l__avm_italics_tl {it}
120
121 \cs_new:Nn \__avm_deinit_first_column:
122 {
123   \tl_if_eq:NNT \f@shape \l__avm_italics_tl {\!/}
124 }
125
126 \cs_new:Nn \__avm_deinit_second_column:
127 {
128   \tl_if_eq:NNT \f@shape \l__avm_italics_tl {\!/}
129 }
```

(End definition for `__avm_deinit_first_column:` and `__avm_deinit_second_column:.`)

`__avm_kern_unused_columns:` A helper macro to fill the horizontal space if a row is ended prematurely, i.e. if no `&` is present.

```
130 \cs_new:Nn \__avm_kern_unused_columns:
131 {
```

```

132   \bool_if:NTF \l__avm_in_first_column
133   { \span\hspace*{-2\tabcolsep} }
134   { }
135 }

```

(End definition for `__avm_kern_unused_columns:`)

`__avm_extra_skip:` This function is used together with the delimiter replacements. It checks whether the delimiter is followed by a line break, in which case an extra skip is automatically inserted

```

136 \cs_new:Nn \__avm_extra_skip:
137 {
138   \peek_meaning_ignore_spaces:NTF \ \ {\vspace*{\l__avm_extra_skip_dim}} {}
139 }

```

(End definition for `__avm_extra_skip:`)

`__avm_module_begin:` The replacement instructions for `__avm_parse:n`

```

\__avm_module_end:
etc.
140 \cs_new:Nn \__avm_module_begin:
141 {
142   \begin{tabular}{@{}}
143       >{\__avm_init_first_column:}l
144       <{\__avm_deinit_first_column:}
145       >{\__avm_init_second_column:}l
146       <{\__avm_deinit_second_column:}
147       @{}
148   }
149 \cs_new:Nn \__avm_module_end:
150 {
151   \__avm_kern_unused_columns:
152   \end{tabular}
153 }
154 \cs_new:Nn \__avm_replace_lbrace:
155 {
156   \__avm_parse_output:nw
157   { \c_math_toggle_token\left\lbrace\__avm_module_begin: }
158 }
159 \cs_new:Nn \__avm_replace_rbrace:
160 {
161   \__avm_parse_output:nw
162   { \__avm_module_end:\right\rbrace\c_math_toggle_token\__avm_extra_skip: }
163 }
164 \cs_new:Nn \__avm_replace_lbrack:
165 {
166   \__avm_parse_output:nw
167   { \c_math_toggle_token\left\lbrack\__avm_module_begin: }
168 }
169 \cs_new:Nn \__avm_replace_rbrack:
170 {
171   \__avm_parse_output:nw
172   { \__avm_module_end:\right\rbrack\c_math_toggle_token\__avm_extra_skip: }
173 }
174 \bool_if:NTF \l__avm_lfg_bool
175 {
176   \cs_new:Nn \__avm_replace_llbrack:

```

```

177     {
178       \__avm_parse_output:nw
179       { \c_math_toggle_token\left\lBrack\__avm_module_begin:  }
180     }
181   \cs_new:Nn \__avm_replace_rrbrack:
182   {
183     \__avm_parse_output:nw
184     { \__avm_module_end:\right\rBrack\c_math_toggle_token\__avm_extra_skip:  }
185   }
186 }
187 {
188   \cs_new:Nn \__avm_replace_llbrack:
189   {
190     \__avm_parse_output:nw
191     {
192       \msg_warning:nn {avm}{!lfgoptionmissing}
193       \c_math_toggle_token\left.\__avm_module_begin:
194     }
195   }
196   \cs_new:Nn \__avm_replace_rrbrack:
197   {
198     \__avm_parse_output:nw
199     {
200       \msg_warning:nn {avm}{!lfgoptionmissing}
201       \__avm_module_end:\right.\c_math_toggle_token\__avm_extra_skip:
202     }
203   }
204 }
205 \cs_new:Nn \__avm_replace_lparen:
206 {
207   \__avm_parse_output:nw
208   { \c_math_toggle_token\left(\__avm_module_begin:  }
209 }
210 \cs_new:Nn \__avm_replace_rparen:
211 {
212   \__avm_parse_output:nw
213   { \__avm_module_end:\right)\c_math_toggle_token\__avm_extra_skip:  }
214 }
215 \cs_new:Nn \__avm_replace_langle:
216 {
217   \__avm_parse_output:nw
218   { \c_math_toggle_token\left<\__avm_module_begin:  }
219 }
220 \cs_new:Nn \__avm_replace_rangle:
221 {
222   \__avm_parse_output:nw
223   { \__avm_module_end:\right>\c_math_toggle_token\__avm_extra_skip:  }
224 }
225 \cs_new:Nn \__avm_replace_plus:
226 {
227   \__avm_parse_output:nw { \ensuremath { \oplus \! } }
228 }
229 \cs_new:Nn \__avm_replace_minus:
230 {

```



```

231   \__avm_parse_output:nw { \ensuremath { \ominus \! } }
232 }
233 \cs_new:Nn \__avm_replace_circle:
234 {
235   \__avm_parse_output:nw { \ensuremath { \bigcirc \! } }
236 }

```

(End definition for `__avm_module_begin:`, `__avm_module_end:`, and etc..)

```

\tag
\type
\pунк
\node
\id
237 \cs_new:Npn \__avm_controls_tag:n #1
238 { \fboxsep.25ex\fbox{\normalfont\__avm_font_tag: #1} }
239 \cs_new:Npn \__avm_controls_type:n #1
240 { \c_group_begin_token\normalfont\__avm_font_type: #1\c_group_end_token }
241 \cs_new_protected:Npn \__avm_controls_type_starred:n #1
242 {
243   \bool_set_false:N \l__avm_in_first_column
244   \normalfont\__avm_font_type: #1
245   \__avm_deinit_second_column:\span\hspace*{-2\tabcolsep}
246   \peek_meaning_ignore_spaces:NTF \ \ {} {\}
247 }
248 \cs_new_protected:Npn \__avm_controls_punk:nn #1 #2
249 {
250   \bool_set_false:N \l__avm_in_first_column
251   \normalfont\c_group_begin_token\__avm_font_attribute:#1%
252   \c_group_end_token\hspace{2\tabcolsep}%
253   \c_group_begin_token\__avm_font_type: #2\c_group_end_token%
254   \__avm_deinit_second_column:\span\hspace*{-2\tabcolsep}
255   \peek_meaning_ignore_spaces:NTF \ \ {} {\}
256 }
257
258 \cs_new:Nn \__avm_initialise_document_commands:
259 {
260   \def\arraystretch{\tl_use:N \l__avm_arraystretch_tl}
261   \dim_set_eq:NN \tabcolsep \l__avm_tabcolsep_dim
262   \int_set_eq:NN \delimiterfactor \l__avm_delimfactor_int
263   \dim_set_eq:NN \delimitershortfall \l__avm_delimshortfall_dim
264   \cs_if_exist:NTF \tag
265     { \RenewDocumentCommand{\tag}{m}{ \__avm_controls_tag:n {##1} } }
266     { \NewDocumentCommand{\tag}{m}{ \__avm_controls_tag:n {##1} } }
267   \cs_if_exist:NTF \0
268     { \RenewDocumentCommand{\0}{}{ \__avm_controls_tag:n {0} } }
269     { \NewDocumentCommand{\0}{}{ \__avm_controls_tag:n {0} } }
270   \cs_if_exist:NTF \1
271     { \RenewDocumentCommand{\1}{}{ \__avm_controls_tag:n {1} } }
272     { \NewDocumentCommand{\1}{}{ \__avm_controls_tag:n {1} } }
273   \cs_if_exist:NTF \2
274     { \RenewDocumentCommand{\2}{}{ \__avm_controls_tag:n {2} } }
275     { \NewDocumentCommand{\2}{}{ \__avm_controls_tag:n {2} } }
276   \cs_if_exist:NTF \3
277     { \RenewDocumentCommand{\3}{}{ \__avm_controls_tag:n {3} } }
278     { \NewDocumentCommand{\3}{}{ \__avm_controls_tag:n {3} } }
279   \cs_if_exist:NTF \4
280     { \RenewDocumentCommand{\4}{}{ \__avm_controls_tag:n {4} } }

```

```

281 { \NewDocumentCommand{\4}{}{ \_avm_controls_tag:n {4} } }
282 \cs_if_exist:NTF \5
283 { \RenewDocumentCommand{\5}{}{ \_avm_controls_tag:n {5} } }
284 { \NewDocumentCommand{\5}{}{ \_avm_controls_tag:n {5} } }
285 \cs_if_exist:NTF \6
286 { \RenewDocumentCommand{\6}{}{ \_avm_controls_tag:n {6} } }
287 { \NewDocumentCommand{\6}{}{ \_avm_controls_tag:n {6} } }
288 \cs_if_exist:NTF \7
289 { \RenewDocumentCommand{\7}{}{ \_avm_controls_tag:n {7} } }
290 { \NewDocumentCommand{\7}{}{ \_avm_controls_tag:n {7} } }
291 \cs_if_exist:NTF \8
292 { \RenewDocumentCommand{\8}{}{ \_avm_controls_tag:n {8} } }
293 { \NewDocumentCommand{\8}{}{ \_avm_controls_tag:n {8} } }
294 \cs_if_exist:NTF \9
295 { \RenewDocumentCommand{\9}{}{ \_avm_controls_tag:n {9} } }
296 { \NewDocumentCommand{\9}{}{ \_avm_controls_tag:n {9} } }
297 \cs_if_exist:NTF \type
298 { \RenewDocumentCommand{\type}{s m}
299 {
300 \IfBooleanTF { ##1 }
301 { \_avm_controls_type_starred:n {##2} }
302 { \_avm_controls_type:n {##2} }
303 }
304 }
305 { \NewDocumentCommand{\type}{s m}
306 {
307 \IfBooleanTF { ##1 }
308 { \_avm_controls_type_starred:n {##2} }
309 { \_avm_controls_type:n {##2} }
310 }
311 }
312 \cs_if_exist:NTF \punk
313 { \RenewDocumentCommand{\punk}{m m}
314 { \_avm_controls_punk:nn {##1}{##2} } }
315 { \NewDocumentCommand{\punk}{m m}
316 { \_avm_controls_punk:nn {##1}{##2} } }
317 \DeclareDocumentCommand{\id}{m m}
318 {%
319 \hcoffin_set:Nw \l_tmpa_coffin
320 \bgroup
321 \def\arraystretch{.5}
322 \begin{tabular}[b]{@{}>{\scriptstyle}i<{\scriptstyle}i@{}}
323 ##1
324 \end{tabular}
325 \egroup
326 \hcoffin_set_end:
327 \hcoffin_set:Nw \l_tmpb_coffin ##2 \hcoffin_set_end:
328 \coffin_join:NnnNnnnn \l_tmpb_coffin {1}{H}
329 \l_tmpa_coffin {r}{H}{ Opt }{ -\coffin_dp:N \l_tmpb_coffin } %-\coffin_wd:N \l_tmpa_coffin
330 \coffin_typeset:Nnnnn \l_tmpb_coffin {1}{vc}{Opt}{Opt}
331 }
332 \bool_if:NT \l_avm_tikz_bool
333 {

```

The last of the bunch is only loaded if Ti&Z is loaded as well:

```

334 \tl_if_eq:VnTF \l__avm_picture_name_tl {automatic}
335 {
336   \stepcounter{l__avm_picture_counter}
337   \tl_set:Nn \l__avm_picture_name_prefix_tl
338     {avm-\tl_use:N \thel__avm_picture_counter}
339 }
340 {
341   \tl_set_eq:NN \l__avm_picture_name_prefix_tl \l__avm_picture_name_tl
342 }
343 \DeclareDocumentCommand{\node}{m m}
344 {
345   \tikz [remember~picture,baseline=(\l__avm_picture_name_prefix_tl-##1.base)]
346   \node [inner~sep=0pt] (\l__avm_picture_name_prefix_tl-##1) {\strut ##2};
347 }
348 }
349 }

```

(End definition for `\tag` and others. These functions are documented on page 3.)

`__avm_wrap:n` The wrapper that first splits the input to `\avm` at each occurrence of `__avm_mode_switch_character` and then inverses `\l__avm_mode_bool`. It then calls the parser (`__avm_parse:n`) for each splitted sequence. This wrapping is necessary because there is no known expandable way to switch a boolean.

```

350 \cs_new_protected:Npn \__avm_wrap:n #1
351 {
352   \seq_set_split:NVn \l__avm_wrapper_seq
353     \__avm_mode_switch_character { #1 }
354   \seq_map_inline:Nn \l__avm_wrapper_seq
355     {
356       \exp_args:No \exp_not:o
357       { \__avm_parse:n {##1} }
358       \bool_set_inverse:N \l__avm_mode_bool
359     }
360 }

```

(End definition for `__avm_wrap:n`.)

`__avm_parse:n` Finally, the parser. It is build on `\@@_act:NNNnn` from 13t1 (see the sub-section *Token by token changes*). Many thanks to Phelype Oleinik for help on this, and in particular on help with expansion.

```

361 \cs_new:Npn \__avm_parse:n #1
362 {
363   \exp:w
364   \group_align_safe_begin:
365   \__avm_parse_loop:w #1
366   \q_recursion_tail \q_recursion_stop
367   \__avm_result:n { }
368 }
369
370 \cs_new:Npn \__avm_end:w \__avm_result:n #1
371 {
372   \group_align_safe_end:
373   \exp_end:
374   #1

```

```

375     }
376
377 \cs_new:Npn \__avm_parse_loop:w #1 \q_recursion_stop
378 {
379     \tl_if_head_is_N_type:nTF {#1}
380     {
381         \__avm_N_type:N #1 \q_recursion_stop
382     }
383     {
384         \tl_if_head_is_group:nTF {#1}
385         { \__avm_replace_group:nw #1 \q_recursion_stop }
386         { \__avm_replace_space:w #1 \q_recursion_stop }
387     }
388 }
389
390 \cs_new:Npn \__avm_N_type:N #1
391 {
392     \quark_if_recursion_tail_stop_do:Nn #1 { \__avm_end:w }
393     \bool_if:NTF \l__avm_mode_bool
394     { \__avm_replace:N #1 }
395     { \__avm_replace_none:N #1 }
396 }
397
398 \cs_new:Npn \__avm_replace_none:N #1
399 {
400     \__avm_parse_output:nw {#1}
401 }
402
403 \cs_new:Npn \__avm_replace:N #1
404 {
405     \str_case:nnF {#1}
406     {
407         { \+ }{ \__avm_replace_plus: }
408         { \- }{ \__avm_replace_minus: }
409         { \shuffle }{ \__avm_replace_circle: }
410         { [ ] }{ \__avm_replace_lbrack: }
411         { ] }{ \__avm_replace_rbrack: }
412         { \[ ] }{ \__avm_replace_llbrack: }
413         { \] ] }{ \__avm_replace_rrbrack: }
414         { ( ) }{ \__avm_replace_lparen: }
415         { ) }{ \__avm_replace_rparen: }
416         { \{ }{ \__avm_replace_lbrace: }
417         { \} }{ \__avm_replace_rbrace: }
418         { < }{ \__avm_replace_langle: }
419         { > }{ \__avm_replace_rangle: }
420     }
421     { \__avm_replace_none:N #1 }
422 }
423
424 \cs_new:Npn \__avm_replace_group:nw #1
425 { \exp_args:NNo \exp_args:No \__avm_replace_group:n { \__avm_parse:n {#1} } }
426
427 \cs_new:Npn \__avm_replace_group:n #1 { \__avm_parse_output:nw { {#1} } }
428

```

```

429 \exp_last_unbraced:NNo
430 \cs_new:Npn \__avm_replace_space:w \c_space_tl { \__avm_parse_output:nw { ~ } }
431
432 \cs_new:Npn \__avm_parse_output:nw #1 #2 \q_recursion_stop \__avm_result:n #3
433 { \__avm_parse_loop:w #2 \q_recursion_stop \__avm_result:n {#3 #1 } }

(End definition for \__avm_parse:n.)
434 \</package>

```