# langsci-avm

Felix Kopecky\*

Version 0.1.0-beta – 9th March 2020

### 1 Introduction

langsci-avm is a LaTeX3 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.

Switching from avm to langsci-avm will require some, though hopefully minimal, changes to the code. In particular, the "active mode" has disappeared, there is now a single way of sorting (see \type), and tags are now produced without @ (\4 instead of @4, etc.)

## 1.1 Example

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

# 1.2 Acknowledgements

Thanks to Phelype Oleinik for help on recursion and expansion with LATEX3. Thanks to Ahmet Bilal Özdemir and Stefan Müller for their help in planning and testing this package.

# 2 User interface

# 2.1 Typing structures and settings

 $\verb| (options)| \{ \langle structure \rangle \}|$ 

The heart of this package and its root document comand is \avm. It currently runs only in text mode, but can be run in math mode if it is escaped with \avm\. In the scope of the command, delimiter characters are processed to open and close (sub-)structures, as described in Section 2.2. For a description of the \langle options \rangle, see \avmsetup.

#### \avmsetup

 $\alpha \{\langle options \rangle\}$ 

 $\langle options \rangle$  is a comma-separated list of key = value settings. The  $\{\langle options \rangle\}$  are the same as in  $\langle options \rangle$ . When inserted in  $\langle options \rangle$ , they apply locally, and globally if given to  $\langle options \rangle$ . Local settings always override global ones, and you can have any feasible number of  $\langle options \rangle$  in your document.

 $stretch = \langle factor \rangle$ 

(initially 1.15)

Define \arraystretch, i.e. the factor of line spacing. Increasing this may be useful if there is too little vertical space between delimiters in subsequent rows.

 $columnsep = \langle length \rangle$ 

(initially 0.5ex)

Define the \arraycolsep, i.e. horizontal space before and after any column. The first and second column will have 1\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 \delimiterfactor. 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 Opt)

Controls \delimitershortfall, i.e. the maximum height that the delimiters can be shorter than the enclosed structure. The default Opt ensure that the delimiters are not shorter than the contents.

 $attributes = \langle font \ settings \rangle$ 

(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  $\t$ ag 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.

## 2.2 Commands available in the scope of \avm

```
[...] [ \langle structure \rangle ] 
 \langle ... \rangle \langle structure \rangle > 
 \langle \langle structure \rangle \rangle \langle \langle \langle \langle \rangle \ran
```

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, they 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.

A  $\langle structure \rangle$  is basically the content of a stylised array: The columns are separated by & and a new line is entered with  $\setminus \setminus$ .

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 exactly one & in your  $\langle structure \rangle$ . In the current beta version, display issues may appear in some structures if none is given.

!...! !  $\langle text \rangle$  !

\9

Escapes the avm mode so that all delimiters can be used as usual characters. If you need ! as a regular character, see \avmsetup to change the switch option.

```
\label{eq:lambda} $$  \ag {\langle identifier \rangle } \\ 1, \ 2, \ 3, \ 4, \ 5, \ 6, \ 7, \ 8, \ 9
```

\tag puts its {\langle identifier \rangle} in a box, more precisely an \fbox. Within the box, the tags font is applied. \1, ..., \9 are shortcuts to \tag and place the respective number in the box. For example, \4 is equivalent to \tag{4}. The shortcuts do not take any arguments.

If you want to use this command outside an AVM, you can obtain, for example, 4, by using \avm{\4}, or the equivalent {\fboxsep.25ex\fbox{\footnotesize 4}}.

```
\type
\type*
```

```
\type(*) \{\langle type \rangle\}
```

Will typeset the  $\langle type \rangle$  in the types font (roman italics by default). The starred variant  $\type*$  will span the complete (sub-)structure and can only be placed in the first column of this structure. After the starred  $\type*$ , a  $\t$  is recommended, but can be omitted in "normal" cases.

\punk

```
\punk \{\langle attribute \rangle\}\{\langle type \rangle\}
```

Some  $\langle attributes \rangle$  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  $\langle attribute \rangle$  and  $\langle type \rangle$ . After \punk, a \\ is recommended, but can be omitted in "normal" cases.

+ In the scope of \avm, + comes out as "⊕". To obtain "+", please use the escape mode, as in !+!.

# 3 Applications

#### 3.1 Disjunctions and other relations

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

#### 3.2 Use as a vector

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

### 3.3 Example environments: gb4e and expex

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:

The same can be achieved with expex using an experimental syntax:

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

A future version of langsci-avm will include a more user-friendly approach. There is currently no known way of adjusting the alignment with linguex.

#### 3.4 forest

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

### 3.5 Tweaking the attribute font with with fontspec

The attributes in your structure are usually type set in SMALL CAPS. That means that your input should usually be lowercase, unless capitals along small capitals in that attribute's description make sense. Some fonts also provide "old style" figures (also known as "text" or "medieval" figures). Those can be enabled with fontspec's Numbers=OldStyle feature.

The following example is output in Libertinus, since the font for this documentation does not support the necessary font features.

# 3.6 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

```
\mathcal{L}_{2}{1}{\ldots}
```

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).

# 4 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.
- 2. The package currently assumes that it is called in text mode.

These features are planned for the future:

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

# 5 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.

# 6 Implementation

```
1 (*package)
2 (@@=avm)
3 \RequirePackage{xparse,array}
4 \ProvidesExplPackage {langsci-avm}
5 {2020-03-09} {0.1.0-beta}
6 {AVMs and feature structures in LaTeX3}
```

\avm This document command initialises an AVM. The first, optional argumet 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-assignemts remain local. It then initialises the commands and shortcuts made locally available, 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.

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

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 2.1.

```
16 \NewDocumentCommand{\avmsetup}{ m }
    { \keys_set:nn { avm } { #1 } }
18
  \keys_define:nn { avm }
19
    {
20
21
      stretch .code:n
                              = {\def\arraystretch{#1}},
      stretch .initial:n
                              = \{1.15\},
23
      columnsep .dim_set:N
                             = \arraycolsep,
      columnsep .initial:n
                              = \{.5ex\},\
      delimfactor .int_set:N = \delimiterfactor,
26
      delimfactor .initial:n = {1000},
      delimfall .dim_set:N = \delimitershortfall,
27
                              = \{0pt\},\
      delimfall .initial:n
28
      attributes .code:n
                              = {\cs_set:Nn \__avm_font_attribute: {#1}},
29
      attributes .initial:n = {\scshape},
30
      types .code:n
                              = {\cs_set:Nn \__avm_font_type: {#1}},
31
      types .initial:n
                              = {\itshape},
32
33
      values .code:n
                              = {\cs_set:Nn \__avm_font_value: {#1}},
      values .initial:n
                              = {\itshape},
      tags .code:n
                              = {\cs_set:Nn \__avm_font_tag: {#1}},
                              = {\footnotesize},
      tags .initial:n
```

```
switch .code:n = {\tl_set:Nn \_avm_mode_switch_character {#1}},
switch .initial:n = { ! }
}
```

(End definition for \avmsetup. This function is documented on page 3.)

\l\_avm\_math\_bool \l\_avm\_mode\_bool \l\_avm\_parens\_tracker We need an auxiliary variable to store the current mode. The math mode boolean is already created, but it will have an effect only in a later version which will include a check whether \avm is called in math mode. \l\_\_avm\_parens\_tracker is a stack for a future check whether the delimiters given to \avm are balanced.

```
40 \bool_new:N \l__avm_math_bool
41 \bool_new:N \l__avm_mode_bool
42 \seq_new:N \l__avm_parens_tracker

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

\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.

```
43 \cs_generate_variant:Nn \seq_set_split:Nnn { NVn } (End definition for \seq_set_split:NVn.)
```

The replacement instructions for \\_\_avm\_parse:n

```
44 \cs_new:Nn \__avm_module_begin:
45
    {
      \begin{array}{
46
        >{\c_math_toggle_token\normalfont\__avm_font_attribute:}l
47
        <\c_math_toggle_token
48
        >{\c_math_toggle_token\normalfont\__avm_font_value:}l
49
        <\c_math_toggle_token}
50
    }
51
52 \cs_new:Nn \__avm_module_end:
    { \end{array} }
54 \cs_new:Nn \__avm_replace_lbrace:
55
56
      \__avm_parse_output:nw
        { \c_math_toggle_token\left\lbrace\__avm_module_begin: }
57
    }
58
  \cs_new:Nn \__avm_replace_rbrace:
59
    {
60
61
      \__avm_parse_output:nw
         { \__avm_module_end:\right\rbrace\c_math_toggle_token }
62
    }
63
  \cs_new:Nn \__avm_replace_lbrack:
66
        _avm_parse_output:nw
        { \c_math_toggle_token\left\lbrack\__avm_module_begin: }
67
    }
68
  \cs_new:Nn \__avm_replace_rbrack:
69
70
      \__avm_parse_output:nw
71
        { \__avm_module_end:\right\rbrack\c_math_toggle_token }
```

```
}
       74 \cs_new:Nn \__avm_replace_lparen:
       75
                _avm_parse_output:nw
       76
               { \c_math_toggle_token\left(\__avm_module_begin: }
       77
           }
       78
         \cs_new:Nn \__avm_replace_rparen:
       79
       80
       81
               _avm_parse_output:nw
               { \__avm_module_end:\right)\c_math_toggle_token }
       82
           }
       83
         \cs_new:Nn \__avm_replace_langle:
       84
           {
       85
       86
                _avm_parse_output:nw
               { \c_math_toggle_token\left<\__avm_module_begin: }
       87
       88
         \cs_new:Nn \__avm_replace_rangle:
       89
       90
           {
       91
              { \__avm_module_end:\right>\c_math_toggle_token }
           }
       93
         \cs_new:Nn \__avm_replace_plus:
       95
              \__avm_parse_output:nw { \ensuremath { \oplus } }
       96
      (End definition for \__avm_module_begin:, \__avm_module_end:, and etc..)
\tag
\type
       98 \cs_new:Npn \__avm_controls_tag:n #1
\punk
           { \fboxsep.25ex\fbox{\normalfont\_avm_font_tag: #1} }
       100 \cs_new:Npn \__avm_controls_type:n #1
           {\normalfont\__avm_font_type: #1}
       102
         \cs_new_protected:Npn \__avm_controls_type_starred:n #1
             \normalfont\__avm_font_type: #1\span
       104
             \peek_meaning_ignore_spaces:NTF \\ {} {\\}
       105
           }
       106
         \cs_new_protected:Npn \__avm_controls_punk:nn #1 #2
       107
           {
       108
              \normalfont\c_group_begin_token\__avm_font_attribute:#1%
       109
              \c_group_end_token\hspace{2\arraycolsep}%
              \c_group_begin_token\__avm_font_type: #2\c_group_end_token%
              \span\peek_meaning_ignore_spaces:NTF \\ {} {\\}
           }
       113
         \cs_new:\n\__avm_initialise_document_commands:
       115
       116
             \cs_if_exist:NTF \tag
               118
                                                  \__avm_controls_tag:n {##1} } }
                  \NewDocumentCommand{\tag}{m}{
       119
              \cs_if_exist:NTF \1
       120
                  \RenewDocumentCommand{\1}{}{
                                                  \_avm_controls_tag:n {1} } }
                  \NewDocumentCommand{\1}{}{
                                                  \_avm_controls_tag:n {1} } }
```

```
\cs_if_exist:NTF \2
         { \RenewDocumentCommand{\2}{}{
                                             \_avm_controls_tag:n {2} } }
124
           \NewDocumentCommand{\2}{}{
                                             \_avm_controls_tag:n {2} } }
125
       \cs_if_exist:NTF \3
126
           \RenewDocumentCommand{\3}{}{
                                              \__avm_controls_tag:n {3} } }
127
           \NewDocumentCommand{\3}{}{
                                              \__avm_controls_tag:n {3} } }
128
       \cs_if_exist:NTF \4
129
           \RenewDocumentCommand{\4}{}{
                                             \__avm_controls_tag:n {4} } }
130
           \NewDocumentCommand{\4}{}{
                                              \_avm_controls_tag:n {4} } }
       \cs_if_exist:NTF \5
132
           \RenewDocumentCommand{\5}{}{
133
                                             \__avm_controls_tag:n {5} } }
           \NewDocumentCommand{\5}{}{
                                              \__avm_controls_tag:n {5} } }
134
       \cs_if_exist:NTF \6
135
           \RenewDocumentCommand{\6}{}{
                                             \_avm_controls_tag:n {6} } }
136
            \NewDocumentCommand{\6}{}{
                                              \_avm_controls_tag:n {6} } }
       \cs_if_exist:NTF \7
138
           \RenewDocumentCommand{\7}{}{
                                             \_avm_controls_tag:n {7} } }
139
            \NewDocumentCommand{\7}{}{
                                             \_avm_controls_tag:n {7} } }
140
       \cs_if_exist:NTF \8
           \RenewDocumentCommand{\8}{}{
                                             \__avm_controls_tag:n {8} } }
           \NewDocumentCommand{\8}{}{
                                             \__avm_controls_tag:n {8} } }
143
       \cs_if_exist:NTF \9
144
         { \RenewDocumentCommand{\9}{}{
                                             \_avm_controls_tag:n {9} } }
145
         { \NewDocumentCommand{\9}{}{
                                             \_avm_controls_tag:n {9} } }
146
       \cs_if_exist:NTF \type
147
         { \RenewDocumentCommand{\type}{s m}
148
149
             \IfBooleanTF { ##1 }
150
                  \_avm_controls_type_starred:n {##2} }
                  \_avm_controls_type:n {##2} }
            }
153
154
          }
             \NewDocumentCommand{\type}{s m}
          {
156
              \IfBooleanTF { ##1 }
                   \_avm_controls_type_starred:n {##2} }
158
                   \_avm_controls_type:n {##2} }
159
             }
160
           }
161
       \cs_if_exist:NTF \punk
           \RenewDocumentCommand{\punk}{m m}
           { \__avm_controls_punk:nn {##1}{##2} } }
           \NewDocumentCommand{\punk}{m m}
165
             \_avm_controls_punk:nn {##1}{##2} } }
166
    }
167
```

(End definition for \tag, \type, and \punk. These functions are documented on page 4.)

\\_\_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.

```
168 \cs_new_protected:Npn \__avm_wrap:n #1
169
     {
```

```
\seq_set_split:NVn \l__avm_wrapper_seq
\_avm_mode_switch_character { #1 }
\seq_map_inline:Nn \l_avm_wrapper_seq
\tag{
\tag{
\texp_args:No \exp_not:o}
\tag{ \_avm_parse:n {\##1} }
\tag{
\tag{\texp_avm_parse:No \l_avm_mode_bool}
\tag{ \_avm_mode_switch_character { #1 }
\tag{\texp_args:No \exp_not:o}
\tag{\texp_avm_parse:n \text{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\mathref{\ma
```

 $(End\ definition\ for\ \_avm\_wrap:n.)$ 

\\_\_avm\_parse:n

Finnaly, the parser. It is build on \\_avm\_act:NNNnn from 13t1 (see the sub-section *To-ken by token changes*). Many thanks to Phelype Oleinik for help on this, and in particular on help with expansion.

```
\cs_new:Npn \__avm_parse:n #1
179
     {
180
         \exp:w
181
         \group_align_safe_begin:
182
            \__avm_parse_loop:w #1
183
           \q_recursion_tail \q_recursion_stop
           \__avm_result:n { }
     }
186
187
   \cs_new:Npn \__avm_end:w \__avm_result:n #1
188
189
       \group_align_safe_end:
190
       \exp_end:
191
192
     }
193
194
  \cs_new:Npn \__avm_parse_loop:w #1 \q_recursion_stop
       \tl_if_head_is_N_type:nTF {#1}
197
198
            \_ = avm_N_type:N #1 q_recursion_stop
199
         }
200
         {
201
            \tl_if_head_is_group:nTF {#1}
202
             { \__avm_replace_group:nw #1 \q_recursion_stop }
203
204
              { \_avm_replace_space:w #1 \q_recursion_stop }
     }
   \cs_new:Npn \__avm_N_type:N #1
209
       \quark_if_recursion_tail_stop_do:Nn #1 { \__avm_end:w }
       \bool_if:NTF \l__avm_mode_bool
211
         { \__avm_replace:N #1 }
         { \__avm_replace_none:N #1 }
214
     }
215
  \cs_new:Npn \__avm_replace_none:N #1
     {
```

```
\__avm_parse_output:nw {#1}
218
219
220
   \cs_new:Npn \__avm_replace:N #1
221
       \str_case:nnF {#1}
223
224
           { + }{ \__avm_replace_plus: }
225
           { [ }{ \_avm_replace_lbrack: }
           { ] }{ \_avm_replace_rbrack: }
227
           { ( }{ \_avm_replace_lparen: }
228
           { \ \ } { \ \ } { \ \ } { \ \ } { \ \ }
229
           { \ \ \ }{ \ \ \ }{ \ \ \ \ }
230
           { \} }{ \__avm_replace_rbrace: }
231
           { < }{ \__avm_replace_langle: }</pre>
232
                }{ \__avm_replace_rangle: }
234
         { \__avm_replace_none:N #1 }
235
     }
236
   \cs_new:Npn \__avm_replace_group:nw #1
     { \exp_args:NNo \exp_args:No \__avm_replace_group:n { \__avm_parse:n {#1} } }
239
240
   \cs_{new:Npn \ \_avm\_replace\_group:n \#1 { \ \_avm\_parse\_output:nw { $ \#1} } }
241
242
   \exp_last_unbraced:NNo
243
   \cs_new:Npn \__avm_replace_space:w \c_space_tl { \__avm_parse_output:nw { ~ } }
244
   \cs_new:Npn \__avm_parse_output:nw #1 #2 \q_recursion_stop \__avm_result:n #3
     { \_avm_parse_loop:w #2 \q_recursion_stop \_avm_result:n {#3 #1 } }
(End definition for \__avm_parse:n.)
248 (/package)
```