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. When you come from avm, please see Section 3.5
for a quick conversion guide.

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

1.1 Example

\avm{
  \begin{array}{c}
  \text{ctxt} \& \begin{array}{c}
  \text{max-qud} \\
  \text{sal-utt} \& \begin{array}{c}
  \text{cat} \\
  \text{cont} \langle \text{ind} & i \rangle
  \end{array}
  \end{array}
  \\
  \end{array}
\}

\[
\begin{bmatrix}
\text{CTXT} & \text{MAX-QUD} \\
\text{SAL-UTT} & \begin{bmatrix}
\text{CAT} \\
\text{CONT} \langle \text{IND} & i \rangle
\end{bmatrix}
\end{bmatrix}
\]

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 contributions in planning and testing
this package.

∗mailto:felix.kopecky@langsci-press.org. Please submit bug reports and feature requests to
2 User interface

2.1 Typing structures and settings

\[ \text{The heart of this package and its root document comand is \texttt{avm}. It currently runs only in text mode, but can be run in math mode if it is escaped with $\texttt{avm}$}. \]

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 \texttt{(options)}, see \texttt{avmsetup}.

\[ \text{\texttt{avmsetup}} \]

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

\begin{itemize}
\item \texttt{stretch} = \langle factor \rangle \quad \text{(initially 1.15)}
  \text{Define \texttt{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.}
\item \texttt{columnsep} = \langle length \rangle \quad \text{(initially 0.5ex)}
  \text{Define the \texttt{arraycolsep}, i.e. horizontal space before and after any column. The first and second column will have 1\texttt{columnsep} to the left and right, respectively. Between the two the distance is 2\texttt{columnsep}. Using relative units (like ex or em) may be a good idea so that \texttt{columnsep} scales well with changes in font size.}
\item \texttt{delimfactor} = \langle factor \rangle \quad \text{(initially 1000)}
  \text{Sets \texttt{delimiterfactor}. The calculation for the minimum height of a delimiter is } y \cdot \frac{f}{1000}, \text{where } y \text{ is the height of the content and } f \text{ the value of \texttt{delimfactor}. The default 1000 ensure that the delimiters’ height is at least that of the structure.}
\item \texttt{delimfall} = \langle length \rangle \quad \text{(initially 0pt)}
  \text{Controls \texttt{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.}
\item \texttt{attributes} = \langle font settings \rangle \quad \text{(initially \texttt{scshape})}
  \text{The font for attributes, i.e. the first column of each structure.}
\item \texttt{values} = \langle font settings \rangle \quad \text{(initially \texttt{itshape})}
  \text{The font for values, i.e. the second column of each structure.}
\item \texttt{types} = \langle font settings \rangle \quad \text{(initially \texttt{itshape})}
  \text{The font used in \texttt{type} and \texttt{type*}.}
\item \texttt{tags} = \langle format settings \rangle \quad \text{(initially \texttt{footnotesize})}
  \text{The font (size) used in \texttt{tag} and the shortcuts \texttt{1}..\texttt{9}.}
\item \texttt{switch} = \langle token \rangle \quad \text{(initially !)}
  \text{Define the escape token. Change this if you need to use “!” as a text glyph.}
\end{itemize}
2.2 Commands available in the scope of \avm

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 \langle\structure\rangle, but this may change in future versions.

A \structure is basically the content of a stylised \texttt{array}: The columns are separated by & and a new line is entered with \texttt{\\}.

\texttt{\la} expects your (sub-)structures to have \textit{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 \structure. In the current beta version, display issues may appear in some structures if none is given.

\begin{verbatim}
\avm{[ < ( \{ ... \} ) > ]}
\end{verbatim}

\begin{verbatim}
\avm{[ \{ ... \} \" < ( ... ), ( ... ) > ]}
\end{verbatim}

\begin{verbatim}
\avm{\{ ... ! (text) ! \}}
\end{verbatim}

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

\begin{verbatim}
\tag {\langle identifier\rangle}
\end{verbatim}

\begin{verbatim}
\1, \2, \3, \4, \5, \6, \7, \8, \9
\end{verbatim}

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

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

\begin{verbatim}
\avm{[ attr1 & \4\}
attr2 & \4[attr3 & val3]\}
attr4 & val4 ]}}
\end{verbatim}

\begin{verbatim}
\tag \{ attr1 \[ attr2 \{ attr3 val3 \}
\end{verbatim}

\begin{verbatim}
\tag \\{ attr4 val4 \}
\end{verbatim}

3
\type \type(*) \{(type)\} \\
\type* Will typeset the \(type\) in the \texttt{types} font (roman italics by default). The starred variant \\type* will span the complete (sub-)structure and \textit{can only be placed in the first column} of this structure. After the starred \texttt{\type*}, a \texttt{\\&} is recommended, but can be omitted in “normal” cases.

\avm{\{
\type*A type spanning a line
attr & {\texttt{\type}}
\}} \quad \begin{array}{|c|}
\hline
A type spanning a line \\
\texttt{ATTR} \quad \texttt{type} \\
\hline
\end{array}

\punk \punk \{(\texttt{attribute})\}{\{(type)\}}

Some \texttt{(attributes)} 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 \texttt{(attribute)} and \texttt{(type)}. After \\punk, a \\& is recommended, but can be omitted in “normal” cases.

\avm{\{
\punk{attr1 & val1} \\
\punk{attr2 & val2} \}
attr3 & val3\
attr4 & val4\}

\avm{\{
attr1 & val1\
attr2 & val2 \\
attr3 & val3\}} \quad \begin{array}{|c|}
\hline
\texttt{ATTR1} \quad \texttt{val1} \\
\texttt{A QUITE LONG} \quad \texttt{ATTR2} \quad \texttt{val2} \\
\texttt{ATTR3} \quad \texttt{val3} \\
\texttt{ATTR4} \quad \texttt{val4} \\
\hline
\end{array}

- \quad In the scope of \texttt{\avm}, + comes out as “⊕”. To obtain “+”, please use the escape mode, as in !+. 

\textit{3 Applications}

\textit{3.1 Disjunctions and other relations}

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

\avm{\{
attr1 \& val1\} \\
attr2 \& val2\}
\texttt{\$\\lor\$} \\
\avm{\{
attr1' \& val1'\} \\
attr2' \& val2'\} \\
attr3' \& val3'\}} \\
\texttt{\textit{sign}} \quad \texttt{\$\to\$} \\
\avm{\{
attribute1 \& value1\} \\
attribute2 \& value2\} \\
attribute3 \& value3\} \\
\texttt{sign} \quad \texttt{\rightarrow} \\
\begin{array}{|c|}
\hline
\texttt{ATTRIBUTE1} \quad \texttt{value1} \\
\texttt{ATTRIBUTE2} \quad \texttt{value2} \\
\texttt{ATTRIBUTE3} \quad \texttt{value3} \\
\hline
\end{array}
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.

\[ \texttt{\textbackslash avm}[\texttt{attributes=\texttt{normalfont}}]{\texttt{\{v1\textbackslash v2\textbackslash v3\}\$\texttt{\$varphi\$}} \]

\[
\begin{bmatrix}
v1 \\
v2 \\
v3 \\
\end{bmatrix} \varphi
\]

3.3 Combinations with 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 \texttt{\textbackslash attop} from gb4e:

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

(1) \[
\begin{bmatrix}
\texttt{\textbackslash attr1 \textbackslash val1} \\
\texttt{\textbackslash attr2 \textbackslash val2} \\
\texttt{\textbackslash attr3 \textbackslash val3} \\
\end{bmatrix}
\]

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

\begin{exe}
\texttt{\textbackslash ex \textbackslash vtop\{\textbackslash strut\textbackslash vskip\textbackslash baselineskip\}
\texttt{\textbackslash avm\{}\texttt{[ attr1 & val1\textbackslash attr2 & val2\textbackslash attr3 & val3]\}}
\texttt{\textbackslash \}}
\end{exe}

\texttt{\textbackslash xe}

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 Combinations with forest

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

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

A

\[
\begin{bmatrix}
\texttt{\textbackslash attr1 \textbackslash val1} \\
\texttt{\textbackslash attr2 \textbackslash val2} \\
\texttt{\textbackslash attr3 \textbackslash val3} \\
\end{bmatrix}
\]

3.5 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, the “active mode” has disappeared, there is now a single way of sorting (see \texttt{\textbackslash type}), and tags are now produced without \texttt{\textbackslash@} (\texttt{\4} instead of \texttt{\4}, etc.). Please refer to Section 4 for features known from avm that are not yet available in langsci-avm.
3.6 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.

\begin{verbatim}
\usepackage{fontspec}
\usepackage{libertinus}
\avm[attributes=\addfontfeatures{
   Numbers=OldStyle,
   Letters=SmallCaps
}]
\end{verbatim}

3.7 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

\begin{verbatim}
\multicolumn{2}{l}{...}
\end{verbatim}

but only in the first column of a (sub-)structure. For a special usage case, see \texttt{\type} and \texttt{\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.

3. A check whether the delimiters are balanced, i.e. whether all (sub-)structures are closed by a \texttt{]}, \texttt{)}, 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 \texttt{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

\begin{verbatim}
\PackageRequirements{\avm}
\ProvidesExplPackage{langsci-avm}{2020-03-11}{0.1.0-beta}{AVMs and feature structures in LaTeX3}
\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 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{verbatim}

\begin{verbatim}
\NewDocumentCommand{\avm}{O{} +m} {
    \c_group_begin_token
    \__avm_initialise_document_commands:
    \bool_set_true:N \l__avm_mode_bool
    \keys_set:nn { avm } { #1 }
    \__avm_wrap:n { #2 }
    \c_group_end_token}
\end{verbatim}

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

\begin{verbatim}
\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 2.1.
\end{verbatim}

\begin{verbatim}
\NewDocumentCommand{\avmsetup}{m} {
    \keys_set:nn { avm } { #1 }
}
\end{verbatim}

\begin{verbatim}
\keys_define:nn { avm }
{
    stretch .code:n = \{\def\arraystretch{#1}\},
    stretch .initial:n = {1.15},
    columnsep .dim_set:N = \arraycolsep,
    columnsep .initial:n = {.5ex},
    delimfactor .int_set:N = \delimiterfactor,
    delimfactor .initial:n = {1000},
    delimfall .dim_set:N = \delimitershortfall,
    delimfall .initial:n = {Opt},
}\end{verbatim}

7
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.

\bool_new:N \l__avm_math_bool
\bool_new:N \l__avm_mode_bool
\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.

`\cs_generate_variant:Nn \seq_set_split:Nnn { NVn }`

(End definition for `\seq_set_split:NVn`.)

The replacement instructions for `\__avm_parse:n`

\begin{array}{>{\c_math_toggle_token}\normalfont\__avm_font_attribute:}l
\left\lbrace\__avm_module_begin:
\end{array}

\begin{array}{>{\c_math_toggle_token}\normalfont\__avm_font_value:}l
\right\rbrace\c_math_toggle_token

\__avm_replace_lbrace:
\__avm_replace_rbrace:
\__avm_replace_lbrack:
\{  
  \__avm_parse_output:nw  
  \{ \c_math_toggle_token\left\lbrack\__avm_module_begin: \}
  \}  
\cs_new:Nn \__avm_replace_rbrack:  
\{  
  \__avm_parse_output:nw  
  \{ \__avm_module_end:\right rbrack\c_math_toggle_token \}  
\}  
\cs_new:Nn \__avm_replace_lparen:  
\{  
  \__avm_parse_output:nw  
  \{ \c_math_toggle_token\left(\__avm_module_begin: \}
  \}  
\cs_new:Nn \__avm_replace_rparen:  
\{  
  \__avm_parse_output:nw  
  \{ \__avm_module_end:\right)\c_math_toggle_token \}  
\}  
\cs_new:Nn \__avm_replace_langle:  
\{  
  \__avm_parse_output:nw  
  \{ \c_math_toggle_token\left<\__avm_module_begin: \}
  \}  
\cs_new:Nn \__avm_replace_rangle:  
\{  
  \__avm_parse_output:nw  
  \{ \__avm_module_end:\right>\c_math_toggle_token \}  
\}  
\cs_new:Nn \__avm_replace_plus:  
\{  
  \__avm_parse_output:nw \{ \ensuremath { \oplus } \}  
\}  
(End definition for \__avm_module_begin:, \__avm_module_end:, and etc.)

\tag \type \punk
\cs_new:Nn \__avm_initialise_document_commands:
\{ 
  \cs_if_exist:NTF \tag
  \{ \RenewDocumentCommand{\tag}{m}{ \__avm_controls_tag:n {##1} } \}
  \{ \NewDocumentCommand{\tag}{m}{ \__avm_controls_tag:n {##1} } \}
  \cs_if_exist:NTF \1
  \{ \RenewDocumentCommand{\1}{m}{ \__avm_controls_tag:n {1} } \}
  \{ \NewDocumentCommand{\1}{m}{ \__avm_controls_tag:n {1} } \}
  \cs_if_exist:NTF \2
  \{ \RenewDocumentCommand{\2}{m}{ \__avm_controls_tag:n {2} } \}
  \{ \NewDocumentCommand{\2}{m}{ \__avm_controls_tag:n {2} } \}
  \cs_if_exist:NTF \3
  \{ \RenewDocumentCommand{\3}{m}{ \__avm_controls_tag:n {3} } \}
  \{ \NewDocumentCommand{\3}{m}{ \__avm_controls_tag:n {3} } \}
  \cs_if_exist:NTF \4
  \{ \RenewDocumentCommand{\4}{m}{ \__avm_controls_tag:n {4} } \}
  \{ \NewDocumentCommand{\4}{m}{ \__avm_controls_tag:n {4} } \}
  \cs_if_exist:NTF \5
  \{ \RenewDocumentCommand{\5}{m}{ \__avm_controls_tag:n {5} } \}
  \{ \NewDocumentCommand{\5}{m}{ \__avm_controls_tag:n {5} } \}
  \cs_if_exist:NTF \6
  \{ \RenewDocumentCommand{\6}{m}{ \__avm_controls_tag:n {6} } \}
  \{ \NewDocumentCommand{\6}{m}{ \__avm_controls_tag:n {6} } \}
  \cs_if_exist:NTF \7
  \{ \RenewDocumentCommand{\7}{m}{ \__avm_controls_tag:n {7} } \}
  \{ \NewDocumentCommand{\7}{m}{ \__avm_controls_tag:n {7} } \}
  \cs_if_exist:NTF \8
  \{ \RenewDocumentCommand{\8}{m}{ \__avm_controls_tag:n {8} } \}
  \{ \NewDocumentCommand{\8}{m}{ \__avm_controls_tag:n {8} } \}
  \cs_if_exist:NTF \9
  \{ \RenewDocumentCommand{\9}{m}{ \__avm_controls_tag:n {9} } \}
  \{ \NewDocumentCommand{\9}{m}{ \__avm_controls_tag:n {9} } \}
  \cs_if_exist:NTF \type
  \{ \RenewDocumentCommand{\type}{s m}{ \__avm_controls_type:nn {##1}{##2} } \}
  \{ \NewDocumentCommand{\type}{s m}{ \__avm_controls_type:nn {##1}{##2} } \}
  \{ \IfBooleanTF { ##1 } 
  \{ \__avm_controls_type_starred:n {##2} \}
  \{ \__avm_controls_type:n {##2} \}
  \}
  \{ \NewDocumentCommand{\type}{s m}{ \__avm_controls_type:nn {##1}{##2} } \}
  \{ \IfBooleanTF { ##1 } 
  \{ \__avm_controls_type_starred:n {##2} \}
  \{ \__avm_controls_type:n {##2} \}
  \}
  \cs_if_exist:NTF \punk
  \{ \RenewDocumentCommand{\punk}{m m}{ \__avm_controls_punk:nn {##1}{##2} } \}
  \{ \NewDocumentCommand{\punk}{m m}{ \__avm_controls_punk:nn {##1}{##2} } \}
  \{ \__avm_controls_punk:nn {##1}{##2} \}
\}
\)

(End definition for \tag, \type, and \punk. These functions are documented on page 3.)
The wrapper that first splits the input to \avm at each occurrence of \__avm_mode_switch_character and then inverts \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.

\cs_new_protected:Npn \__avm_wrap:n #1
  \seq_set_split:NVn \l__avm_wrapper_seq \__avm_mode_switch_character { #1 }
  \seq_map_inline:Nn \l__avm_wrapper_seq
    \exp_args:No \exp_not:o
    \__avm_parse:n {##1}
    \bool_set_inverse:N \l__avm_mode_bool
\end{definition}

Finally, the parser. It is build on \__avm_act:NNNnn from \texttt{l3tl} (see the sub-section \texttt{Token by token changes}). Many thanks to Phelype Oleinik for help on this, and in particular on help with expansion.
\begin{verbatim}
{ \_avm_replace:N #1 }
{ \_avm_replace_none:N #1 }

\cs_new:Npn \_avm_replace_none:N #1
{
 \_avm_parse_output:nw {#1}
}

\cs_new:Npn \_avm_replace:N #1
{
 \str_case:nnF {#1}
{ + }{ \_avm_replace_plus: }
{ [ }{ \_avm_replace_lbrack: }
{ ] }{ \_avm_replace_rbrack: }
{ ( }{ \_avm_replace_lparen: }
{ ) }{ \_avm_replace_rparen: }
{ \{ }{ \_avm_replace_lbrace: }
{ \} }{ \_avm_replace_rbrace: }
{ < }{ \_avm_replace_langle: }
{ > }{ \_avm_replace_rangle: }
}{ \_avm_replace_none:N #1 }
}

\cs_new:Npn \_avm_replace_group:nw #1
{ \exp_args:NNo \exp_args:No \_avm_replace_group:n { \_avm_parse:n {#1} } }

\cs_new:Npn \_avm_parse:n { #1 }
{ \_avm_parse_output:nw { {#1} } }

\exp_last_unbraced:NNo
\cs_new:Npn \_avm_replace_space:w \c_space_tl
{ \_avm_parse_output:nw { ~ } }

\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{verbatim}

\textit{(End definition for \_avm_parse:n.)}

\textit{(package)}