# The xparse package Document command parser\*

The LATEX3 Project<sup>†</sup>

Released 2017/03/11

The xparse package provides a high-level interface for producing document-level commands. In that way, it is intended as a replacement for the LATEX  $2_{\mathcal{E}}$  \newcommand macro. However, xparse works so that the interface to a function (optional arguments, stars and mandatory arguments, for example) is separate from the internal implementation. xparse provides a normalised input for the internal form of a function, independent of the document-level argument arrangement.

At present, the functions in xparse which are regarded as "stable" are:

- \NewDocumentCommand
- \RenewDocumentCommand
- \ProvideDocumentCommand
- \DeclareDocumentCommand
- \NewDocumentEnvironment
- \RenewDocumentEnvironment
- \ProvideDocumentEnvironment
- \DeclareDocumentEnvironment
- \NewExpandableDocumentCommand
- \RenewExpandableDocumentCommand
- \ProvideExpandableDocumentCommand
- \DeclareExpandableDocumentCommand
- \IfNoValue(TF)
- \IfValue(TF)
- \IfBoolean(TF)

with the other functions currently regarded as "experimental". Please try all of the commands provided here, but be aware that the experimental ones may change or disappear.

<sup>\*</sup>This file describes v6988, last revised 2017/03/11.

<sup>&</sup>lt;sup>†</sup>E-mail: latex-team@latex-project.org

#### 0.1 Specifying arguments

Before introducing the functions used to create document commands, the method for specifying arguments with xparse will be illustrated. In order to allow each argument to be defined independently, xparse does not simply need to know the number of arguments for a function, but also the nature of each one. This is done by constructing an argument specification, which defines the number of arguments, the type of each argument and any additional information needed for xparse to read the user input and properly pass it through to internal functions.

The basic form of the argument specifier is a list of letters, where each letter defines a type of argument. As will be described below, some of the types need additional information, such as default values. The argument types can be divided into two, those which define arguments that are mandatory (potentially raising an error if not found) and those which define optional arguments. The mandatory types are:

- m A standard mandatory argument, which can either be a single token alone or multiple tokens surrounded by curly braces. Regardless of the input, the argument will be passed to the internal code surrounded by a brace pair. This is the xparse type specifier for a normal T<sub>F</sub>X argument.
- 1 An argument which reads everything up to the first open group token: in standard LATEX this is a left brace.
- r Reads a "required" delimited argument, where the delimiters are given as  $\langle token1 \rangle$  and  $\langle token2 \rangle$ : r $\langle token1 \rangle \langle token2 \rangle$ . If the opening  $\langle token \rangle$  is missing, the default marker -NoValue- will be inserted after a suitable error.
- R As for r, this is a "required" delimited argument but has a user-definable recovery  $\langle default \rangle$ , given as  $R\langle token1 \rangle \langle token2 \rangle \{\langle default \rangle\}$ .
- u Reads an argument "until"  $\langle tokens \rangle$  are encountered, where the desired  $\langle tokens \rangle$  are given as an argument to the specifier:  $u\{\langle tokens \rangle\}$ .
- v Reads an argument "verbatim", between the following character and its next occurrence, in a way similar to the argument of the LATEX  $2_{\varepsilon}$  command \verb. Thus a v-type argument is read between two matching tokens, which cannot be any of %, \, #, {,} or  $\square$ . The verbatim argument can also be enclosed between braces, { and }. A command with a verbatim argument will not work when it appears within an argument of another function.

The types which define optional arguments are:

- o A standard LATEX optional argument, surrounded with square brackets, which will supply the special -NoValue- marker if not given (as described later).
- d An optional argument which is delimited by  $\langle token1 \rangle$  and  $\langle token2 \rangle$ , which are given as arguments:  $d\langle token1 \rangle \langle token2 \rangle$ . As with o, if no value is given the special marker -NoValue- is returned.
- O As for o, but returns  $\langle default \rangle$  if no value is given. Should be given as  $O\{\langle default \rangle\}$ .
- D As for d, but returns  $\langle default \rangle$  if no value is given:  $D\langle token1 \rangle \langle token2 \rangle \{\langle default \rangle\}$ . Internally, the o, d and O types are short-cuts to an appropriated-constructed D type argument.

- s An optional star, which will result in a value \BooleanTrue if a star is present and \BooleanFalse otherwise (as described later).
- t An optional  $\langle token \rangle$ , which will result in a value \BooleanTrue if  $\langle token \rangle$  is present and \BooleanFalse otherwise. Given as  $t\langle token \rangle$ .
- g An optional argument given inside a pair of TEX group tokens (in standard I♣TEX, { ...}), which returns -NoValue- if not present.
- **G** As for **g** but returns  $\langle default \rangle$  if no value is given:  $G\{\langle default \rangle\}$ .
- e An optional set of *embellishments*, each of which requires a *value*. If a key is not present, -NoValue- is returned. The returned data is a token list comprising one braced entry per key, ordered as for the key list in the argument specification. Given as e{\langle tokens \rangle}. All \langle tokens \rangle must be distinct. This is an experimental type.
- E As for e but returns one or more  $\langle defaults \rangle$  if values are not given:  $E\{\langle tokens \rangle\}\{\langle defaults \rangle\}$ . See Section 0.6 for more details.

Using these specifiers, it is possible to create complex input syntax very easily. For example, given the argument definition 's o o m O{default}', the input '\*[Foo]{Bar}' would be parsed as:

- #1 = \BooleanTrue
- #2 = Foo
- #3 = -NoValue-
- #4 = Bar
- #5 = default

whereas '[One] [Two] {} [Three]' would be parsed as:

- #1 = \BooleanFalse
- #2 = One
- #3 = Two
- #4 =
- #5 = Three

Delimited argument types (d, o and r) are defined such that they require matched pairs of delimiters when collecting an argument. For example

Also note that  $\{$  and  $\}$  cannot be used as delimiters as they are used by  $T_EX$  as grouping tokens. Arguments to be grabbed inside these tokens must be created as either m- or g-type arguments.

Within delimited arguments, non-balanced or otherwise awkward tokens may be included by protecting the entire argument with a brace pair

These braces will be stripped if they surround the  $\it entire$  content of the optional argument

Two more tokens have a special meaning when creating an argument specifier. First, + is used to make an argument long (to accept paragraph tokens). In contrast to  $\LaTeX$  2 $_{\mathcal{E}}$ 's \newcommand, this applies on an argument-by-argument basis. So modifying the example to 's o o +m O{default}' means that the mandatory argument is now \long, whereas the optional arguments are not.

Secondly, the token > is used to declare so-called "argument processors", which can be used to modify the contents of an argument before it is passed to the macro definition. The use of argument processors is a somewhat advanced topic, (or at least a less commonly used feature) and is covered in Section 0.9.

When an optional argument is followed by a mandatory argument with the same delimiter, xparse issues a warning because the optional argument could not be omitted by the user, thus becoming in effect mandatory. This applies to o, d, D, s, t, e, and E type arguments followed by r or R-type required arguments, but also to g or G type arguments followed by m type arguments.

## 0.2 Spacing and optional arguments

TEX will find the first argument after a function name irrespective of any intervening spaces. This is true for both mandatory and optional arguments. So \foo[arg] and \foo\_\lumbda [arg] are equivalent. Spaces are also ignored when collecting arguments up to the last mandatory argument to be collected (as it must exist). So after

```
\NewDocumentCommand \foo { m o m } { ... }
```

the user input  $\lceil arg1 \rceil \lceil arg2 \rceil \rceil$  and  $\lceil arg1 \rceil \rfloor \lceil arg2 \rceil \rceil$  will both be parsed in the same way. However, spaces are *not* ignored when parsing optional arguments after the last mandatory argument. Thus with

```
\NewDocumentCommand \foobar { m o } { ... }
```

\foobar{arg1}[arg2] will find an optional argument but \foobar{arg1}\_\[arg2] will not. This is so that trailing optional arguments are not picked up "by accident" in input.

There is one major exception to the rules listed above: when xparse is used to define what TEX defines as "control symbols" in which the function name is made up of a single character, such as "\\", spaces are *not* ignored directly after them even for mandatory arguments.

#### 0.3 Required delimited arguments

The contrast between a delimited (D-type) and "required delimited" (R-type) argument is that an error will be raised if the latter is missing. Thus for example

```
\NewDocumentCommand {\foobaz} {r()m} {}
\foobaz{oops}
```

will lead to an error message being issued. The marker <code>-NoValue-(r-type)</code> or user-specified default (for R-type) will be inserted to allow error recovery.

Users should note that support for required delimited arguments is somewhat experimental. Feedback is therefore very welcome on the LaTeX-L mailing list.

#### 0.4 Verbatim arguments

Arguments of type v are read in verbatim mode, which will result in the grabbed argument consisting of tokens of category codes 12 ("other") and 13 ("active"), except spaces, which are given category code 10 ("space"). The argument is delimited in a similar manner to the LATEX  $2\varepsilon$  \verb function, or by (correctly nested) pairs of braces.

Functions containing verbatim arguments cannot appear in the arguments of other functions. The v argument specifier includes code to check this, and will raise an error if the grabbed argument has already been tokenized by TEX in an irreversible way.

By default, an argument of type v must be at most one line. Prefixing with + allows line breaks within the argument.

Users should note that support for verbatim arguments is somewhat experimental. Feedback is therefore very welcome on the LaTeX-L mailing list.

## 0.5 Default values of arguments

Uppercase argument types (0, D, ...) allow to specify a default value to be used when the argument is missing; their lower-case counterparts use the special marker <code>-NoValue-</code>. The default value can be expressed in terms of the value of any other arguments by using #1, #2, and so on.

The default values may refer to arguments that appear later in the argument specification. For instance a command could accept two optional arguments, equal by default:

Users should note that support for default arguments referring to other arguments is somewhat experimental. Feedback is therefore very welcome on the LaTeX-L mailing list.

## 0.6 Default values for "embellishments"

The E-type argument allows one default value per test token. This is achieved by giving a list of defaults for each entry in the list, for example:

```
E{^_}{{UP}{DOWN}}
```

If the list of default values is *shorter* than the list of test tokens, the special \NoValue marker will be returned (as for the e-type argument). Thus for example

```
E{^_}{{UP}}
```

has default UP for the ^ test token, but will return the \NoValue marker as a default for \_. This allows mixing of explicit defaults with testing for missing values.

## 0.7 Declaring commands and environments

With the concept of an argument specifier defined, it is now possible to describe the methods available for creating both functions and environments using xparse.

The interface-building commands are the preferred method for creating document-level functions in  $\LaTeX$  3. All of the functions generated in this way are naturally robust (using the  $\varepsilon$ -T<sub>F</sub>X \protected mechanism).

\NewDocumentCommand \RenewDocumentCommand \ProvideDocumentCommand \DeclareDocumentCommand

```
\verb|\NewDocumentCommand| \langle Function \rangle \ \{\langle arg \ spec \rangle\} \ \{\langle code \rangle\}|
```

This family of commands are used to create a document-level  $\langle function \rangle$ . The argument specification for the function is given by  $\langle arg \ spec \rangle$ , and expanding to be replaced by the  $\langle code \rangle$ .

As an example:

```
NewDocumentCommand \chapter { s o m }
{
   \IfBooleanTF {#1}
      { \typesetstarchapter {#3} }
      { \typesetnormalchapter {#2} {#3} }
}
```

would be a way to define a \chapter command which would essentially behave like the current LATEX  $2_{\varepsilon}$  command (except that it would accept an optional argument even when a \* was parsed). The \typesetnormalchapter could test its first argument for being -NoValue- to see if an optional argument was present.

The difference between the  $\New...$   $\Renew...$ ,  $\Provide...$  and  $\Declare...$  versions is the behaviour if  $\langle function \rangle$  is already defined.

- \NewDocumentCommand will issue an error if  $\langle function \rangle$  has already been defined.
- \RenewDocumentCommand will issue an error if  $\langle function \rangle$  has not previously been defined.
- \ProvideDocumentCommand creates a new definition for  $\langle function \rangle$  only if one has not already been given.
- \DeclareDocumentCommand will always create the new definition, irrespective of any existing \( \langle function \rangle \) with the same name. This should be used sparingly.

**TEX**hackers note: Unlike  $\text{LATEX } 2\varepsilon$ 's \newcommand and relatives, the \NewDocumentCommand family of functions do not prevent creation of functions with names starting \end....

\NewDocumentEnvironment \RenewDocumentEnvironment \ProvideDocumentEnvironment \DeclareDocumentEnvironment

These commands work in the same way as  $\ensuremath{\mbox{NewDocumentCommand}}$ , etc., but create environments ( $\ensuremath{\mbox{\mbox{begin}}{\langle environment\rangle}}$ ). Both the  $\ensuremath{\langle start\ code\rangle}$  and  $\ensuremath{\langle end\ code\rangle}$  may access the arguments as defined by  $\ensuremath{\langle arg\ spec\rangle}$ . The arguments will be given following  $\ensuremath{\mbox{\mbox{begin}}{\langle environment\rangle}}$ .

#### 0.8 Testing special values

Optional arguments created using xparse make use of dedicated variables to return information about the nature of the argument received.

```
\IfNoValueT
\IfNoValueF
\IfNoValueTF
```

```
\IfNoValueTF {\langle argument \rangle} {\langle true\ code \rangle} \IfNoValueT {\langle argument \rangle} {\langle true\ code \rangle} \IfNoValueF {\langle argument \rangle} {\langle false\ code \rangle}
```

The \IfNoValue(TF) tests are used to check if  $\langle argument \rangle$  (#1, #2, etc.) is the special -NoValue- marker For example

```
\NewDocumentCommand \foo { o m }
   {
     \IfNoValueTF {#1}
      { \DoSomethingJustWithMandatoryArgument {#2} }
      { \DoSomethingWithBothArguments {#1} {#2} }
}
```

will use a different internal function if the optional argument is given than if it is not present.

Note that three tests are available, depending on which outcome branches are required: \IfNoValueTF, \IfNoValueT and \IfNoValueF.

As the \IfNoValue(TF) tests are expandable, it is possible to test these values later, for example at the point of typesetting or in an expansion context.

It is important to note that -NoValue- is constructed such that it will not match the simple text input -NoValue-, i.e. that

```
\IfNoValueTF{-NoValue-}
```

will be logically false.

When two optional arguments follow each other (a syntax we typically discourage), it can make sense to allow users of the command to specify only the second argument by providing an empty first argument. Rather than testing separately for emptyness and for <code>-NoValue-</code> it is then best to use the argument type <code>O</code> with an empty default value, and simply test for emptyness using the <code>expl3</code> conditional <code>\tl\_if\_blank:nTF</code> or its <code>etoolbox</code> analogue <code>\ifblank</code>.

\IfValueT \
\IfValueF \
\IfValueTF \

The reverse form of the \IfNoValue(TF) tests are also available as \IfValue(TF). The context will determine which logical form makes the most sense for a given code scenario.

\BooleanFalse \BooleanTrue The true and false flags set when searching for an optional token (using s or  $t\langle token \rangle$ ) have names which are accessible outside of code blocks.

```
\IfBooleanT *
\IfBooleanF *
\IfBooleanTF *
```

checks for a star as the first argument, then chooses the action to take based on this information.

## 0.9 Argument processors

}

xparse introduces the idea of an argument processor, which is applied to an argument after it has been grabbed by the underlying system but before it is passed to  $\langle code \rangle$ . An argument processor can therefore be used to regularise input at an early stage, allowing the internal functions to be completely independent of input form. Processors are applied to user input and to default values for optional arguments, but not to the special  $\NoValue$  marker.

Each argument processor is specified by the syntax  $\{processor\}$  in the argument specification. Processors are applied from right to left, so that

```
>{\ProcessorB} >{\ProcessorA} m
```

would apply \ProcessorA followed by \ProcessorB to the tokens grabbed by the m argument.

#### \ProcessedArgument

xparse defines a very small set of processor functions. In the main, it is anticipated that code writers will want to create their own processors. These need to accept one argument, which is the tokens as grabbed (or as returned by a previous processor function). Processor functions should return the processed argument as the variable \ProcessedArgument.

### \ReverseBoolean

#### \ReverseBoolean

This processor reverses the logic of \BooleanTrue and \BooleanFalse, so that the example from earlier would become

```
\NewDocumentCommand \foo { > { \ReverseBoolean } s m }
{
  \IfBooleanTF #1
     { \DoSomethingWithoutStar {#2} }
     { \DoSomethingWithStar {#2} }
}
```

\SplitArgument

 $SplitArgument {\langle number \rangle} {\langle token \rangle}$ 

Updated: 2012-02-12

This processor splits the argument given at each occurrence of the  $\langle token \rangle$  up to a maximum of  $\langle number \rangle$  tokens (thus dividing the input into  $\langle number \rangle + 1$  parts). An error is given if too many  $\langle tokens \rangle$  are present in the input. The processed input is placed inside  $\langle number \rangle + 1$  sets of braces for further use. If there are fewer than  $\{\langle number \rangle\}$  of  $\{\langle tokens \rangle\}$  in the argument then \NoValue markers are added at the end of the processed argument.

```
\NewDocumentCommand \foo
{ > { \SplitArgument { 2 } { ; } } m }
{ \InternalFunctionOfThreeArguments #1 }
```

Any category code 13 (active)  $\langle tokens \rangle$  will be replaced before the split takes place. Spaces are trimmed at each end of each item parsed.

\SplitList

```
SplitList {\langle token(s) \rangle}
```

This processor splits the argument given at each occurrence of the  $\langle token(s) \rangle$  where the number of items is not fixed. Each item is then wrapped in braces within #1. The result is that the processed argument can be further processed using a mapping function.

```
\NewDocumentCommand \foo
{ > { \SplitList { ; } } m }
{ \MappingFunction #1 }
```

If only a single  $\langle token \rangle$  is used for the split, any category code 13 (active)  $\langle token \rangle$  will be replaced before the split takes place.

\ProcessList  $\star$ 

```
\ProcessList \{\langle list \rangle\}\ \{\langle function \rangle\}
```

To support \SplitList, the function \ProcessList is available to apply a  $\langle function \rangle$  to every entry in a  $\langle list \rangle$ . The  $\langle function \rangle$  should absorb one argument: the list entry. For example

```
\NewDocumentCommand \foo
{ > { \SplitList { ; } } m }
{ \ProcessList {#1} { \SomeDocumentFunction } }
```

This function is experimental.

#### \TrimSpaces

#### \TrimSpaces

Removes any leading and trailing spaces (tokens with character code 32 and category code 10) for the ends of the argument. Thus for example declaring a function

\foo{ hello world }

will show hello world at the terminal, with the space at each end removed. \TrimSpaces will remove multiple spaces from the ends of the input in cases where these have been included such that the standard TeX conversion of multiple spaces to a single space does not apply.

This function is experimental.

### 0.10 Fully-expandable document commands

There are *very rare* occasion when it may be useful to create functions using a fully-expandable argument grabber. To support this, xparse can create expandable functions as well as the usual robust ones. This imposes a number of restrictions on the nature of the arguments accepted by a function, and the code it implements. This facility should only be used when *absolutely necessary*; if you do not understand when this might be, *do not use these functions*!

\NewExpandableDocumentCommand \RenewExpandableDocumentCommand \ProvideExpandableDocumentCommand \DeclareExpandableDocumentCommand This family of commands is used to create a document-level  $\langle function \rangle$ , which will grab its arguments in a fully-expandable manner. The argument specification for the function is given by  $\langle arg\ spec \rangle$ , and the function will execute  $\langle code \rangle$ . In general,  $\langle code \rangle$  will also be fully expandable, although it is possible that this will not be the case (for example, a function for use in a table might expand so that **\omit** is the first non-expandable non-space token).

Parsing arguments expandably imposes a number of restrictions on both the type of arguments that can be read and the error checking available:

- The last argument (if any are present) must be one of the mandatory types m or r.
- All short arguments appear before long arguments.
- The mandatory argument types 1 and u may not be used after optional arguments.
- The optional argument types g and G are not available.
- The "verbatim" argument type v is not available.
- Argument processors (using >) are not available.
- It is not possible to differentiate between, for example \foo[ and \foo{[}: in both cases the [ will be interpreted as the start of an optional argument. As a result, checking for optional arguments is less robust than in the standard version.

xparse will issue an error if an argument specifier is given which does not conform to the first six requirements. The last item is an issue when the function is used, and so is beyond the scope of xparse itself.

## 0.11 Access to the argument specification

The argument specifications for document commands and environments are available for examination and use.

\GetDocumentCommandArgSpec \GetDocumentEnvironmentArgSpec  $\label{lem:commandArgSpec} $$\operatorname{CommandArgSpec} \ \langle \operatorname{function} \rangle $$ \operatorname{CotDocumentEnvironmentArgSpec} \ \langle \operatorname{environment} \rangle $$$ 

These functions transfer the current argument specification for the requested  $\langle function \rangle$  or  $\langle environment \rangle$  into the token list variable \ArgumentSpecification. If the  $\langle function \rangle$  or  $\langle environment \rangle$  has no known argument specification then an error is issued. The assignment to \ArgumentSpecification is local to the current  $T_{EX}$  group.

\ShowDocumentCommandArgSpec \ShowDocumentEnvironmentArgSpec

These functions show the current argument specification for the requested  $\langle function \rangle$  or  $\langle environment \rangle$  at the terminal. If the  $\langle function \rangle$  or  $\langle environment \rangle$  has no known argument specification then an error is issued.

# 1 Load-time options

log-declarations

The package recognises the load-time option log-declarations, which is a key-value option taking the value true and false. By default, the option is set to true, meaning that each command or environment declared is logged. By loading xparse using

\usepackage[log-declarations=false]{xparse}

this may be suppressed and no information messages are produced.

# Index

The italic numbers denote the pages where the corresponding entry is described, numbers underlined point to the definition, all others indicate the places where it is used.

Symbols	\long 4
\\ 4	N
A	\New 6
\ArgumentSpecification 11, 11	\newcommand 1, 4, 6
(	NewDocumentCommand $\dots 1, 6, 6, 6, 6$
В	\NewDocumentEnvironment 1, 6, 6
\begin 6, 6	\NewExpandableDocumentCommand 1, 11, 11
\BooleanFalse 3, 3, 7, 8, 8	\NoValue 5, 5, 8, 9
\BooleanTrue 3, 3, 7, 8, 8	
	О
$\mathbf{C}$	\omit 11
\chapter 6	options:
Th.	log-declarations
D	P
\Declare	<del>-</del>
\DeclareDocumentCommand 1, 6, 6	\ProcessedArgument
\DeclareDocumentEnvironment 1, 6	\ProcessorA 8
\DeclareExpandableDocumentCommand 1, 11	\ProcessorB 8
${f E}$	\protected 6
\end 6	\Provide
\end 6	\ProvideDocumentCommand 1, 6, 6
(022011)	\ProvideDocumentEnvironment 1, 6
${f G}$	\ProvideExpandableDocumentCommand 1, 11
\GetDocumentCommandArgSpec 11, 11	,
\GetDocumentEnvironmentArgSpec . 11, 11	${f R}$
	\Renew 6
I	\RenewDocumentCommand 1, 6, 6
\IfBoolean(TF) 1	\RenewDocumentEnvironment 1, 6
\IfBooleanTF 8, 8	\RenewExpandableDocumentCommand 1, 11
\IfNoValue(TF) 1, 7, 7, 7	\ReverseBoolean
\IfNoValueTF 7, 7, 7, 7, 7, 7, 7	a
\IfValue(TF)	S 11 11
\IfValueTF	\ShowDocumentCommandArgSpec 11, 11
т	\ShowDocumentEnvironmentArgSpec . 11, 11
L	\SplitArgument
log-declarations (option) 12	\SplitList 9, 9, 9

${f T}$	\TrimSpaces 10, 10, 10
$T_{EX}$ and $\LaTeX 2_{\varepsilon}$ commands:	\typesetnormalchapter 6
\ifblank 7	-
tl commands:	$\mathbf{V}$
\tl_if_blank:nTF 7	\verb 2, 5