

Painless Digit Separation

Introduction

N2281 ("Digit Separators", by Lawrence Crowl) and N3342 ("Digit Separators coming back", by Jens Maurer) propose allowing an underscore as an optional digit separator in numeric literals. N3342 further proposes addressing a conflict with user-defined literal syntax by disallowing user-defined literals that start with an underscore followed by a digit. However, an ambiguity remains with hexadecimal literals. Consider:

```
0xdead_beef_db
```

Is "`_db`" a suffix indicating a user-defined literal or two additional hexadecimal digits? What about "`_beef_db`"?

To avoid the problem altogether, this paper proposes an alternative separator glyph: The single quote. It avoids the ambiguities mentioned above, but it is also nearly identical to the "upper comma" that is somewhat commonly used as a digits separator in print and handwriting (particularly for currency values, but also in other contexts). The example above could for example be unambiguously written:

```
0xdead'beef_db
```

The motivating examples of N3342 become:

```
"pronounce 7'237'498'123"
```

```
"compare 237'498'123 with 237'499'123 for equality"
```

```
"decide whether 237'499'123 or 20'249'472 is larger"
```

Proposed Wording Changes

Amend the token-grammar of 2.10 [lex.ppnumber] as indicated:

```
pp-number :
    digit
    . digit
    pp-number digit
    pp-number ' digit
    pp-number identifier-nondigit
    pp-number e sign
    pp-number E sign
    pp-number .
```

Amend the token-grammar of 2.14.2 [lex.icon] as indicated:

```

decimal-literal:
    nonzero-digit
    decimal-literal 'opt' digit

octal-literal:
    0
    octal-literal 'opt' octal-digit

hexadecimal-literal:
    0x hexadecimal-digit
    0X hexadecimal-digit
    hexadecimal-literal 'opt' hexadecimal-digit

```

Amend 2.14.2 [lex.icon] paragraph 1 as indicated:

- 1 An *integer literal* is a sequence of digits that has no period or exponent part, with optional separating single quotes that are ignored when determining its value. ...
 [*Example*: The number twelve can be written 12, 014, or 0XC. The literals 1048576, 1'048'576, 0X100000, 0x10'0000, and 0'004'000'000 all have the same value. — *end example*]

Amend the token-grammar of 2.14.4 [lex.fcon] as indicated:

```

digit-sequence:
    digit
    digit-sequence 'opt' digit

```

Amend 2.14.4 [lex.fcon] paragraph 1 as indicated:

- 1 ... The integer and fraction parts both consist of a sequence of decimal (base ten) digits, with optional separating single quotes that are ignored when determining their value. [*Example*: The literals 1.602'176'565e-19 and 1.602176565e-19 have the same value. — *end example*] ...

Acknowledgments

The wording section was heavily inspired by N3342. Thanks to Mike Miller for helpful review comments on an early draft.