Choosing correct grammatical form & capitalization

Dear language/locale experts: For some languages, and in certain contexts, it appears that current CLDR/LDML structure does not provide enough flexibility to handle the variety of grammatical forms required for certain terms such as month names, and may also not provide enough information to correctly address capitalization/lettercasing requirements. For example, to translate the date forms in the following, some languages may require more than 2 different grammatical forms for “October” (and the various forms may have different capitalization requirements as well):

1. “Today is Thursday, October 22, 2009”
2. “October 2009” (in a calendar user-interface widget, to select which calendar page to show)
3. “October” (by itself as the title of a calendar page)
4. “October” (an item in a popup menu)

LDML currently provides the following two styles per <http://www.unicode.org/reports/tr35/>:

• format (the default), the form used within a date format string [as in #1 above]
• stand-alone, the form used independently, such as in calendar headers [as in #3 above, or for day names, as in the heading of a column on a calendar page]

Finnish, for example, may require three different forms for wide month name, as in the following examples: “maaliskuu”, “9. maaliskuuta”, “Heisei 21, maaliskuun 9.” (for Japanese calendar)

In some cases the combination of grammatical form and appropriate capitalization or punctuation may also require more than 2 forms. In Russian, for example, different contexts require 3 different forms for month or day name:

• stand-alone “Декабрь”, perhaps a new format-alone form “декабрь”, format form “декабря” (where grammatical changes may be required, e.g. for genitive)
• stand-alone “Среда”, perhaps a new format-alone form “среда”, format form “среду”

Catalan uses format-style months with different prepositions, but stand-alone months without such prepositions: format “15 d’abril” and “15 de maig”, versus stand-alone “abril” and “maig”.

cldrbug 2265 is for these issues: <http://unicode.org/cldr/trac/ticket/2265>.

Other capitalization issues

LDML has data for localized display names of scripts, languages, territories, currencies, calendars, collation orders, and measurement system names. However, appropriate capitalization of these has been an ongoing issue and source of confusion. In some cases there may also be issues of appropriate grammatical form.

As described in <http://www.unicode.org/reports/tr35/>, these display names are intended “…for use in user-interfaces for displaying menu lists. In addition, the localized names for currency items may also be suitable for use in user-interfaces involving flowing text.”

The guidance in the Survey Tool for language capitalization also states “A common question is whether to capitalize or not. With a new locale, use whatever is normal practice for what should occur in menus. …This is the practice for scripts, territories and other types of items too.”

Beyond that, LDML also specifies the inList and inText elements that can be used for further control of capitalization in various circumstances:
• inList “controls whether display names (language, territory, etc) are title cased in GUI menu
lists and the like. It is only used in languages where the normal display is lower case, but title
case is used in lists.” Possible values include “titlecase-firstword”, “titlecase-words”, and
“mixed”. Note, since the basic names of the display items are already supposed to be cased
as in GUI menus per the instructions above, it is not clear what additional value this element
provides.
• inText “indicates the casing of the data in the category identified by the inText type attribute,
when that data is written in text or how it would appear in a dictionary”; types to which this
can separately apply include languages, territories, scripts, measurementSystemNames,
currency, etc. Possible values include “lowercase-words”, “titlecase-firstword”, “titlecase-
words”, and “mixed”.

It has also been suggested that perhaps instead, these items should be capitalized as they
would be in a dictionary, which may also be how they would appear in the middle of text (not at
the beginning of a sentence). In that case the inList item would indicate how to display them in
a menu.

What we would like you to do
1. For the examples in section A, we need to understand when the current “format” and
“stand-alone” forms are insufficient to cover the language requirements (both in terms of
grammatical form and capitalization). Please look over the examples and let us know
whether there are some that cannot be accommodated by the current LDML capabilities.
Alternatively, you can provide a translation of each of these examples, with some guidance
about why a particular grammatical form or capitalization was used in each case. Note: The
eamples in A11 are particularly difficult, and there may not be a way of making these work
well in certain languages.

2. For the examples in sections B, C, and D, we primarily need to understand what
capitalization is required in each case. For some languages different grammatical forms of
a display name may be required for different display contexts; please let us know.

3. Regarding the examples in section E, CLDR already has a mechanism to match the noun
plural form to the number, and to provide for capitalization of the unit name in this particualr
context. Please let us know whether there are any other issues with unit representation that
cannot be handled by the current mechanisms.

4. Finally, we would like feedback on this document and these examples. Please let us know if
the instructions are not clear, the examples are inadequate or not well-chosen, etc. (And if
you have additional examples in the form of text or a screenshot, please provide them!)

Note that many of these examples are specific to, or at least oriented to, Mac OS X, since
those were what I had available. If these do not cover cases needed for other platforms,
please send examples.

Thanks for your help!

Peter Edberg (Apple Inc.), for the CLDR team
A. Date formats and date elements

A1. (EEE h:mm:ss a)?
A2. (EEEE, MMMM d, y)?
A3. (EEE MMM d h:mm:ss a)?
A4. (EEEE MMM d)?

A5.1. (cccc)?
A5.2. (cccc)?
A6.1. (MMMM y) or (LLLL y)?
A6.2. (MMMM y) or (LLLL y)?
A7. (cccccc)?
A11. Here, popup menus can be dragged around to create a date format, and literal characters can be typed in between as separators. The popups can be used to select different styles for day and month name, etc. This may difficult to localize, since for example elements are changed in isolation from one another. Any suggestions for appropriate formats for these elements?
**Date formats in text:**

<table>
<thead>
<tr>
<th>Date Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Today is Thursday, October 22, 2009.</td>
<td>A15. (EEEE, MMMM d, y)?</td>
</tr>
<tr>
<td>Today is October 22.</td>
<td>A16. (MMMM d)?</td>
</tr>
<tr>
<td>Today is Thursday.</td>
<td>A17. (EEEE)?</td>
</tr>
<tr>
<td>This month is October.</td>
<td>A18. (MMMM) or (LLLL)?</td>
</tr>
<tr>
<td>This month is October 2009.</td>
<td>A19. (MMMM y) or (LLLL y)?</td>
</tr>
</tbody>
</table>

**Date element lists (need to know correct grammatical form & capitalization):**

<table>
<thead>
<tr>
<th>Element List</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>The months are January, February, March, April, May, June, July…</td>
<td>A20. Month name list in text</td>
</tr>
<tr>
<td>The days of the week are are Monday, Tuesday, Wednesday, Thursday…</td>
<td>A21. Day name list in text</td>
</tr>
<tr>
<td>First day of week</td>
<td>A22. Day name list in menu</td>
</tr>
</tbody>
</table>
B. Lists/names of languages (need to know correct grammatical form & capitalization)

B1. User-interface language preferences 1; mixed languages (special case). This data is not from CLDR.

B2. User-interface language preferences 2; mixed languages (special case). This data is not from CLDR.

B3. Menu of languages/locales with special word break options
B4. Menu of languages for spell checking

B5. Menu of collation languages and keys

B6. Collation language/key preference
Germanic languages include English, German, Dutch, Norwegian, Danish, Swedish…

**C. Lists/names of regions, cities** (need to know correct grammatical form & capitalization)

The G8 countries include France, Germany, Italy, Japan, the United Kingdom…
D. Lists/names of other items (need to know correct grammatical form & capitalization)

D1. List of calendars in a menu.

D2. List of calendars in text

Gregorian-based calendars include Japanese, Buddhist, Republic of China...

D3. List of measurement units in menu.

D4. List of currencies in menu.

D5. List of currencies in text.

Asian currencies include Chinese Yuan Renminbi, Japanese Yen, South Korean Won, and Indian Rupee.

D6. Currencies in dialog.

(need example)
E. Lists/names of units (current structure can already handle matching of unit plural form to number; need to know correct capitalization and whether there are other issues)

E1. Currency units in text.

Tickets are available for 15 US dollars.

E2. Time units in text (hours).

The examination will take 3 hours.

E3. Time units in text (minutes).

The trip will take 30 minutes.

E4. Time units (days).

The offer will expire in 3 days.

E5. Length units (meters).

The minimum length of a soccer field is 100 meters.