



图 16-9 语言标记被分隔为子标记

## 16.4.5 大小写

所有的标记都是不区分大小写的，也就是说，标记 en 和 eN 是等价的。但是，习惯上用全小写来表示一般的语言，而用全大写来表示特定的国家。例如，fr 表示所有分类为法语的语言，而 FR 表示国家法国<sup>19</sup>。

## 16.4.6 IANA语言标记注册

第一个和第二个语言子标记的值由各种标准文档以及相关的维护组织定义。IANA<sup>20</sup>依据 RFC 3066 中概括的规则来管理标准的语言标记列表。

如果语言标记由标准的国家和语言值组成，标记就不需要专门注册。只有那些无法用标准的国家和语言值构成的语言标记才需要专门向 IANA 注册<sup>21</sup>。下面几节概括了 RFC 3066 中关于第一子标记和第二子标记的标准。

386

## 16.4.7 第一个子标记——名字空间

第一个子标记通常是标准化的语言记号，选自 ISO 639 中的语言标准集合。不过也可以用字母 i 来标识在 IANA 中注册的名字，或用 x 表示私有的或者扩展的名字。下面是各种规则。

如果第一个子标记含有：

- 2 个字符，那就是来自 ISO 639<sup>22</sup> 和 639-1 标准的语言代码；

注 19：这种惯用法是 ISO 标准 3166 中推荐的。

注 20：参见 <http://www.iana.org> 和 RFC 2860。

注 21：编写本书的时候，IANA 中只显式注册了 21 种语言标记，包括汉语中的广东话（标记是 zh-yue）、新挪威语（标记是 no-nyn）、卢森堡语（标记是 i-lux）以及克林根语（语言标记是 i-klingon）。在因特网上使用的其余数百种口语的标记均由标准的组件构成。

注 22：参见 ISO 标准 639，“Codes for the representation of names of languages—Part2: Alpha-3code”（表示语言名字的代码——第 2 部分：Alpha-3 代码）。