

Groupe de travail Réseau  
**Request for Comments : 5646**  
**BCP 47**  
RFC rendue obsolète : 4646  
Catégorie : Bonnes pratiques actuelles

A. Phillips, éditeur, Lab126  
M. Davis, éditeur, Google  
septembre 2009

Traduction Claude Brière de l'Isle

# Étiquettes d'identification des langages

## Résumé

Le présent document décrit la structure, le contenu, la construction, et la sémantique des étiquettes de langage à utiliser dans les cas où il est souhaitable d'indiquer le langage utilisé dans un objet d'information. Il décrit aussi comment enregistrer les valeurs à utiliser dans les étiquettes de langage et la création d'extensions définies par l'utilisateur pour un échange privé.

## Statut de ce mémoire

Ce document spécifie les bonnes pratiques actuelles sur l'Internet pour la communauté de l'Internet, et demande des discussions et suggestions pour son amélioration. La diffusion du présent mémoire n'est soumise à aucune restriction.

## Notice de droits de reproduction

Copyright (c) 2009 IETF Trust et les personnes identifiées comme auteurs du document. Tous droits réservés.

Le présent document est soumis au BCP 78 et aux dispositions légales de l'IETF Trust qui se rapportent aux documents de l'IETF (<http://trustee.ietf.org/license-info>) en vigueur à la date de publication de ce document. Prière de revoir ces documents avec attention, car ils décrivent vos droits et obligations par rapport à ce document. Les composants de code extraits du présent document doivent inclure le texte de licence simplifié de BSD comme décrit au paragraphe 4.e des dispositions légales du Trust et sont fournis sans garantie comme décrit dans la licence de BSD simplifiée.

Le présent document peut contenir des matériaux provenant de documents de l'IETF ou de contributions à l'IETF publiées ou rendues disponibles au public avant le 10 novembre 2008. La ou les personnes qui ont le contrôle des droits de reproduction sur tout ou partie de ces matériaux peuvent n'avoir pas accordé à l'IETF Trust le droit de permettre des modifications de ces matériaux en dehors du processus de normalisation de l'IETF. Sans l'obtention d'une licence adéquate de la part de la ou des personnes qui ont le contrôle des droits de reproduction de ces matériaux, le présent document ne peut pas être modifié en dehors du processus de normalisation de l'IETF, et des travaux dérivés ne peuvent pas être créés en dehors du processus de normalisation de l'IETF, excepté pour le formater en vue de sa publication comme RFC ou pour le traduire dans une autre langue que l'anglais.

## Table des matières

1. Introduction.....	2
2. L'étiquette de langage.....	2
2.1 Syntaxe.....	2
2.2. Sources et interprétation d'étiquette de langage.....	5
2.2.1 Sous étiquette principale de langage.....	5
3. Format et maintenance de registre.....	12
3.1 Format du registre IANA des sous étiquettes de langage.....	12
3.2 Réviseur de sous étiquette de langage.....	19
3.3 Maintenance du registre.....	19
3.4 Stabilité des entrées du registre de l'IANA.....	20
3.5 Procédure d'enregistrement des sous étiquettes.....	23
3.6. Possibilités d'enregistrement.....	25
3.7 Extensions et registre des extensions.....	27
3.8 Mise à jour du registre des sous étiquettes de langage.....	28
3.9 Applicabilité du registre des sous étiquettes de langage.....	28
4. Formation et traitement des étiquettes de langage.....	29
4.1 Choix de l'étiquette de langage.....	29
4.2 Signification de l'étiquette de langage.....	33
4.3 Listes de langages.....	34
4.4 Considérations de longueur.....	34
4.5 Canonisation des étiquettes de langage.....	35
4.6 Considérations sur les sous étiquettes d'utilisation privée.....	37
5. Considérations relatives à l'IANA.....	37
5.1 Registre des sous étiquettes de langages.....	37

5.2 Registre des extensions.....	38
6. Considérations sur la sécurité.....	38
7. Considérations de jeux de caractères.....	39
8. Changements par rapport à la RFC 4646.....	39
9. Références.....	41
9.1 Références normatives.....	41
9.2 Références pour information.....	42
Appendice A. Exemples d'étiquettes de langages (information).....	43
Appendice B. Exemples de formulaire d'enregistrement.....	44
Appendice C. Remerciements.....	44
Adresse des auteurs.....	45

## 1. Introduction

Les êtres humains sur notre planète ont, dans le passé et à présent, utilisé un certain nombre de langages. Il y a de nombreuses raisons pour vouloir identifier le langage utilisé quand on présente ou demande des informations.

Le langage d'un élément d'information ou les préférences de langage d'un utilisateur ont souvent besoin d'être identifiées afin que le traitement approprié puisse être appliqué. Par exemple, les préférences de langage de l'utilisateur dans un navigateur de la Toile peuvent être utilisées pour choisir les pages appropriées de la Toile. Les informations de langage peuvent aussi être utilisées pour choisir les outils (comme des dictionnaires) pour aider au traitement ou à la compréhension du contenu dans différents langages. La connaissance du langage particulier utilisé par certains éléments du contenu de l'information pourrait être utile ou même exigée par certains types de traitements, par exemple, la vérification de l'orthographe, la parole synthétisée par ordinateur, la transcription du Braille, ou du rendu d'impression de haute qualité.

Un moyen d'indiquer le langage utilisé est d'étiqueter le contenu des informations avec un identifiant ou "étiquette". Ces étiquettes peuvent aussi être utilisées pour spécifier les préférences de l'utilisateur lors du choix du contenu des informations ou pour étiqueter des attributs supplémentaires du contenu et des ressources associées.

Parfois des étiquettes de langage sont utilisées pour indiquer des attributs supplémentaires du langage du contenu. Par exemple, indiquer des informations spécifiques du dialecte, du système d'écriture, ou de l'orthographe utilisés dans un document ou ressource peut permettre à l'utilisateur d'obtenir des informations sous une forme qu'il peut comprendre, ou elles peuvent être importantes dans le traitement ou le rendu du contenu donné dans une forme ou style approprié.

Le présent document spécifie un mécanisme d'identifiant particulier (l'étiquette de langage) et une fonction d'enregistrement pour les valeurs à utiliser pour former les étiquettes. Il définit aussi un mécanisme pour les valeurs d'utilisation privée et de futures extensions.

Le présent document remplace la [RFC4646] (qui rendait obsolète la [RFC3066] qui, à son tour, remplaçait la [RFC1766]). Le présent document, combiné avec la [RFC4647], compose le BCP 47. Pour une liste des changements dans le présent document, voir la Section 8.

Les mots clés "DOIT", "NE DOIT PAS", "EXIGE", "DEVRA", "NE DEVRA PAS", "DEVRAIT", "NE DEVRAIT PAS", "RECOMMANDE", "PEUT", et "FACULTATIF" en majuscules dans ce document sont à interpréter comme décrit dans le BCP 14, [RFC2119].

## 2. L'étiquette de langage

Les étiquettes de langage sont utilisées pour aider à identifier les langages, parlés, écrits, par signes, ou signalés autrement, pour les besoins de la communication. Cela inclut les langages construits et artificiels mais exclut les langages non destinés principalement à la communication humaine, comme les langages de programmation.

### 2.1 Syntaxe

Une étiquette de langage est composée d'une séquence d'une ou plusieurs "sous étiquettes", dont chacune précise ou rétrécit la gamme de langages identifiée par l'étiquette globale. Les sous étiquettes, à leur tour, sont une séquence de caractères alphanumériques (lettres et chiffres) distingués et séparés des autres sous étiquettes dans une étiquette par un tiret ("-", [Unicode] U+002D).

Il y a différents types de sous étiquette, dont chacune est distinguée par sa longueur, sa position dans l'étiquette, et son contenu : chaque type de sous étiquette peut être reconnu seulement par ces caractéristiques. Cela rend possible d'extraire et

allouer des informations de signification aux sous étiquettes, même si les valeurs spécifiques de sous étiquette ne sont pas reconnues. Donc, un processeur d'étiquette de langage n'a pas besoin d'avoir une liste des étiquettes ou sous étiquettes valides (c'est-à-dire, une copie du registre IANA des sous étiquettes de langage) afin d'effectuer les recherches courantes et les opérations de confrontation. Les seules exceptions à cette capacité de déduire la signification de la structure des sous étiquettes sont les étiquettes faisant l'objet d'une clause d'antériorité dans les productions "regular" et "irregular" mentionnées plus loin. Ces étiquettes étaient enregistrées sous la [RFC3066] et sont une liste fixe qui ne peut jamais changer.

La syntaxe de l'étiquette de langage en ABNF [RFC5234] est :

```

Language-Tag = langtag                ; étiquettes de langage normales
               / privateuse           ; étiquette d'utilisation privée
               / grandfathered        ; étiquettes faisant l'objet d'une clause d'antériorité

langtag = language
         ["-" script]
         ["-" region]
         *("-" variant)
         *("-" extension)
         ["-" privateuse]

language = 2*3ALPHA                    ; plus court code ISO 639
          ["-" extlang]                 ; parfois suivi par des sous étiquettes étendues de langage,
          / 4ALPHA                      ; ou réservé pour utilisation future,
          / 5*8ALPHA                    ; ou sous étiquette de langage enregistrée.

extlang = 3ALPHA                       ; codes ISO 639 choisis
         *2("-" 3ALPHA)                 ; réservé de façon permanente

script = 4ALPHA                         ; code ISO 15924

region = 2ALPHA                         ; code ISO 3166-1
        / 3DIGIT                       ; code UN M.49

variant = 5*8alphanumeric              ; variantes enregistrées
        / (DIGIT 3alphanumeric)

extension = singleton 1*("-" (2*8alphanumeric))

singleton = DIGIT                       ; alphanumériques seuls, "x" réservé pour utilisation privée
          / %x41-57                     ; 0 - 9
          / %x59-5A                     ; A - W
          / %x61-77                     ; Y - Z
          / %x79-7A                     ; a - w
          / %x79-7A                     ; y - z

privateuse = "x" 1*("-" (1*8alphanumeric))

grandfathered = irregular / regular     ; étiquettes non redondantes enregistrées à l'époque de la RFC 3066

irregular = "en-GB-oed"                  ; étiquettes irrégulières qui ne correspondent pas à la production
        / "i-ami"                       ; 'langtag' et ne seraient autrement pas considérées comme
        / "i-bnn"                       ; "bien formées. Ces étiquettes sont toutes valides, mais la plupart
        / "i-default"                   ; sont déconseillées en faveur d'étiquettes ou combinaison de
        / "i-enochian"                   ; sous étiquettes plus modernes.
        / "i-hak"
        / "i-klingon"
        / "i-lux"
        / "i-mingo"
        / "i-navajo"
        / "i-pwn"
        / "i-tao"
        / "i-tay"
        / "i-tsu"

```

```

    / "sgn-BE-FR"
    / "sgn-BE-NL"
    / "sgn-CH-DE"

regular = "art-lojban"           ; ces étiquettes correspondent à la production 'langtag',
    / "cel-gaulish"             ; mais leurs sous étiquettes ne sont pas des sous étiquettes
    / "no-bok"                  ; de langage étendu ou de variantes :
    / "no-nyn"                  ; leur signification est définie par leur enregistrement, et toutes
    / "zh-guoyu"                ; sont déconseillées en faveur d'une sous étiquette
    / "zh-hakka"                ; ou séquence de sous étiquettes plus moderne.
    / "zh-min"
    / "zh-min-nan"
    / "zh-xiang"

alphanum = (ALPHA / DIGIT)      ; lettres et nombres

```

**Figure 1 : ABNF des étiquettes de langage**

Pour des exemples d'étiquettes de langage, voir l'Appendice A.

Toutes les sous étiquettes ont une longueur maximum de huit caractères. Les espaces ne sont pas permises dans une étiquette de langage. Il y a une subtilité dans la production ABNF "variant" : une variante commençant par un chiffre a une longueur minimum de quatre caractères, tandis que celles commençant par une lettre ont une longueur minimum de cinq caractères.

Bien que la [RFC5234] se réfère à des octets, les étiquettes de langage décrites dans le présent document sont des séquences de caractères du répertoire US-ASCII [ISO646]. Les étiquettes de langage PEUVENT être utilisées dans des documents et applications qui utilisent d'autres codages, pour autant qu'ils englobent la partie pertinente du répertoire de l'US-ASCII. Un exemple de cela serait un document XML qui utilise le codage UTF-16LE [RFC2781] de [Unicode].

### 2.1.1 Formatage de l'étiquette de langage

Les étiquettes de langage et leurs sous étiquettes, y compris d'utilisation privée et d'extensions, sont toujours à traiter comme insensibles à la casse : il existe des conventions pour la mise en majuscules de certaines des sous étiquettes, mais cela NE DOIT PAS être pris comme portant une signification.

Donc, l'étiquette "mn-Cyrl-MN" n'est pas distincte de "MN-cYrL-mn" ou "mN-cYrL-Mn" (ou toute autre combinaison) et chacune de ces variations porte la même signification : mongol écrit en écriture cyrillique comme utilisé en Mongolie.

La syntaxe d'ABNF ne distingue pas non plus entre majuscules et minuscules : les lettres US-ASCII majuscules dans la gamme "A" à "Z" sont toujours considérées équivalentes et transposées directement en leur équivalent US-ASCII minuscule dans la gamme de "a" à "z". Donc l'étiquette "I-AMI" est considérée comme équivalente à cette valeur "i-ami" dans la production "irregular".

Bien que les distinctions de casse ne portent pas de signification dans les étiquettes de langage, un formatage et une présentation cohérente des étiquettes de langage va aider les utilisateurs. Le format des sous étiquettes dans le registre est RECOMMANDÉ comme la forme pour utiliser les étiquettes de langage. Ce format correspond généralement aux conventions courantes pour les diverses normes ISO d'où les sous étiquettes sont dérivées.

Ces conventions incluent :

- o [ISO639-1] recommande que les codes de langage soient écrits en minuscules ("mn" mongolien).
- o [ISO15924] recommande que les codes de script utilisent les minuscules avec la lettre initiale en majuscule ("Cyrl" pour cyrillique).
- o [ISO3166-1] recommande que les codes de pays soient en majuscules ("MN" Mongolie).

Une mise en œuvre peut reproduire ce format sans accéder au registre comme suit. Toutes les sous étiquettes, incluant les sous étiquettes d'extension et d'utilisation privée, utilisent des lettres minuscules avec deux exceptions : les sous étiquettes de deux et quatre lettres qui n'apparaissent ni au début de l'étiquette ni après des singletons. De telles sous étiquettes de deux lettres sont toutes en majuscules (comme dans les étiquettes "en-CA-x-ca" ou "sgn-BE-FR") et les sous étiquettes de quatre lettres sont avec la première lettre en majuscule (comme dans l'étiquette "az-Latn-x-latn").

Note : Le repli de casse de lettres ASCII dans certaines situations, si il n'est pas traité avec soin, produit parfois des valeurs de caractère non ASCII. Le fichier de la base de données des caractères Unicode "SpecialCasing.txt" [SpecialCasing] définit les cas spécifiques qui sont connus pour causer des problèmes avec cela. En particulier, la

lettre "i" (U+0069) en Turc et en Azéri est mise en majuscule U+0130 (LATIN CAPITAL LETTER I WITH DOT ABOVE, *I majuscule latin avec point au dessus*). Les mises en œuvre DEVRAIENT spécifier une opération locale neutre à l'égard de la casse pour s'assurer que le repli de casse des sous étiquettes ne produit pas cette valeur, qui est illégale dans les étiquettes de langage. Par exemple, si on mettaient en majuscule la sous étiquette de région 'in' en utilisant les règles locales du Turc, la séquence U+0130 U+004E résulterait, à la place du "IN" attendu.

## 2.2. Sources et interprétation d'étiquette de langage

L'espace de noms des étiquettes de langage et de leurs sous étiquettes est administré par l'Autorité d'allocation des numéros de l'Internet (IANA, *Internet Assigned Numbers Authority*) conformément aux règles de la Section 5 de ce document. Le registre des sous étiquettes de langues tenu par l'IANA est la source des sous étiquettes valides : d'autres normes référencées dans ce paragraphe fournissent le matériel de source pour ce registre.

Terminologie utilisée dans le document :

"Étiquette" se réfère à une étiquette de langage complète, comme dans "sr-Latn-RS" ou "az-Arab-IR". Des exemples d'étiquettes dans le présent document sont encloses entre guillemets ("en-US").

"Sous étiquette" se réfère à une section spécifique d'une étiquette, délimitée par un tiret, comme dans les sous étiquettes 'zh', 'Hant', et 'CN' dans l'étiquette "zh-Hant-CN". Les exemples de sous étiquettes dans le présent document sont encloses entre des apostrophes ('Hant').

"Code" se réfère aux valeurs définies dans des normes externes (et qui sont utilisées comme des sous étiquettes dans le présent document). Par exemple, 'Hant' est le code d'écriture [ISO15924] qui était utilisé pour définir la sous étiquette d'écriture 'Hant' à utiliser dans une étiquette de langage. Les exemples de codes dans le présent document sont enclos entre des apostrophes ('en', 'Hant').

Les étiquettes de langage sont conçues de telle façon que chaque type de sous étiquette ait une longueur et des restrictions de contenu uniques. Cela rend possible l'identification du type de sous étiquette, même si le contenu de la sous étiquette elle-même n'est pas reconnu. Cela permet l'analyse des étiquettes et leur traitement sans référence à la dernière version des normes sous-jacentes ou du registre de l'IANA et rend plus simple le traitement des exceptions associées lors de l'analyse des étiquettes.

Certaines des sous étiquettes dans le registre de l'IANA ne viennent pas d'une norme sous-jacente. Elles peuvent seulement apparaître dans des positions spécifiques dans une étiquette : elles peuvent seulement se produire comme sous étiquettes principales de langage ou comme sous étiquettes de variante.

Les séquences de sous étiquettes d'utilisation privée et d'extension DOIVENT se produire à la fin de la séquence de sous étiquettes et NE DOIVENT PAS être intercalées avec les sous étiquettes définies ailleurs dans le présent document. Ces séquences sont introduites par des sous étiquettes d'un seul caractère, qui sont réservées comme suit :

- o La sous étiquette d'une seule lettre 'x' introduit une séquence de sous étiquettes d'utilisation privée. L'interprétation de toute sous étiquette d'utilisation privée est définie seulement par accord privé et n'est pas définie par les règles de ce paragraphe ni aucune norme ou registre défini dans le présent document.
- o La sous étiquette d'une seule lettre 'i' est utilisée par certaines étiquettes faisant l'objet d'une clause d'antériorité, comme "i-default", où elle apparaît toujours en première position et ne peut pas être confondue avec une extension.
- o Toutes les autres sous étiquettes d'une seule lettre et d'un seul chiffre sont réservées pour introduire des séquences de sous étiquette d'extension normalisées comme décrit au paragraphe 3.7.

### 2.2.1 Sous étiquette principale de langage

La sous étiquette principale de langage est la première sous étiquette dans une étiquette de langage et ne peut pas être omise, avec deux exceptions :

- o La sous étiquette d'un seul caractère 'x' comme sous étiquette principale indique que l'étiquette de langage consiste seulement en sous étiquettes dont la signification est définie par un accord privé. Par exemple, dans l'étiquette "x-fr-CH", les sous étiquettes 'fr' et 'CH' ne représentent pas le langage français ou le pays de Suisse (ou toute autre valeur du registre de l'IANA) sauf si il y a un accord privé pour le faire . Voir le paragraphe 4.6.
- o La sous étiquette d'un seul caractère 'i' est utilisée par certaines étiquettes faisant l'objet d'une clause d'antériorité (voir le paragraphe 2.2.8) comme "i-klinton" et "i-bnn". (D'autres étiquettes faisant l'objet d'une clause d'antériorité ont une sous étiquette de langage principale dans leur première position.)

Les règles suivantes s'appliquent à la sous étiquette principale de langage :

1. Des sous étiquettes principales de langage de deux caractères ont été définie dans le registre de l'IANA conformément aux allocations qui se trouvent dans la norme ISO 639-1:2002, "Codes pour la représentation des noms de langages -- Partie 1 : Code Alpha-2" [ISO639-1], ou en utilisant les allocations faites ensuite par l'autorité d'enregistrement de ISO 639-1 ou les organes qui gouvernent la normalisation.

2. Des sous étiquettes principales de langage de trois caractères ont été définies dans le registre de l'IANA conformément aux allocations qui se trouvent dans une des parties additionnelles de ISO 639 ou aux allocations faites ensuite par les autorités d'enregistrement pertinentes de ISO 639 ou les organes qui gouvernent la normalisation :
  - A. ISO 639-2:1998 "Codes pour la représentation des noms de langages -- Partie 2 : Code Alpha-3 - édition 1", [ISO639-2]
  - B. ISO 639-3:2007 "Codes pour la représentation des noms de langages -- Partie 3 : Code Alpha-3 pour la couverture complète des langages", [ISO639-3]
  - C. ISO 639-5:2008 "Codes pour la représentation des noms de langages -- Partie 5 : Code Alpha-3 pour les familles et groupes de langages", [ISO639-5]
3. Les sous étiquettes dans la gamme 'qaa' à 'qtz' sont réservées pour utilisation privée dans les étiquettes de langage. Ces sous étiquettes correspondent aux codes réservés par la norme ISO 639-2 pour utilisation privée. Ces codes PEUVENT être utilisés pour des sous étiquettes non enregistrées de langages principaux (à la place d'utiliser des sous étiquettes d'utilisation privée suivant 'x-'). Se reporter au paragraphe 4.6 pour plus d'informations sur les sous étiquettes d'utilisation privée.
4. Les sous étiquettes de langage de quatre caractères sont réservées pour une future normalisation possible.
5. Toute sous étiquette de langage de quatre à huit caractères dans le registre de l'IANA a été définie via le processus d'enregistrement du paragraphe 3.5 et PEUT être utilisée pour former une sous étiquette principale de langage. Un exemple de ce qu'un tel enregistrement pourrait inclure est l'enregistrement IANA faisant l'objet d'une clause d'antériorité "i-enochian". La sous étiquette 'enochian' pourrait être enregistrée dans le registre IANA comme une sous étiquette principale de langage (en supposant que ISO 639 n'ait pas enregistré ce langage d'abord) rendant valides des étiquettes comme "enochian-AQ" et "enochian-Latn". Au moment de la création du présent document, il n'y avait pas d'exemple de cette sorte de sous étiquette. De futurs enregistrements de ce type sont déconseillés : une tentative d'enregistrement de tout nouveau langage principal proposé DOIT être faite auprès de l'autorité d'enregistrement de la norme ISO 639. Les propositions rejetées par l'autorité d'enregistrement de la norme ISO 639 ont peu de chances de satisfaire les critères pour les sous étiquettes principales de langage et ont donc peu de chances d'être enregistrées.
6. D'autres valeurs NE DOIVENT PAS être allouées à une sous étiquette principale sauf par révision ou mise à jour du présent document.

Quand les langages ont à la fois le code ISO 639-1 à deux caractères et un code à trois caractères (alloué par ISO 639-2, ISO 639-3, ou ISO 639-5) seul le code ISO 639-1 à deux caractères est défini dans le registre de l'IANA.

Quand un langage n'a pas de code ISO 639-1 à deux caractères et que le code ISO 639-2/T (Terminologie) et le code ISO 639-2/B (Bibliographie) pour ce langage diffèrent, seul le code de terminologie est défini dans le registre IANA. Au moment de la création de ce document, tous les langages qui avaient les deux sortes de codes à trois caractères avaient aussi un code à deux caractères alloué ; on suppose que de futures allocations de cette nature ne se produiront plus.

Afin d'éviter l'instabilité de la forme canonique des étiquettes, si un code à deux caractères est ajouté à ISO 639-1 pour un langage pour lequel un code à trois caractères a déjà été inclus dans ISO 639-2 ou ISO 639-3, le code à deux caractères NE DOIT PAS être enregistré. Voir le paragraphe 3.4.

Par exemple, si un contenu a été étiqueté avec 'haw' (hawaïen) qui n'a actuellement pas de code à deux caractères, l'étiquette n'aura pas besoin d'être changée si ISO 639-1 devait ultérieurement allouer un code à deux caractères au langage hawaïen.

Pour éviter ces problèmes avec le choix de version et de sous étiquette (comme rencontré durant la transition entre les RFC 1766 et RFC 3066) ainsi que pour assurer la nature canonique des sous étiquettes définie par le présent document, le comité consultatif conjoint d'autorités d'enregistrement de la norme ISO 639 (ISO 639/RA-JAC, *Registration Authority Joint Advisory Committee*) a inclus la déclaration suivante dans [iso639.prin] : "Un code de langage déjà dans ISO 639-2 au point de geler ISO 639-1 ne devra pas être ultérieurement ajouté à ISO 639-1. C'est pour assurer la cohérence des usages au fil du temps, comme les utilisateurs sont conduits dans les applications de l'Internet à employer le code alpha-3 quand un code alpha-2 n'est pas disponible pour ce langage."

### 2.2.2 Sous étiquettes de langage étendues

Des sous étiquettes de langage étendues sont utilisées pour identifier certains langages spécialement choisis qui, pour diverses raisons historiques et de compatibilité sont étroitement identifiés avec, ou étiquetés en utilisant une sous étiquette de langage principale existante. Les sous étiquettes de langage étendues sont toujours utilisées avec leur sous étiquette de langage principale englobante (indiquée avec un champ "Prefix" dans le registre) quand elles sont utilisées pour former l'étiquette de langage. Tous les langages qui ont une sous étiquette de langage étendue dans le registre ont aussi un

enregistrement de sous étiquette de langage principal identique dans le registre. Cette sous étiquette principale de langage est RECOMMANDÉE pour former l'étiquette de langage. Les règles suivantes s'appliquent aux sous étiquettes de langage étendues :

1. Les sous étiquettes de langage étendues consistent seulement en sous étiquettes de trois lettres. Tous les enregistrements de sous étiquette de langage étendue définis dans le registre ont été définis en accord avec les allocations de [ISO639-3]. Les collections et groupements de langage, comme définis dans [ISO639-5], sont spécifiquement exclus des sous étiquettes de langage étendues.
2. Les enregistrements de sous étiquette de langage étendue DOIVENT inclure exactement un champ "Prefix" indiquant une sous étiquette ou séquence de sous étiquettes appropriée pour cette sous étiquette de langage étendue.
3. Les enregistrements de sous étiquette de langage étendue DOIVENT inclure une "Valeur préférée". Les champs "Valeur préférée" et "Sous étiquette" DOIVENT être identiques.
4. Bien que la production ABNF "extlang" permette jusqu'à trois étiquettes de langage étendues dans l'étiquette de langage, les sous étiquettes de langage étendues NE DOIVENT PAS inclure une autre sous étiquette de langage étendue dans leur "Prefix". C'est-à-dire, la seconde et la troisième position de sous étiquette de langage étendue dans une étiquette de langage sont réservées de façon permanente et les étiquettes qui incluent ces sous étiquettes dans cette position sont, et vont toujours rester, invalides.

Par exemple, le langage macrole chinois ('zh') englobe un certain nombre de langages. Pour des raisons de compatibilité, chacun de ces langages a à la fois une sous étiquette de langage principale et étendue dans le registre. Des exemples de cela incluent le chinois gan ('gan'), le chinois cantonais ('yue'), et le chinois mandarin ('cmn'). Chacun est englobé par le langage macrole 'zh' (Chinois). Donc, ils ont chacun le préfixe "zh" dans leurs enregistrements du registre. Donc, le chinois gan est représenté avec des étiquettes commençant par "zh-gan" ou "gan", le cantonais avec des étiquettes commençant soit par "yue" soit par "zh-yue", et le chinois mandarin avec "zh-cmn" ou "cmn". La sous étiquette de langage 'zh' peut encore être utilisée sans sous étiquette de langage étendue pour étiqueter une ressource comme une variété non spécifiée de chinois, alors que la sous étiquette de langage principale ('gan', 'yue', 'cmn') est préférée pour utiliser la forme de langage étendue ("zh-gan", "zh-yue", "zh-cmn").

### 2.2.3 Sous étiquette de script

Des sous étiquettes de script sont utilisées pour indiquer les variantes de système de script ou d'écriture qui distinguent les formes écrites d'un langage ou de ses dialectes. Les règles suivantes s'appliquent aux sous étiquettes de script :

1. Les sous étiquettes de script DOIVENT suivre toutes les sous étiquettes de langage principale et étendue et DOIVENT précéder tous les autres types de sous étiquette.
2. Les sous étiquettes de script consistent en quatre lettres et ont été définies en accord avec les allocations de la norme [ISO15924] ("Information et documentation -- Codes pour la représentation des noms d'écritures") ou allouées ensuite par l'autorité d'enregistrement de la norme ISO 15924 ou les organisations qui gouvernent la normalisation. Seuls les codes alloués par ISO 15924 seront pris en compte pour l'enregistrement.
3. Les sous étiquettes de script 'Qaaa' à 'Qabx' sont réservées pour utilisation privée dans les étiquettes de langage. Ces sous étiquettes correspondent aux codes réservés par la norme ISO 15924 pour utilisation privée. Ces codes PEUVENT être utilisés pour des valeurs de script non enregistrées. Voir au paragraphe 4.6 plus d'informations sur les sous étiquettes d'utilisation privée.
4. Il DOIT y avoir au plus une sous étiquette de script dans une étiquette de langage, et la sous étiquette de script DEVRAIT être omise quand elle n'ajoute pas de valeur distinctive à l'étiquette ou quand l'enregistrement de la sous étiquette de langage principale ou étendue dans le registre de sous étiquettes inclut un champ "Suppress-Script" faisant la liste des sous étiquettes de script applicables.

Par exemple : "sr-Latn" représente le serbe écrit en utilisant l'écriture latine.

### 2.2.4 Sous étiquette de région

Les sous étiquettes de région sont utilisées pour indiquer les variations linguistiques associées à, ou appropriées pour, un pays, territoire, ou région spécifique. Normalement, une sous étiquette de région est utilisée pour indiquer des variations comme des dialectes ou usages régionaux, ou des conventions orthographiques spécifiques d'une région. Elles peuvent aussi être utilisées pour indiquer que le contenu est exprimé d'une façon qui est appropriée pour l'usage à travers une région, par exemple, un contenu espagnol agencé pour être utile en Amérique latine.

Les règles suivantes s'appliquent aux sous étiquettes de région :

1. Les sous étiquettes de région DOIVENT suivre toutes les sous étiquettes de langage principal, langage étendu, ou script et DOIVENT précéder tous les autres types de sous étiquette.
2. Les sous étiquettes de région de deux lettres ont été définies en accord avec les allocations de la norme [ISO3166-1] ("Codes pour la représentation des noms de pays et leurs subdivisions -- Partie 1 : Codes de pays") en utilisant la liste alpha-2 des codes de pays ou en utilisant les allocations faites ensuite par l'agence de maintenance de ISO 3166-1 ou les organismes qui gouvernent la normalisation. De plus, les codes qui sont "exceptionnellement réservés" (par opposition à "alloués") dans la norme ISO 3166-1 ont aussi été définis dans le registre, à l'exception de 'UK', qui est un synonyme exact du code alloué 'GB'.
3. Les sous étiquettes de région 'AA', 'QM'-'QZ', 'XA'-'XZ', et 'ZZ' sont réservées pour utilisation privée dans les étiquettes de langage. Ces sous étiquettes correspondent aux codes réservés par ISO 3166 pour utilisation privée. Ces codes PEUVENT être utilisés pour des sous étiquettes de région d'utilisation privée (à la place d'utiliser une séquence de sous étiquettes d'utilisation privée). Voir au paragraphe 4.6 plus d'informations sur les sous étiquettes d'utilisation privée.
4. Les sous étiquettes de région de trois caractères consistent seulement en caractères de chiffres (nombre) et ont été définies en accord avec les allocations qui se trouvent dans les codes standard de pays ou zones pour utilisation statistique [UN\_M.49] ou les allocations faites ultérieurement par les organes qui gouvernent la normalisation. Tous les codes de UN M.49 ne sont pas définis dans le registre de l'IANA. Les règles suivantes définissent quels codes sont entrés dans le registre comme des sous étiquettes valides :
  - A. Les codes numériques UN alloués aux régions "macro géographiques" (continentales) ou aux sous régions DOIVENT être enregistrés dans le registre. Ces codes ne sont pas associés à un code alpha-2 alloué par ISO 3166-1 et représentent des zones supra-nationales, couvrant généralement plus d'une nation, état, province, ou territoire.
  - B. Les codes numériques UN pour les "groupements économiques" ou "autres groupements" NE DOIVENT PAS être enregistrés dans le registre IANA et NE DOIVENT PAS être utilisés pour former des étiquettes de langage.
  - C. Quand ISO 3166-1 réalloue un code anciennement utilisé pour un pays ou zone à un autre pays ou zone et que ce code déjà est présent dans le registre, le code numérique UN pour ce pays ou zone DOIT être enregistré dans le registre comme décrit au paragraphe 3.4 et DOIT être utilisé pour former des étiquettes de langage qui représentent le pays ou région pour lequel elles sont définies (plutôt que le code ISO 3166-1 recyclé).
  - D. Les codes numériques UN pour les pays ou zones pour lesquels il y a un code alpha-2 ISO 3166-1 associé dans le registre NE DOIVENT PAS être entrés dans le registre et NE DOIVENT PAS être utilisés pour former des étiquettes de langage. Noter que la sous étiquette fondée sur ISO 3166 dans le registre DOIT en fait être associée au code UN M.49 en question.
  - E. Pour des raisons historiques, le code numérique UN 830 (Îles anglo-normandes) qui n'a pas été enregistré au moment où ce document a été adopté et n'avait, à ce moment, aucun code ISO 3166-1 correspondant, PEUT être entré dans le registre IANA via le processus décrit au paragraphe 3.5, pourvu qu'aucun code ISO 3166-1 avec cette signification exacte n'ait été précédemment enregistré.
  - F. Pour tous les autres codes numériques UN pour des pays ou zones qui n'ont pas de code alpha-2 ISO 3166-1 associé, ils NE DOIVENT PAS être entrés dans le registre et NE DOIVENT PAS être utilisés pour former des étiquettes de langage. Pour plus d'informations sur ces codes, voir le paragraphe 3.4.
5. Les codes alphanumériques de l'Appendice X du document des Nations Unies NE DOIVENT PAS être entrés dans le registre et NE DOIVENT PAS être utilisés pour former des étiquettes de langage. (Au moment de la création du présent document, ces valeurs correspondaient aux codes alpha-2 de ISO 3166-1.)
6. Il DOIT y avoir au plus une sous étiquette de région dans une étiquette de langage et une sous étiquette de région PEUT être omise quand elle n'ajoute aucune valeur perceptible à l'étiquette.

Par exemple :

"de-AT" représente l'allemand ('de') comme utilisé en Autriche ('AT').

"sr-Latn-RS" représente le serbe ('sr') écrit en utilisant l'écriture latine ('Latn') comme utilisée en Serbie ('RS').

"es-419" représente l'espagnol ('es') approprié pour la région Amérique Latine et Caraïbes définie par les Nations Unies ('419').

### 2.2.5 Sous étiquettes de variante

Les sous étiquettes de variantes sont utilisées pour indiquer des variantes supplémentaires, bien reconnues, qui définissent un langage ou ses dialectes qui ne sont pas couverts par les autres sous étiquettes disponibles. Les règles suivantes s'appliquent aux sous étiquettes de variantes :

1. Les sous étiquettes de variantes DOIVENT suivre toutes sous étiquettes de langage principal, de langage étendu, de script, ou de région et DOIVENT précéder toutes séquences de sous étiquettes d'extension ou d'utilisation privée.
2. Les sous étiquettes de variantes, comme une collection, ne sont pas associées à une norme externe particulière. La signification d'une sous étiquette de variante dans le registre est définie dans le cours du processus d'enregistrement défini au paragraphe 3.5. Noter que toute sous étiquette de variante particulière pourrait être associée à une norme externe. Cependant, l'association à une norme n'est pas exigée pour l'enregistrement.
3. Plus d'une variante PEUT être utilisée pour former l'étiquette de langage.
4. Les sous étiquettes de variantes DOIVENT être enregistrées par l'IANA conformément aux règles du paragraphe 3.5 du présent document avant d'être utilisées pour former des étiquettes de langage. Afin de distinguer les variantes des autres types de sous étiquettes, les enregistrements DOIVENT satisfaire les restrictions suivantes de longueur et contenu :
  1. Les sous étiquettes de variantes qui commencent par une lettre (a-z, A-Z) DOIVENT faire au moins cinq caractères.
  2. Les sous étiquettes de variantes qui commencent par un chiffre (0-9) DOIVENT faire au moins quatre caractères.
5. La même sous étiquette de variante NE DOIT PAS être utilisée plus d'une fois au sein d'une étiquette de langage.
  - \* Par exemple, l'étiquette "de-DE-1901-1901" n'est pas valide.

Les enregistrements de sous étiquette de variante dans le registre des sous étiquettes de langage PEUVENT inclure un ou plusieurs champs "Prefix" (paragraphe 3.1.8). Chaque "Prefix" indique une séquence convenable de sous étiquettes pour former (avec les autres sous étiquettes appropriées) une étiquette de langage quand on utilise la variante.

La plupart des variantes qui partagent un préfixe sont mutuellement exclusives. Par exemple, les variations orthographiques allemandes '1996' et '1901' NE DEVRAIENT PAS être utilisées dans la même étiquette, car elles représentent les dates de réforme de l'orthographe différentes. Une variante qui peut significativement être utilisée en combinaison avec une autre variante DEVRAIT inclure un champ "Prefix" dans son enregistrement de registre qui fasse la liste des autres variantes. Par exemple, si une autre variante de l'allemand "exemple" était créée qui puisse raisonnablement être utilisée avec '1996', alors "exemple" devrait inclure deux champs "Prefix" : "de" et "de-1996".

Par exemple : "sl-nedis" représente le dialecte natisone ou nadiza du slovène ; "de-CH-1996" représente l'allemand utilisé en Suisse et écrit en utilisant la réforme de l'orthographe qui a débuté en 1996.

### 2.2.6 Sous étiquettes d'extension

Les extensions fournissent un mécanisme pour étendre les étiquettes de langage à utiliser dans diverses applications. Elles sont destinées à identifier les informations couramment utilisées en association avec les langages ou étiquettes de langage mais ne font pas partie de l'identification du langage (voir au paragraphe 3.7). Les règles suivantes s'appliquent aux extensions :

1. Une extension DOIT suivre au moins une sous étiquette de langage principale. C'est-à-dire, une étiquette de langage ne peut pas commencer par une extension. Les extensions étendent les étiquettes de langage, elles ne les outrepassent ni ne les remplacent. Par exemple, "a-value" n'est pas une étiquette de langage bien formée, tandis que "de-a-value" l'est. Noter que les extensions ne peuvent pas être utilisées dans des étiquettes qui sont entièrement d'utilisation privée (c'est-à-dire, des étiquettes commençant par "x-").
2. Les sous étiquettes d'extension sont séparées des autres sous étiquettes définies dans le présent document par une sous étiquette d'un seul caractère (appelée un "singleton"). Le singleton DOIT être celui alloué à une autorité d'enregistrement via le mécanisme décrit au paragraphe 3.7 et NE DOIT PAS être la lettre 'x', qui est réservée pour les séquences de sous étiquettes d'utilisation privée.
3. Chaque sous étiquette singleton DOIT apparaître au plus une fois dans chaque étiquette (autres que des sous étiquettes d'utilisation privée). C'est-à-dire, les sous étiquettes singleton NE DOIVENT PAS être répétées. Par exemple, l'étiquette "en-a-bbb-a-ccc" est invalide parce que la sous étiquette 'a' apparaît deux fois. Noter que l'étiquette "en-a-bbb-x-a-ccc" est valide parce que la seconde apparition du singleton 'a' est dans une séquence d'utilisation privée.
4. Les sous étiquettes d'extension DOIVENT satisfaire toutes les exigences établies par le document qui définit leur préfixe singleton et toutes les exigences fournies par l'autorité de maintenance. Noter qu'il pourrait ne pas y avoir de registre de ces sous étiquettes et que les processeurs de validation ne sont pas obligés de valider les extensions.
5. Chaque sous étiquette d'extension DOIT être de deux à huit caractères et consister seulement en lettres ou chiffres, avec chaque sous étiquette séparée par un seul '-'. Les distinctions de casse sont ignorées dans les extensions (comme avec toute sous étiquette de langage) et les sous étiquettes normalisées de ce type sont supposées être en minuscules.

6. Chaque singleton DOIT être suivi par au moins une sous étiquette d'extension. Par exemple, l'étiquette "tlh-a-b-foo" est invalide parce que le premier singleton "a" est suivi immédiatement par un autre singleton "b".
7. Les sous étiquettes d'extension DOIVENT suivre toutes les sous étiquettes principales de langage, de langage étendu, de script, de région, et de variantes dans une étiquette et DOIVENT précéder toute séquence de sous étiquette d'utilisation privée.
8. Toutes les sous étiquettes qui suivent le singleton et avant un autre singleton font partie de l'extension. Exemple : dans l'étiquette "fr-a-Latn", la sous étiquette 'Latn' ne représente pas la sous étiquette de script 'Latn' définie dans le registre des sous étiquettes de langage de l'IANA. Sa signification est définie par l'extension 'a'.
9. Dans le cas où plus d'une extension apparaît dans une seule étiquette, l'étiquette DEVRAIT être canonisée comme décrit au paragraphe 4.5, en rangeant les diverses séquences d'extension en ordre ASCII insensible à la casse.

Par exemple, si une extension était définie pour le singleton "r" et qu'elle définisse les sous étiquettes montrées, alors l'étiquette suivante serait un exemple valide : "en-Latn-GB-boont-r-extended-sequence-x-private".

### 2.2.7 Sous étiquettes d'utilisation privée

Les sous étiquettes d'utilisation privée sont utilisées pour indiquer des distinctions dans le langage qui sont importantes dans un contexte donné par un accord privé. Les règles suivantes s'appliquent aux sous étiquettes d'utilisation privée :

1. Les sous étiquettes d'utilisation privée sont séparées des autres sous étiquettes définies dans le présent document par la sous étiquette réservée d'un seul caractère 'x'.
2. Les sous étiquettes d'utilisation privée DOIVENT se conformer aux contraintes de format et de contenu définies dans l'ABNF pour toutes les sous étiquettes ; c'est-à-dire, elles DOIVENT consister seulement en lettres et chiffres et ne pas excéder huit caractères.
3. Les sous étiquettes d'utilisation privée DOIVENT suivre toutes les sous étiquettes de langage principales, de langage étendu, de script, de région, de variante, et d'extension, dans l'étiquette. Une autre façon de dire cela est que toutes les sous étiquettes suivant le singleton 'x' DOIVENT être considérées comme d'utilisation privée. Exemple : la sous étiquette 'US' dans l'étiquette "en-x-US" est une sous étiquette d'utilisation privée.
4. Une étiquette PEUT consister entièrement en sous étiquettes d'utilisation privée.
5. Aucune source n'est définie pour les sous étiquettes d'utilisation privée. L'utilisation de sous étiquettes d'utilisation privée est seulement faite par accord privé.
6. Les sous étiquettes d'utilisation privée sont NON RECOMMANDÉES lorsque des solutions de remplacement existent ou pour des échanges généraux. Voir au paragraphe 4.6 plus d'informations sur le choix de sous étiquette d'utilisation privée.

Par exemple, supposons qu'un groupe d'étudiants examine des textes en grec médiéval. Ils pourraient s'accorder pour utiliser une collection de sous étiquettes d'utilisation privée pour identifier différents styles d'écritures dans les textes. Par exemple, ils pourraient utiliser 'el-x-koine' pour les documents dans le style "courant" et utiliser 'el-x-attic' pour les autres documents qui imitent le style Attique. Ces sous étiquettes ne seraient pas reconnues par des processus ou systèmes externes, mais pourraient être utiles pour catégoriser divers textes à étudier par ceux du groupe.

Dans le registre, il y a aussi des sous étiquettes dérivées de codes réservés par ISO 639, ISO 15924, ou ISO 3166 pour utilisation privée. Ne pas les confondre avec les séquences de sous étiquette d'utilisation privée qui suivent la sous étiquette 'x'. Voir le paragraphe 4.6.

### 2.2.8 Enregistrements faisant l'objet d'une clause d'antériorité et redondants

Avant la RFC 4646, les étiquettes de langage complètes étaient enregistrées selon les règles des RFC 1766 et/ou RFC 3066. Toutes ces étiquettes enregistrées restent valides comme étiquettes de langage.

Beaucoup de ces étiquettes enregistrées sont devenues redondantes par l'arrivée de la RFC 4646 ou du présent document. Une étiquette redondante est un enregistrement faisant l'objet d'une clause d'antériorité dont les sous étiquettes individuelles apparaissent avec la même signification sémantique dans le registre. Par exemple, l'étiquette "zh-Hant" (chinois traditionnel) peut maintenant être composée à partir des sous étiquettes 'zh' (chinois) et 'Hant' (variante d'écriture Han

traditionnelle). Ces étiquettes redondantes sont maintenues dans le registre comme des enregistrements de type "redondant", principalement à titre de curiosité historique.

Le reste des étiquettes précédemment enregistrées sont "l'objet d'une clause d'antériorité". Ces étiquettes sont classées dans deux groupes : "regular" et "irregular".

Les étiquettes faisant l'objet d'une clause d'antériorité qui (paraissent) correspondre à la production 'langtag' dans la Figure 1 sont considérées comme des étiquettes faisant l'objet d'une clause d'antériorité "régulière". Ces étiquettes contiennent une ou plusieurs sous étiquettes qui soit n'apparaissent pas individuellement dans le registre, soit apparaissent mais avec une signification différente : chaque étiquette, dans sa totalité, représente un langage ou collection de langages.

Les étiquettes faisant l'objet d'une clause d'antériorité qui ne correspondent pas à la production "langtag" dans l'ABNF et seraient autrement invalides sont considérées comme des étiquettes faisant l'objet d'une clause d'antériorité "irrégulières". À l'exception de "en-GB-oed", qui est une variante de "en-GB", chacune d'elles, dans sa totalité, représente un langage.

Beaucoup des étiquettes faisant l'objet d'une clause d'antériorité ont été supplantées par l'ajout ultérieur de nouvelles sous étiquettes : chaque enregistrement supplanté contient un champ "Valeur préférée" qui devrait être utilisé pour former les étiquettes de langage représentant cette valeur. Par exemple, l'étiquette 'art-lojban' est supplantée par la sous étiquette de langage principale 'jbo'.

### 2.2.9 Classes de conformité

Les mises en œuvre ont parfois besoin de décrire leurs capacités à l'égard des règles et pratiques décrites dans le présent document. Les étiquettes peuvent être vérifiées de différentes façons, mais deux classes particulières de conformité d'étiquette sont définies formellement ici.

Une étiquette est considérée comme "bien formée" si elle se conforme à l'ABNF (paragraphe 2.1). Les étiquettes de langage peuvent être bien formées en termes de syntaxe mais non valides en termes de contenu. Cependant, de nombreuses opérations impliquant des étiquettes de langage fonctionnent bien sans rien savoir de la signification ou de la validité des sous étiquettes.

Une étiquette est considérée comme "valide" si elle satisfait ces conditions :

- o l'étiquette est bien formée,
- o soit l'étiquette est dans la liste des étiquettes faisant l'objet d'une clause d'antériorité, soit toutes ses sous étiquettes principales de langage, de langage étendu, de script, région, et variante apparaissent dans le registre des sous étiquettes de langage de l'IANA à la date du registre,
- o il n'y a pas de sous étiquette de variante dupliquée,
- o il n'y a pas de sous étiquette de singleton (extension) dupliquée.

Noter que la validité d'une étiquette dépend de la date du registre utilisée pour valider l'étiquette. Une copie plus récente du registre pourrait contenir une sous étiquette qu'une plus ancienne version n'a pas.

Une étiquette est considérée valide pour une extension donnée (paragraphe 3.7) (comme une version, révision, et date particulière) si elle satisfait les critères de "valide" ci-dessus et satisfait aussi cette condition : chaque sous étiquette utilisée dans la partie "extension" de l'étiquette est valide en accord avec l'extension.

Les plus anciennes spécifications ou les mises en œuvre d'étiquette de langage font parfois référence à la [RFC3066]. Un plus large ensemble d'étiquettes était considérées comme bien formées dans ce document. Toutes les étiquettes qui étaient valides sous la RFC 3066 sont à la fois bien formées et valides selon la syntaxe du présent document ; seule les étiquettes invalides ou illégales étaient bien formées selon la définition antérieure mais ne le sont plus. La syntaxe des étiquettes de langage selon la RFC 3066 était :

```
obs-language-tag = primary-subtag *( "-" subtag )
primary-subtag  = 1*8ALPHA
subtag          = 1*8(ALPHA / DIGIT)
```

**Figure 2 : Syntaxe d'étiquette de langage de la RFC 3066**

Les sous étiquettes destinées à utilisation privée ainsi que les séquences d'utilisation privée introduites par la sous étiquette 'x' sont disponibles pour les cas dans lesquels aucune sous étiquette allouée n'est disponible et où l'enregistrement n'est pas une option convenable. Par exemple, on pourrait utiliser une étiquette comme "no-QQ", où 'QQ' est dans une gamme de codes d'utilisation privée ISO 3166-1 pour indiquer une région par ailleurs non définie. Les utilisateurs NE DOIVENT PAS allouer des étiquettes de langage qui utilisent des sous étiquettes qui n'apparaissent pas dans le registre autrement que dans

des séquences d'utilisation privée (comme la sous étiquette "personal" dans l'étiquette "en-x-personal"). En plus de n'être pas valide, l'utilisateur risque aussi une collision avec une éventuelle future allocation ou enregistrement possible.

Note : bien que la production "Language-Tag" qui apparaît dans le présent document soit fonctionnellement équivalente à celle de la [RFC4646], elle a été changée pour prévenir certaines erreurs de formation provenant d'anciennes productions faisant l'objet d'une clause d'antériorité.

### 3. Format et maintenance de registre

Le registre IANA des sous étiquettes de langue ("le registre") contient une liste complète de toutes les sous étiquettes valides dans les étiquettes de langage. Cela permet aux mises en œuvre une façon directe et fiable de valider les étiquettes de langage. Le registre va être maintenu de telle façon que, sauf pour les sous étiquettes d'extension, il soit possible de valider toutes les sous étiquettes qui apparaissent dans l'étiquette de langage dans les dispositions du présent document ou ses révisions ou successeurs. De plus, la signification des diverses sous étiquettes va être sans ambiguïté et stable dans le temps. (La signification des sous étiquettes d'utilisation privée n'est, bien sûr pas définie par le registre.)

Cette Section définit le registre avec les procédures de maintenance et mise à jour qui lui sont associées, ainsi qu'un registre pour les extensions aux étiquettes de langage (paragraphe 3.7).

#### 3.1 Format du registre IANA des sous étiquettes de langage

Le registre IANA des sous étiquettes de langue est un fichier lisible par la machine dans un format décrit par ce paragraphe, plus des copies des formulaires d'enregistrement approuvés en accord avec le processus décrit au paragraphe 3.5.

Les formulaires d'enregistrement existants pour les étiquettes faisant l'objet d'une clause d'antériorité et redondantes tirées de la RFC 3066 ont été conservés au titre du registre obsolète de la RFC 3066. Les sous étiquettes ajoutées au registre par la [RFC4645] ou la [RFC5645] n'ont pas de formulaire d'enregistrement séparé (donc aucun formulaire n'est archivé pour ces ajouts).

##### 3.1.1 Format de fichier

Le registre est un fichier de texte [Unicode] et consiste en une série d'enregistrements dans un format fondé sur un "record-jar" (décrit dans [record-jar]). Chaque enregistrement, à son tour, consiste en une série de champs qui décrivent les diverses sous étiquettes et étiquettes. Le fichier du registre réel est codé en utilisant le codage de caractères UTF-8 [RFC3629].

Chaque champ peut être considéré comme une seule ligne logique de caractères. Chaque champ contient un "nom de champ" et un "corps de champ". Ils sont séparés par un "séparateur de champ". Le séparateur de champ est un caractère COLON (U+003A) (*deux-points*, :) plus des espaces blanches qui l'entourent. Chaque champ est terminé par la séquence CRLF de nouvelle ligne. Le texte de chaque champ DOIT être en forme C de normalisation (NFC, *Normalization Form C*) Unicode.

Une collection de champs forme un "enregistrement". Les enregistrements sont séparés par des lignes contenant seulement la séquence "%%" (U+0025 U+0025).

Bien que les champs soient logiquement une seule ligne de texte, chaque ligne de texte dans le format de fichier est limitée à une longueur de 72 octets. Pour s'accommoder de cela, le corps de champ peut être étalé sur plusieurs lignes ; c'est appelé le "repli à la ligne". Le repli à la ligne est fait en accord avec les conventions habituelles pour le saut à la ligne. C'est normalement sur une limite d'espace, mais peut se produire entre d'autres caractères quand la valeur n'inclut pas d'espaces, comme quand un langage n'utilise pas d'espace entre les mots. En tout cas, il NE DOIT PAS y avoir de coupure à l'intérieur d'une séquence UTF-8 multi-octets ou au milieu d'une séquence de caractères combinés. Pour plus d'informations, voir [UAX14].

Bien que le format de fichier utilise le jeu de caractères Unicode et que le fichier lui-même soit codé en utilisant le codage UTF-8, les champs sont restreints aux caractères imprimables du répertoire US-ASCII [ISO646] sauf mention contraire dans la description d'un champ spécifique (paragraphe 3.1.2).

Le format du registre est décrit par l'ABNF suivant [RFC5234]. Les numéros de caractères (codets) sont tirés de Unicode, et les terminaux dans les productions ABNF sont en termes de caractères plutôt que d'octets.

```
registry = record *("%%" CRLF record) ; registre = enregistrement
record   = 1*field ; enregistrement = 1 champ
```

```

field = ( field-name field-sep field-body CRLF ) ; champ = nom de champ, séparateur, corps de champ
field-name = (ALPHA / DIGIT) [*(ALPHA / DIGIT / "-") (ALPHA / DIGIT)] ; nom de champ
field-sep = *SP ":" *SP ; séparateur de champ
field-body = *(([*SP CRLF] 1*SP] 1*CHARS) ; corps de champ
CHARS = (%x21-10FFFF) ; codets Unicode

```

**Figure 3 :ABNF du format de registre**

La séquence "." (U+002E U+002E) dans un corps de champ note une gamme de valeurs. Une telle gamme représente toutes les sous étiquettes de la même longueur qui sont en ordre alphabétique ou numérique au sein de cette gamme, y compris les valeurs explicitement mentionnées. Par exemple, "a..c" note les valeurs "a", "b", et "c", et "11..13" note les valeurs "11", "12", et "13".

Tous les champs dont le corps de champ contient une valeur de date utilisent le format "full-date" (*date complète*) spécifiée dans la [RFC3339]. Par exemple, "2004-06-28" représente le 28 juin 2004, dans le calendrier grégorien.

### 3.1.2 Définitions d'enregistrement et de champ

Il y a trois types d'enregistrements dans le registre : "File-Date" (*date du fichier*), "Subtag" (*sous étiquette*), et "Tag" (*étiquette*).

Le premier enregistrement dans le registre est toujours l'enregistrement "File-Date". Cet enregistrement se produit seulement une fois dans le fichier et contient un seul champ dont le nom de champ est "File-Date". Le corps de champ de cet enregistrement contient une date (voir le paragraphe 5.1) rendant possible de reconnaître facilement les différentes versions du registre.

```

File-Date: 2004-06-28
%%

```

**Figure 4 : Exemple de l'enregistrement "File-Date"**

Les enregistrements suivants contiennent plusieurs champs et représentent les informations sur les sous étiquettes ou étiquettes. Les deux types d'enregistrements ont une structure identique, sauf que les enregistrements "Subtag" contiennent un champ avec un nom de champ de "Subtag", tandis que, sans surprise, les enregistrements "Tag" contiennent un champ avec un nom de champ de "Tag". Les noms de champ NE DOIVENT PAS se produire plus d'une fois par enregistrement, à l'exception des champs "Description", "Comments", et "Prefix".

Chaque enregistrement DOIT contenir au moins un de chaque champ suivant :

- o "Type"
  - \* Le corps de champ d'un type DOIT consister en une des chaînes suivantes : "language", "extlang", "script", "region", "variant", "grandfathered", et "redundant" ; il note le type d'étiquette ou sous étiquette.
- o "Subtag" ou "Tag"
  - \* Les corps de champ "Subtag" contiennent la sous étiquette définie. Ce champ DOIT apparaître dans tous les enregistrements dont le "Type" a une de ces valeurs : "language", "extlang", "script", "region", ou "variant".
  - \* Les corps de champ "Tag" contiennent une étiquette de langage complète. Ce champ DOIT apparaître dans tous les enregistrements dont le "Type" a une de ces valeurs : "grandfathered" ou "redundant". Si le "Type" est "grandfathered", alors le corps de champ "Tag" va être une des étiquettes mentionnées dans la production "regular" ou "irregular" du paragraphe 2.1.
- o "Description"
  - \* Les corps de champ "Description" contiennent une description non normative de la sous étiquette ou étiquette.
- o "Added"
  - \* Les corps de champ "Added" contiennent la date de l'enregistrement qui a été enregistré ou, dans le cas d'étiquettes faisant l'objet d'une clause d'antériorité ou redondante, la date à laquelle l'étiquette correspondante a été enregistrée selon les règles de la [RFC1766] ou [RFC3066].

Chaque enregistrement PEUT aussi contenir les champs suivants :

- o "Deprecated" (*déconseillé*)
  - \* Les corps de champ "Deprecated" contiennent la date à laquelle l'enregistrement a été déconseillé. Dans certains cas, cette valeur est plus tôt que celle du champ "Added" dans le même enregistrement. C'est-à-dire, la date où il est déconseillé précédée de l'ajout de l'enregistrement au registre.
- o "Valeur préférée"
  - \* Les corps de champ "Valeur préférée" contiennent une transposition canonique de la valeur de cet enregistrement en un équivalent moderne qui est préféré à sa place. Selon la valeur du champ "Type", cette valeur peut prendre différentes formes :

- + Pour les champs de type "language", "Valeur préférée" contient la sous étiquette de langage principale qui est préférée quand on forme l'étiquette de langage.
- + Pour les champs de type "script", "region", ou "variant", "Valeur préférée" contient la sous étiquette du même type qui est préférée pour former l'étiquette de langage.
- + Pour les champs de type "extlang", "grandfathered", ou "redundant", "Valeur préférée" contient un "gamme de langages étendue" [RFC4647] qui est préférée pour former l'étiquette de langage. C'est-à-dire, l'étiquette de langage préférée va contenir, dans l'ordre, chacune des sous étiquettes qui apparaissent dans la "Valeur préférée" ; des champs supplémentaires peuvent être inclus dans une étiquette de langage, comme décrit ailleurs dans le présent document. Par exemple, le remplacement de l'étiquette faisant l'objet d'une clause d'antériorité "zh-min-nan" (chinoid Min Nan) est "nan", qui peut être utilisée comme base pour des étiquettes comme "nan-Hant" ou "nan-TW" (noter que la forme étendue de sous étiquette de langage comme "zh-nan-Hant" ou "zh-nan-TW" peut aussi être utilisée).
- o "Prefix"
  - \* Le corps de champ "Prefix" contient une étiquette de langage valide qui est RECOMMANDÉE comme préfixe possible de cette sous étiquette d'enregistrement. Ce champ PEUT apparaître dans des enregistrements dont le corps de champ "Type" est soit "extlang", soit "variant" (il NE DOIT PAS apparaître dans d'autres types d'enregistrement).
- o "Suppress-Script"
  - \* Le corps de champ "Suppress-Script" contient une sous étiquette de script qui NE DEVRAIT PAS être utilisée pour former des étiquettes de langage avec la sous étiquette de langage principale associée ou étendue. Ce champ DOIT apparaître seulement dans des enregistrements dont le corps de champ "Type" est "langage" ou "extlang". Voir le paragraphe 4.1.
- o "Macrole language"
  - \* Le corps de champ "Macrole language" contient une sous étiquette de langage principal définie par la norme ISO 639 comme le "langage macrole" qui englobe cette sous étiquette de langage. Ce champ DOIT apparaître seulement dans les enregistrements dont le corps de champ "Type" est soit "language", soit "extlang".
- o "Scope"
  - \* Le corps de champ "Scope" contient des informations sur une sous étiquette de langage principale ou étendue qui indique le type du code de langage conforme à ISO 639. Les valeurs permises dans ce champ sont "macrole language", "collection", "special", et "private-use". Ce champ apparaît seulement dans les enregistrements dont le corps de champ "Type" est soit "language", soit "extlang". Quand ce champ est omis, le langage est un langage individuel.
- o "Comments"
  - \* Le corps de champ "Comments" contient des informations supplémentaires sur la sous étiquette, comme réputé approprié pour comprendre le registre et mettre en œuvre les étiquettes de langage en utilisant la sous étiquette ou étiquette.

De futures versions du présent document pourraient ajouter des champs supplémentaires au registre ; les mises en œuvre DEVRAIENT ignorer les champs trouvés dans le registre qui ne sont pas définis dans le présent document.

### 3.1.3 Champ "Type"

Le champ "Type" contient la chaîne qui identifie le type d'enregistrement dans lequel elle apparaît. Les valeurs pour le corps de champs "Type" sont : "language" (paragraphe 2.2.1) "extlang" (paragraphe 2.2.2) "script" (paragraphe 2.2.3) "region" (paragraphe 2.2.4) "variant" (paragraphe 2.2.5) "grandfathered" ou "redundant" (paragraphe 2.2.8).

### 3.1.4 Champs "Sous étiquette" et "Étiquette"

Le champ "Subtag" contient la sous étiquette définie dans l'enregistrement. Le champ "Tag" apparaît dans les enregistrements dont le "Type" est soit "grandfathered", soit "redundant" et contient une étiquette enregistrée dans la [RFC3066].

Le corps de champ "Subtag" DOIT suivre les conventions de casse décrites au paragraphe 2.1.1. Toutes les sous étiquettes utilisent des lettres minuscules dans le corps de champ, avec deux exceptions :

Les sous étiquettes dont le champ "Type" est "script" (en d'autres termes, les sous étiquettes définies par ISO 15924) DOIVENT utiliser la première lettre en majuscule.

Les sous étiquettes dont le champ "Type" est "region" (en d'autres termes, les sous étiquettes région non numériques définies par ISO 3166-1) DOIVENT être tout en majuscules.

Le corps de champ "Tag" DOIT être formaté en accord avec les règles décrites au paragraphe 2.1.1.

### 3.1.5 Champs "Description"

Le champ "Description" contient une description de l'étiquette ou sous étiquette dans l'enregistrement. Le champ "Description" PEUT apparaître plus d'une fois par enregistrement. Le champ "Description" PEUT inclure la gamme

complète des caractères Unicode. Au moins un des champs "Description" DOIT être écrit ou transcrit dans l'écriture latine ; des champs "Description" supplémentaires PEUVENT être dans toute écriture ou langage.

Le champ "Description" est utilisé à des fins d'identification. Les descriptions DEVRAIENT contenir toutes et seulement les informations nécessaires pour distinguer une sous étiquette des autres avec lesquelles elle pourrait être confondue. Elle ne sont pas destinées à fournir des informations générales ou à fournir tous les autres noms ou désignations possibles. Les champs "Description" ne représentent pas nécessairement le nom d'origine réel de l'élément dans l'enregistrement, et il n'est pas non plus garanti que les descriptions soient dans une langue particulière (comme l'anglais ou le français par exemple).

Les descriptions dans le registre qui correspondent aux codes ISO 639, ISO 15924, ISO 3166-1, ou UN M.49 sont destinées seulement à indiquer la signification de cet identifiant comme défini dans la norme de source au moment où il a été ajouté au registre ou comme modifié ensuite, dans la limite des règles de stabilité (paragraphe 3.4) via des enregistrements ultérieurs. La "Description" ne remplace pas le contenu de la norme de source elle-même. Les champs "Description" ne sont pas destinés à être le nom anglais localisé pour les sous étiquettes. La localisation ou la traduction des descriptions d'étiquette et sous étiquette de langage sort du domaine d'application du présent document.

Pour les sous étiquettes provenant d'une source standard (comme ISO 639 ou ISO 15924) les champs "Description" dans l'enregistrement sont aussi initialement pris de cette source standard. Plusieurs descriptions dans la source standard sont étalées sur des champs "Description" séparés. Les descriptions de la source standard PEUVENT être éditées ou modifiées, soit avant l'insertion, soit via le processus d'enregistrement, et des descriptions supplémentaires ou étrangères omises ou supprimées. Chaque champ "Description" DOIT être unique au sein de l'enregistrement dans lequel il apparaît, et formater des variations de la même description NE DEVRAIT PAS se produire dans cet enregistrement spécifique. Par exemple, bien que le code ISO 639-1 'fy' ait les deux descriptions "frisien occidental" et "frisien, occidental" dans cette norme, seulement une de ces descriptions apparaît dans le registre.

Pour s'assurer que les utilisateurs sont au clair sur la sous étiquette à utiliser, les champs "Description" alloués à un enregistrement de tout type spécifique ("langage", "extlang", "script", et autres) DOIVENT être uniques au sein de ce type d'enregistrement donné avec l'exception suivante : si un champ "Description" particulier se produit dans plusieurs enregistrements d'un type donné, alors au plus un des enregistrements peut omettre le champ "Deprecated". Tous les enregistrements déconseillés qui partagent une "Description" DOIVENT avoir la même "Valeur préférée", et tous les enregistrements non déconseillés DOIVENT avoir cette "Valeur préférée". Cela signifie que deux enregistrements de même type qui partagent une "Description" sont aussi sémantiquement équivalents et pas plus d'un enregistrement avec une "Description" donnée n'est préféré pour cette signification.

Par exemple, considérons les sous étiquettes de "langage" 'zza' (Zaza) et 'diq' (Dimli). Il se trouve que 'zza' est un langage macrole englobant 'diq' et a donc aussi une description dans ISO 639-3 de "Dimli". Cette description a été éditée pour se lire "Dimli (langage macrole)" dans l'enregistrement du registre pour 'zza' pour empêcher une collision.

À l'opposé, les sous étiquettes 'he' et 'iw' partagent une valeur de "Description" de "Hébreux" ; ceci est permis parce que 'iw' est déconseillé et sa "Valeur préférée" est 'he'.

Pour les champs de type "language", le premier champ "Description" qui apparaît dans le registre correspond chaque fois que possible au nom de référence alloué par ISO 639-3. Cela facilite les références croisées entre ISO 639 et le registre.

Quand on crée ou met à jour un enregistrement à cause de l'action d'une des sources standard, le réviseur de sous étiquettes de langage PEUT éditer des descriptions pour corriger des irrégularités de formatage (comme une faute d'orthographe, des apostrophes inappropriées ou autres fautes de ponctuation, ou d'espaces excessives ou manquantes) avant de soumettre l'enregistrement proposé à la liste [ietf-languages@iana.org](mailto:ietf-languages@iana.org) pour examen.

### 3.1.6 Champ "Deprecated"

Le champ "Deprecated" contient la date à laquelle l'enregistrement a été déconseillé et PEUT être ajouté, changé, ou supprimé de tout enregistrement via le processus de maintenance décrit au paragraphe 3.3 ou via le processus d'enregistrement décrit au paragraphe 3.5. Généralement, l'ajout d'un champ "Deprecated" est dû à l'action d'un des organismes de normalisation, comme ISO 3166, qui retire un code. Bien que valides dans les étiquettes de langage, les sous étiquettes et étiquettes avec un champ "Deprecated" sont déconseillées, et les processeurs de validation NE DEVRAIENT PAS générer ces sous étiquettes. Noter qu'un enregistrement qui contient un champ "Deprecated" et pas de champ "Valeur préférée" correspondant n'a pas de transposition de remplacement.

Dans certains cas historiques, il se pourrait qu'il n'ait pas été possible de reconstruire la date originale de dépréciation. Dans ces cas, une date approximative apparaît dans le registre. Certaines sous étiquettes et certaines étiquettes faisant l'objet d'une clause d'antériorité ou redondantes ont été déconseillées avant la création initiale du registre. Les règles exactes pour elles apparaissent à la Section 2 de la [RFC4645]. Noter que ces enregistrements ont un champ "Deprecated" avec une date antérieure au champ "Added" correspondant !

### 3.1.7 Champ "Preferred-Value"

Le champ "Valeur préférée" contient une transposition entre l'enregistrement dans lequel il apparaît et une autre étiquette ou sous étiquette (selon le type de l'enregistrement). La valeur de ce champ est utilisée pour la canonisation (voir le paragraphe 4.5). Dans les cas où la sous étiquette ou étiquette a aussi un champ "Deprecated", alors la "Valeur préférée" est RECOMMANDÉE comme meilleur choix pour représenter la valeur de cet enregistrement lors du choix d'une étiquette de langage.

Les enregistrements qui contiennent une "Valeur préférée" entrent dans un des quatre groupes suivants :

1. Codes de langage ISO 639 qui ont ensuite été supprimés en faveur d'autres codes. Ces valeurs sont pour la plupart des curiosités historiques. L'appariement 'he'/'iw' cité ci-dessus en est un exemple.
2. Des sous étiquettes (avec des types autres que langage ou extlang) tirées de codes ou valeurs qui ont été retirées en faveur d'un nouveau code. En particulier, cela s'applique aux sous étiquettes de région tirées de ISO 3166-1, parce que parfois un pays va changer son nom ou son administration d'une façon telle que cela entraîne un nouveau code de région. Dans certains cas, les pays sont revenus à un ancien nom, qui pourrait avoir été déjà codé. Par exemple, la sous étiquette 'ZR' (Zaire) a été remplacée par la sous étiquette 'CD' (République démocratique du Congo) quand le nom de ce pays a été changé.
3. Les étiquettes ou sous étiquettes qui sont devenues obsolètes parce que les valeurs qu'elles représentent ont été codées ultérieurement. Beaucoup des étiquettes qui faisaient l'objet d'une clause d'antériorité ou étaient redondantes ont été ensuite codées par ISO 639, par exemple, et entrent dans ce groupe. Par exemple, "i-klinton" a été déconseillé quand la sous étiquette 'tlh' a été ajoutée. L'enregistrement pour "i-klinton" a une "Valeur préférée" de 'tlh'.
4. Les sous étiquettes de langage étendues ont toujours une transposition identique à leur sous étiquette de langage principale. Par exemple, la sous étiquette de langage étendue 'yue' (cantonais) peut être utilisée pour former l'étiquette "zh-yue". Elle a une transposition de "Valeur préférée" en la sous étiquette de langage principal 'yue', signifiant qu'une étiquette comme "zh-yue-Hant-HK" peut être canonisée en "yue-Hant-HK".

Les enregistrements autres que ceux de type "extlang" qui contiennent un champ "Valeur préférée" DOIVENT aussi avoir un champ "Deprecated". Ce champ contient la date à laquelle l'étiquette ou sous étiquette a été déconseillée en faveur de la valeur préférée.

Pour les enregistrements de type "extlang", le champ "Valeur préférée" apparaît sans un champ "Deprecated" correspondant. Une mise en œuvre PEUT ignorer ces transpositions de valeur préférée, mais si elle ignore la transposition, elle DEVRAIT le faire de façon cohérente. Elle DEVRAIT aussi traiter la "Valeur préférée" comme équivalente à l'élément transposé. Par exemple, les étiquettes "zh-yue-Hant-HK" et "yue-Hant-HK" sont sémantiquement équivalentes et devraient être traitées comme si elles étaient la même étiquette.

Occasionnellement, le code déconseillé est préféré dans certains contextes. Par exemple, "iw" et "he" peuvent toutes deux être utilisées dans le langage de programmation Java, mais "he" est converti en entrée à "iw", qui est donc la forme canonique dans Java.

Les transpositions de "Valeur préférée" dans les enregistrements de type "region" ne représentent pas parfois exactement la même signification que la valeur originale. Il y a de nombreuses raisons pour changer un code de pays, et l'effet que cela a sur la formation des étiquettes de langage va dépendre de la nature du changement en question. Par exemple, la sous étiquette de région 'YD' (Yémen démocratique) a été déconseillée en faveur de la sous étiquette 'YE' (Yémen) quand ces deux pays ont fusionné en 1990.

Une "Valeur préférée" PEUT être ajoutée, changée, ou supprimée des enregistrements en accord avec les règles du paragraphe 3.3. L'ajout, modification, ou suppression d'un champ "Valeur préférée" dans un enregistrement n'implique pas que le contenu qui utilise la sous étiquette affectée doit être ré-étiqueté.

Les champs "Valeur préférée" dans les enregistrements de type "grandfathered" et "redundant" contiennent chacun une "gamme étendue de langages" [RFC4647] dont l'utilisation est fortement RECOMMANDÉE à la place de la valeur de l'enregistrement. Dans de nombreux cas, ces transpositions ont été créées via la dépréciation des étiquettes durant la période avant l'adoption de la [RFC4646]. Par exemple, l'étiquette "no-nyn" a été déconseillée en faveur du code de langage "nn" défini par la norme ISO 639-1.

Le champ "Valeur préférée" dans les enregistrements de sous étiquette de type "extlang" contient aussi une "gamme étendue de langages". Cela permet que la sous étiquette soit déconseillée en faveur d'une seule sous étiquette de langage principal ou d'une nouvelle séquence "extlang" de langages.

Généralement, l'ajout, la suppression, ou le changement d'un champ "Valeur préférée" pour une sous étiquette est fait pour refléter des changements dans une des sources standard. Par exemple, si un code de région ISO 3166-1 est déconseillé en faveur d'un autre code, cela DEVRAIT résulter en l'ajout d'un champ "Valeur préférée".

Les changements à une sous étiquette peuvent affecter aussi d'autres sous étiquettes : quand il propose des changements au registre, le réviseur de sous étiquettes de langage DOIT revoir le registre à la recherche de tels effets et proposer les changements nécessaires en utilisant le processus du paragraphe 3.5, mais n'importe qui PEUT demander de tels changements. Par exemple :

Supposons que la sous étiquette 'XX' ait une "Valeur préférée" de 'YY'. Si 'YY' change plus tard pour avoir une "Valeur préférée" de 'ZZ', alors la "Valeur préférée" pour 'XX' DOIT aussi changer pour être 'ZZ'.

Supposons qu'une sous étiquette de langage enregistrée 'dialect' représente un langage non encore disponible dans ISO 639.

L'ajout ultérieur d'un code de langage correspondant dans ISO 639 DEVRAIT résulter en l'ajout d'une "Valeur préférée" pour 'dialect'.

### 3.1.8 Champ "Prefix"

Le champ "Prefix" contient une étiquette de langage valide qui est RECOMMANDÉE comme préfixe possible de cette sous étiquette d'enregistrement, peut-être avec d'autres sous étiquettes. C'est-à-dire, quand on inclut un langage étendu ou une sous étiquette de variante qui a au moins un "Prefix" dans une étiquette de langage, l'étiquette résultante DEVRAIT correspondre à au moins un des champs "Prefix" de la sous étiquette en utilisant l'algorithme "Filtrage étendu" (voir la [RFC4647]) et chacune des sous étiquettes dans ce "Prefix" DEVRAIT apparaître avant la sous étiquette elle-même.

Le champ "Prefix" DOIT apparaître exactement une fois dans un enregistrement de type "extlang". Le champ "Prefix" PEUT apparaître plusieurs fois (ou pas du tout) dans les enregistrements de type "variant". Des champs supplémentaires de ce type PEUVENT être ajoutés à un enregistrement "variant" via le processus d'enregistrement, pourvu que l'enregistrement "variant" ait déjà au moins un champ "Prefix".

Chaque champ "Prefix" indique une séquence particulière de sous étiquettes qui forme une étiquette significative avec cette sous étiquette. Par exemple, la sous étiquette de langage étendue 'cmn' (chinois mandarin) n'a de sens qu'avec son préfixe 'zh' (chinois). De même, 'rozaj' (résien, un dialecte slovène) serait approprié quand il est utilisé avec son préfixe 'sl' (slovène) tandis que des étiquettes comme "est-1994" ne sont pas appropriées (et probablement sans signification). Bien que le "Prefix" pour 'rozaj' soit "sl", d'autres sous étiquettes pourrait apparaître entre elles. Par exemple, l'étiquette "sl-IT-rozaj" (slovène, Italie, résien) correspond au "Prefix" "sl".

Le "Prefix" indique aussi quand des sous étiquettes de variante ont un sens utilisées ensemble (beaucoup de celles qui par ailleurs partagent un "Prefix" sont mutuellement exclusives) et ce que l'ordre relatif des variantes est supposé être. Par exemple, la variante '1994' (orthographe résienne normalisée) a plusieurs champs "Prefix" dans le registre ("sl-rozaj", "sl-rozaj-biske", "sl-rozaj-njiva", "sl-rozaj-osojs", et "sl-rozaj-solba"). Cela indique non seulement que "1994" est approprié avec chacune de ces cinq sous étiquettes de variantes résiennes ('rozaj', 'biske', 'njiva', 'osojs', et 'solba') mais aussi qu'il DEVRAIT apparaître à la suite de chacune de ces variantes dans une étiquette. Donc, l'étiquette de langage devrait prendre la forme "sl-rozaj-biske-1994", plutôt que "sl-1994-rozaj-biske" ou "sl-rozaj-1994-biske".

Si un enregistrement n'inclut pas de champ "Prefix", un champ "Prefix" NE DOIT PAS être ajouté à l'enregistrement à une date ultérieure. Autrement, les changements (ajouts, suppressions, ou modifications) à l'ensemble de champs "Prefix" PEUVENT être enregistrés, pour autant qu'ils élargissent strictement la gamme des étiquettes de langage qui sont recommandées. Par exemple, un "Prefix" avec la valeur "be-Latn" (biélorusse, écriture latine) pourrait être remplacé par la valeur "be" (biélorusse) mais pas par la valeur "ru-Latn" (Russe, écriture latine) ou la valeur "be-Latn-BY" (biélorusse, écriture latine, Biélorussie) ces dernières changent ou rétrécissent la gamme des étiquettes suggérées.

Le corps de champ du champ "Prefix" NE DOIT PAS être en conflit avec un "Prefix" déjà enregistré pour un enregistrement donné. Un tel conflit se produirait quand aucune étiquette valide ne pourrait être construite qui contiendrait le préfixe, comme quand deux sous étiquettes ont chacune un "Prefix" qui contient les autres sous étiquettes. Par exemple, supposons que la sous étiquette "avariant" ait le préfixe "es-bvariant". Alors la sous étiquette "bvariant" ne peut pas recevoir le préfixe "avariant", parce que cela exigerait une étiquette de la forme "es-avariant-bvariant-avariant", qui ne serait pas valide.

### 3.1.9 Champ "Suppress-Script"

Le champ "Suppress-Script" contient une sous étiquette de script (dont l'enregistrement apparaît dans le registre). Le champ "Suppress-Script" DOIT apparaître seulement dans les enregistrements dont le corps de champ "Type" est "language" ou "extlang". Ce champ NE DOIT PAS apparaître plus d'une fois dans un enregistrement.

Ce champ indique une écriture utilisée pour rédiger l'immense majorité des documents dans le langage donné. La sous étiquette pour une telle écriture n'ajoute aucune information distinctive à une étiquette de langage et donc NE DEVRAIT PAS être utilisée pour la plupart des documents dans ce langage. Omettre la sous étiquette de script indiquée par ce champ aide à assurer une plus grande compatibilité entre les étiquettes de langage générées en accord avec les règles du document et les étiquettes de langage et processeurs ou consommateurs d'étiquettes sur la base de la RFC 3066. Par exemple, virtuellement tous les documents islandais sont écrits en écriture latine, rendant la sous étiquette "Latn" redondante dans l'étiquette "est-Latn".

De nombreux enregistrements de sous étiquette de langage n'ont pas de champ "Suppress-Script". L'absence d'un champ "Suppress-Script" pourrait indiquer que le langage est coutumièrement écrit dans plus d'une écriture ou que le langage n'est coutumièrement pas écrit du tout. Cela pourrait aussi signifier que des informations suffisantes n'étaient pas disponibles quand l'enregistrement a été créé et donc reste candidat pour un enregistrement futur.

### 3.1.10 Champ "Macrole langage"

Le champ "Macrole langage" contient une sous étiquette de langage principale (dont l'enregistrement apparaît dans le registre). Ce champ indique un langage qui englobe le langage de cette sous étiquette en accord avec les allocations faites dans ISO 639-3.

ISO 639-3 étiquette certaines des langues dans le registre comme des "macrole languages". ISO 639-3 définit le terme "macrole language" comme signifiant des "grappes de variétés de langage en étroite relation qui [...] peuvent être considérés comme des langages individuels distincts, bien que dans certains contextes d'usage une seule identité de langage soit nécessaire pour toutes". Cela correspond aux codes enregistrés dans ISO 639-2 comme des langages individuels qui ont été trouvés correspondre à plus d'une langue dans ISO 639-3.

Un langage contenu dans un langage macrole est appelé un "langage englobé". L'enregistrement pour chaque langage englobé contient un champ "Macrole langage" dans le registre ; les langages macrole eux-mêmes ne sont pas spécialement marqués. Noter que certains langages englobés ont des codes ISO 639-1 ou ISO 639-2.

Le champ "Macrole langage" peut seulement se produire dans des enregistrements de type "language" ou "extlang". Seules les valeurs allouées dans ISO 639-3 vont être considérées pour inclusion. Les champs "Macrole langage" PEUVENT être ajoutés ou supprimés via le processus normal d'enregistrement chaque fois que ISO 639-3 définit de nouvelles valeurs ou supprime des anciennes valeurs. Les langages Macrole sont pour information, et PEUVENT être supprimés ou changés si ISO 639-3 change les valeurs. Pour plus d'informations sur l'utilisation de ce champ et le choix entre les sous étiquettes de langage macrole et de langage englobé, voir le paragraphe 4.1.1.

Par exemple, les sous étiquettes de langage 'nb' (Bokmal norvégien) et 'nn' (Nynorsk norvégien) ont chacune un champ "Macrole langage" avec une valeur de 'no' (norvégien). Pour plus d'informations, voir le paragraphe 4.1.

### 3.1.11 Champ "Scope"

Le champ 'Scope' contient des informations de classement sur une sous étiquette de langage principal ou étendu dérivée de la norme ISO 639. La plupart des langages ont une portée de "individual", qui signifie que le langage n'est pas une langue macrole, une collection, un code spécial ou d'utilisation privée. C'est-à-dire, il est ce qu'on pourrait normalement considérer comme étant un "langage". Toute sous étiquette de langage principal ou étendu qui n'a pas de champ "Scope" est un langage individuel.

Les informations de "Scope" peuvent parfois être utiles pour choisir des étiquettes de langage, car elles indiquent l'objet ou la "portée" de l'allocation de code dans ISO 639. Les valeurs disponibles sont :

- o "macrole language" - Indique un langage macrole comme défini dans ISO 639-3 (voir le paragraphe 3.1.10). Un langage macrole est une grappe de langages en étroite relation qui sont parfois considérés comme étant un seul langage.
- o "collection" - Indique une sous étiquette qui représente une collection de langages, normalement reliés par un type d'association historique, géographique, ou linguistique. À la différence d'un langage macrole, une collection peut contenir des langages qui sont seulement vaguement en relation et une collection ne peut pas être utilisée de façon interchangeable avec les langages qui lui appartiennent.
- o "special" - Indique un code de langage spécial. Ce sont des sous étiquettes utilisées pour identifier des attributs linguistiques non associés particulièrement à un langage concret. Cela inclut des codes pour quand le langage est indéterminé ou pour un contenu non linguistique.
- o "private-use" - Indique un code réservé pour utilisation privée dans la norme sous-jacente. Les sous étiquettes avec cette portée peuvent être utilisées pour indiquer le langage principal pour lequel il n'existe pas d'allocation dans ISO 639 ou enregistrée.

Le champ "Scope" PEUT apparaître dans des enregistrements de type "language" ou "extlang". Noter que beaucoup de préfixes pour des sous étiquettes de langage étendues vont avoir une portée de "language macrole" (bien que certaines ne l'aient pas) et que beaucoup de langages qui ont une portée de "macrole language" vont avoir des sous étiquettes de langage étendu associées.

Le champ "Scope" PEUT être ajouté, modifié, ou supprimé via le processus d'enregistrement, pourvu que le changement reflète les changements faits par ISO 639 à la classification des allocations. De tels changements sont supposés être rares.

Par exemple, la sous étiquette de langage principal 'zh' (chinois) a une portée de "macrole language", tandis que son langage englobé 'nan' (chinois Min Nan) a une portée de "individual". La valeur spéciale 'und' (indéterminé) a une portée de "special". La collection ISO 639-5 'gem' (langages germaniques) a une portée de "collection".

### 3.1.12 Champ "Comments"

Le champ "Comments" contient des informations supplémentaires sur l'enregistrement et PEUT apparaître plus d'une fois par enregistrement. Le corps de champ PEUT inclure la gamme complète de caractères Unicode et ne se restreint pas à une écriture particulière. Ce champ PEUT être inséré ou changé via le processus d'enregistrement, et aucune garantie de stabilité n'est fournie.

Le contenu de ce champ n'est pas restreint, sauf par le besoin d'enregistrer les informations, la pertinence de la demande, et par des limitations de taille pratique raisonnables. La principale raison du champ "Comments" est l'identification de la sous étiquette -- pour aider à distinguer la sous étiquette des autres avec lesquelles elle pourrait être confondue comme aide à l'utilisation. De grandes quantités d'informations sur l'utilisation, l'histoire, ou les bases générales d'une sous étiquette sont déconseillées, car elles relèvent généralement d'une demande d'enregistrement plutôt que du registre.

## 3.2 Réviseur de sous étiquettes de langage

Le réviseur de sous étiquettes de langage est le modérateur de la liste de diffusion [ietf-languages@iana.org](mailto:ietf-languages@iana.org), il répond aux demandes d'enregistrement, et effectue les autres tâches de maintenance du registre décrites au paragraphe 3.3. Seul le réviseur de sous étiquette de langage a la permission de demander à l'IANA de changer, mettre à jour, ou ajouter des enregistrements au registre des sous étiquettes de langage. Le réviseur de sous étiquettes de langage PEUT déléguer la modération de la liste et autres tâches d'écriture en tant que de besoin.

Le réviseur de sous étiquettes de langage est appointé par l'IESG pour une durée indéfinie, sous réserve de suppression ou remplacement à la discrétion de l'IESG. L'IESG sollicitera des candidats pour le poste (à l'adoption du présent document ou en cas de vacance) et sollicitera ensuite des retours sur les qualifications des candidats. Les candidats qualifiés devraient être familiarisés avec le BCP 47 et ses exigences, être volontaires pour administrer équitablement de façon responsable et judicieuse le processus d'enregistrement, et être convenablement informés des questions d'identification de langages afin que le réviseur puisse juger des revendications et en tirer des conclusions sur les contributions des experts de langage et des demandeurs de sous étiquettes.

Les performances ou décisions suivantes du réviseur d'étiquettes de langues PEUVENT faire l'objet d'un appel auprès de l'IESG selon les mêmes règles que les autres décisions de l'IETF (voir la [RFC2026]). L'IESG peut revenir ou inverser les décisions du réviseur d'étiquettes de langues, fournir des directives, ou prendre d'autres actions appropriées.

## 3.3 Maintenance du registre

La maintenance du registre exige que, comme les codes sont alloués ou supprimés par les normes ISO 639, ISO 15924, ISO 3166, et UN M.49, le réviseur d'étiquettes de langues DOIT évaluer chaque changement et déterminer l'action appropriée en accord avec les règles du présent document. Ces mises à jour suivent le processus d'enregistrement décrit au paragraphe 3.5. Généralement, le réviseur d'étiquettes de langues va commencer le processus pour le nouvel enregistrement ou sa mise à jour en remplissant le formulaire d'enregistrement et en le soumettant. Si un changement d'une de ces normes a lieu et si le réviseur d'étiquettes de langues ne fait pas cela en temps utile, toute partie intéressée PEUT soumettre le formulaire. Ensuite, le processus d'enregistrement se continue normalement.

Noter que certains enregistrements affectent d'autres sous étiquettes -- peut-être plus d'une -- comme quand une sous étiquette de région est déconseillée en faveur d'une nouvelle valeur. Le réviseur d'étiquettes de langues est responsable de s'assurer que tout changement est enregistré de façon appropriée, avec chaque changement exigeant son propre formulaire d'enregistrement.

Le réviseur d'étiquettes de langues DOIT s'assurer que les nouvelles sous étiquettes satisfont les autres exigences du présent document (et plus spécialement du paragraphe 3.4) ou soumettent un formulaire d'enregistrement approprié pour une sous étiquette de remplacement, comme décrit dans ce paragraphe. Chaque sous étiquette individuelle affectée par un

changement DOIT être envoyée à la liste de diffusion [ietf-languages@iana.org](mailto:ietf-languages@iana.org) avec son propre formulaire d'enregistrement et dans un message distinct.

### 3.4 Stabilité des entrées du registre de l'IANA

La stabilité des entrées et de leur signification dans le registre est critique pour la stabilité à long terme des étiquettes de langage. Les règles de ce paragraphe garantissent que la signification d'une étiquette de langage spécifique est stable dans le temps et ne va pas changer.

Ces règles traitent spécifiquement de la façon dont les changements de codes (incluant la suppression et la dépréciation de codes) tenus par les normes ISO 639, ISO 15924, ISO 3166, et UN M.49 sont reflétés dans le registre IANA des sous étiquettes de langue. Les allocations au registre IANA des sous étiquettes de langue DOIVENT suivre les règles de stabilité suivantes :

1. Les valeurs dans les champs "Type", "Subtag", "Tag", et "Added" NE DOIVENT PAS être changées et leur stabilité est garantie dans le temps.
2. Les valeurs dans les champs "Valeur préférée" et "Deprecated" PEUVENT être ajoutées, altérées, ou supprimées via le processus d'enregistrement. Ces changements DEVRAIENT être limités à ceux nécessaires pour refléter des changements dans une des normes sous-jacentes (ISO 639, ISO 15924, ISO 3166-1, ou UN M.49) et normalement, l'altération ou la suppression d'une "Valeur préférée" est limitée spécifiquement aux codes de région.
3. Les valeurs dans le champ "Description" NE DOIVENT PAS être changées d'une façon qui invaliderait une étiquette existante. La description PEUT être un peu élargie dans sa portée, changée pour ajouter des informations, ou adaptée à l'usage moderne le plus courant. Par exemple, des pays changent occasionnellement leur nom ; un exemple historique est celui de la "Haute Volta" qui s'est changé en "Burkina Faso".
4. Les valeurs dans le champ "Prefix" PEUVENT être ajoutées aux enregistrements de type "variant" existants via le processus d'enregistrement, pourvu que le "variant" ait déjà au moins un "Prefix". Un champ "Prefix" NE DEVRA PAS être enregistré pour un "variant" qui n'a pas un champ "Prefix" existant. Si un préfixe est ajouté à un enregistrement de variante, des champs "Comment" PEUVENT être utilisés pour expliquer les différents usages avec les divers préfixes.
5. Les valeurs dans le champ "Prefix" dans les enregistrements de type "variant" PEUVENT aussi être modifiées, pour autant que les modifications élargissent l'ensemble de préfixes. C'est-à-dire, un préfixe PEUT être remplacé par un de ses propres préfixes. Par exemple, le préfixe "en-US" pourrait être remplacé par "en", mais pas par des préfixes "en-Latn", "fr", ou "en-US-boont". Si une de ces valeurs de préfixe était nécessaire, elle devrait être enregistrée séparément.
6. Les valeurs dans le champ "Prefix" dans les enregistrements de type "extlang" NE DOIVENT PAS être ajoutées, modifiées, ou supprimées.
7. Le champ "Prefix" NE DOIT PAS être supprimé d'un enregistrement dans lequel il apparaît. Ce champ DEVRAIT être inclus dans l'enregistrement initial de tout enregistrement de type "variant" et DOIT être inclus dans tout enregistrement de type "extlang".
8. Le champ "Comments" PEUT être ajouté, changé, modifié, ou supprimé via le processus d'enregistrement ou tout processus ou considération décrit dans ce paragraphe.
9. Le champ "Suppress-Script" PEUT être ajouté ou supprimé via le processus d'enregistrement.
10. Le champ "Macrole language" PEUT être ajouté ou supprimé via le processus d'enregistrement, mais seulement en réponse aux changements faits par ISO 639. Le champ "Macrole language" apparaît chaque fois qu'un langage a un langage macrole correspondant dans ISO 639. C'est-à-dire, les champs "Macrole language" dans le registre correspondent exactement à ceux de ISO 639. Aucune autre transposition de "Macrole language" ne va être soumise à enregistrement.
11. Le champ "Scope" PEUT être ajouté ou supprimé d'une sous étiquette de langage principal ou étendu après l'enregistrement initial, et il PEUT être modifié afin de correspondre à tout changement fait par ISO 639. Les changements du champ "Scope" DOIVENT refléter les changements faits par ISO 639. Noter que les sous étiquettes de langage principal ou étendu dont les enregistrements ne contiennent pas un champ "Scope" (c'est-à-dire, la plupart d'entre elles) sont des langages individuels comme décrit au paragraphe 3.1.11.
12. Les sous étiquettes de langage principal ou étendu (autres que les valeurs enregistrées indépendamment en utilisant le processus d'enregistrement) sont créées en accord avec les allocations des diverses parties de ISO 639, comme suit :

- A. Les codes alloués par ISO 639-1 qui n'entrent pas en conflit avec des sous étiquettes de deux lettres existantes et n'ont pas de sous étiquettes de langage principal de trois lettres correspondantes définies dans le registre, sont entrées dans le registre IANA comme de nouveaux enregistrements de type "language". Noter que les langages qui ont reçu un code dans ISO 639-1 ne peuvent pas recevoir de sous étiquettes de langage étendu, même si elles sont englobées dans un macro langage.
- B. Les codes alloués par ISO 639-3 ou ISO 639-5 qui ne sont pas en conflit avec des sous étiquettes principales de langage de trois lettres existantes et qui n'ont pas de code ISO 639-1 alloué (ou espéré) sont entrés dans le registre de l'IANA comme de nouveaux enregistrements de type "language". Noter que ces deux normes comportent maintenant un sur ensemble de codes ISO 639-2. Les codes qui ont une transposition "macrole language" définie au moment de leur enregistrement DOIVENT contenir un champ "Macrole language".
- C. Les codes alloués par ISO 639-3 PEUVENT aussi être considérés pour un enregistrement de sous étiquette de langage étendu. Noter qu'il DOIT leur être alloué un enregistrement de sous étiquette de langage principal de type "language" même quand un enregistrement "extlang" est proposé. Quand on envisage une allocation de sous étiquette de langage étendu, ces critères s'appliquent :
1. Si un langage a une transposition de langage macrole, et si ce langage macrole a d'autres langages englobés qui ont allouées des sous étiquettes de langage étendu, alors le nouveau langage DEVRAIT avoir aussi alloué un enregistrement "extlang". Par exemple, tout langage avec un langage macrole de 'zh' ou 'ar' va se voir allouer un enregistrement "extlang".
  2. Les enregistrements "extlang" NE DEVRAIENT PAS être créés pour des langages si d'autres langages englobés par le langage macrole n'incluent pas aussi des enregistrements "extlang". Par exemple, si un nouveau langage serbo-croate ('sh') était enregistré, il n'aurait pas d'enregistrement "extlang" parce que d'autres langages englobés, comme le serbe ('sr') n'en incluent pas dans le registre.
  3. Les langages de signes DEVRAIENT avoir un enregistrement "extlang" avec un préfixe de "sgn".
  4. Les enregistrements "extlang" NE DOIVENT PAS être créés pour des éléments déjà dans le registre. Les sous étiquettes de langage étendu vont seulement être considérées au moment de l'enregistrement initial.
  5. Les enregistrements de sous étiquette de langage étendu DOIVENT inclure les champs "Prefix" et "Valeur préférée" avec les valeurs de champ allouées comme décrit au paragraphe 2.2.2.
- D. Tous les autres codes alloués par ISO 639-2 qui ne sont pas en conflit avec les sous étiquettes de langage principal ou étendu de trois lettres existantes et qui n'ont pas de code ISO 639-1 à deux lettres alloué sont entrées dans le registre IANA comme de nouveaux enregistrements de type "language". Ce type d'enregistrement n'est pas supposé se produire à l'avenir.
13. Les codes alloués par ISO 15924 et ISO 3166-1 qui ne sont pas en conflit avec les sous étiquettes existantes du type associé et dont la signification n'est pas la même que celle d'une sous étiquette existante du même type sont entrés dans le registre IANA comme de nouveaux enregistrements.
14. Les codes alloués par ISO 639, ISO 15924, ou ISO 3166-1 qui sont retirés par leurs autorités respectives de maintenance ou d'enregistrement restent valides dans les étiquettes de langage. Un champ "Deprecated" contenant la date de retrait DOIT être ajouté à l'enregistrement. Si un nouvel enregistrement de même type est ajouté représentant une valeur de remplacement, alors un champ "Valeur préférée" PEUT aussi être ajouté. Le processus d'enregistrement PEUT être utilisé pour ajouter des commentaires sur le retrait du code par les normes respectives. Par exemple, le code de région 'TL' a été alloués au pays "Timor-Leste", remplaçant le code 'TP' (qui était alloué au "Timor oriental" quand il était sous administration du Portugal). La sous étiquette 'TP' reste valide dans les étiquettes de langage, mais son enregistrement contient la "Valeur préférée" de 'TL' et son champ "Deprecated" contient la date à laquelle le nouveau code a été alloué ("2004-07-06").
15. Les codes alloués par ISO 639, ISO 15924, ou ISO 3166-1 qui sont en conflit avec les sous étiquettes existantes du type associé, y compris les sous étiquettes qui sont déconseillées, NE DOIVENT PAS être entrés dans le registre. Les considérations supplémentaires suivantes s'appliquent aux valeurs de sous étiquette qui sont réallouées :
- A. Pour les codes ISO 639, si la signification du nouveau code alloué n'est pas représentée par une sous étiquette dans le registre IANA, le réviseur de sous étiquettes de langue, comme décrit au paragraphe 3.5, DEVRA préparer une proposition d'entrée dans le registre IANA, aussitôt que praticable, d'une sous étiquette de langage enregistrée comme valeur de remplacement pour le nouveau code. La forme de la sous étiquette de langage enregistrée va être à la discrétion du réviseur de sous étiquettes de langue et DOIT se conformer aux autres restrictions sur les sous étiquettes de langage dans le présent document.
- B. Pour toutes les sous étiquettes dont la signification est dérivée d'une norme externe (c'est-à-dire, par ISO 639, ISO 15924, ISO 3166-1, ou UN M.49) si une nouvelle signification est attribuée à un code existant et si la nouvelle signification élargit celle de ce code, alors la signification de la sous étiquette associée PEUT être changée pour y correspondre. La signification d'une sous étiquette NE DOIT cependant PAS être rétrécie, car il pourrait en résulter qu'une proportion inconnue des utilisations existantes d'une sous étiquette devienne invalide. Noter que l'autorité d'enregistrement de ISO 639 a adopté une politique de stabilité similaire.
- C. Pour les codes ISO 15924, si la signification du nouveau code alloué n'est pas représentée par une sous étiquette dans le registre IANA, le réviseur de sous étiquettes de langue, comme décrit au paragraphe 3.5, DEVRA préparer une proposition pour l'entrée dans le registre IANA, aussitôt que praticable, d'une variante enregistrée de sous

étiquette comme valeur de remplacement pour le nouveau code. La forme de la variante enregistrée de sous étiquette sera à la discrétion du réviseur de sous étiquettes de langue et DOIT se conformer aux autres restrictions sur les sous étiquettes de variantes dans le présent document.

- D. Pour les codes ISO 3166-1, si la signification du nouveau code alloué est associée au même code UN M.49 comme une autre sous étiquette de "region", alors la sous étiquette de région existante reste la valeur préférée pour cette région et aucune nouvelle entrée n'est créée. Un commentaire PEUT être ajouté à la sous étiquette de région existante indiquant la relation au nouveau code ISO 3166-1.
- E. Pour les codes ISO 3166-1, si la signification du nouveau code alloué est associée à un code UN M.49 qui n'est pas représenté par une sous étiquette de région existante, alors le réviseur de sous étiquettes de langue, comme décrit au paragraphe 3.5, DEVRA préparer une proposition d'entrée du code de pays UN M.49 approprié comme entrée dans le registre IANA.
- F. Pour les codes ISO 3166-1, si il n'y a pas de code numérique des Nations Unies associé, alors le réviseur de sous étiquettes de langue DEVRA demander aux Nations Unies d'en créer un. Si il n'y a pas de réponse des Nations Unies dans les 90 jours de l'envoi de la demande, le réviseur de sous étiquettes de langue DEVRA préparer une proposition pour l'entrée dans le registre IANA, aussitôt que praticable, d'une sous étiquette de variante enregistrée comme valeur de remplacement pour le nouveau code. La forme de la sous étiquette de variante enregistrée va être à la discrétion du réviseur de sous étiquettes de langue et DOIT se conformer aux autres restrictions sur les sous étiquettes de variante du présent document. Cette situation a très peu de chances de jamais se produire.
16. Le répertoire UN M.49 a des codes à la fois pour des "pays et zones" (comme "276" pour Allemagne) et "régions et sous régions géographiques" (comme "150" pour Europe). Les codes de pays ou zone de UN M.49 pour lesquels il n'y a pas de code ISO 3166-1 correspondant NE DOIVENT PAS être enregistrés, sauf comme remplacement d'un code ISO 3166-1 qui est bloqué à l'enregistrement par une sous étiquette existante. Si un tel code devient nécessaire, alors l'agence de maintenance de ISO 3166-1 DEVRA d'abord être sollicitée pour allouer un code à la région. Si la sollicitation d'allocation de code à ISO 3166-1 est refusée ou n'a pas de réponse en temps utile, le processus d'enregistrement décrit au paragraphe 3.5 peut alors être utilisé pour enregistrer le code UN M.49 correspondant. De cette façon, les codes UN M.49 restent disponibles comme valeur en dernier ressort dans les cas où ISO 3166-1 réalloue une valeur déconseillée dans le registre.
17. Les entrées redondantes et faisant l'objet d'une clause d'antériorité forment ensemble la liste complète des étiquettes enregistrées sous la [RFC3066]. Les étiquettes redondantes sont celles précédemment enregistrées qui peuvent maintenant être formées en utilisant les sous étiquettes définies dans le registre. Les entrées faisant l'objet d'une clause d'antériorité incluent celles qui ne peuvent jamais être légales parce que elles sont "irregular" (c'est-à-dire, elle ne correspondent pas à la production "langtag" dans la Figure 1) sont limitées par la règle (des sous étiquettes comme 'nyn' et 'min' ressemblent à la production "extlang", mais ne peuvent pas être enregistrées comme des sous étiquettes de langage étendu) ou dont les sous étiquettes sont inappropriées pour l'enregistrement. Toutes les étiquettes faisant l'objet d'une clause d'antériorité sont dans la liste des productions "regular" ou "irregular" dans l'ABNF. Sous la [RFC4646] il était possible à des étiquettes faisant l'objet d'une clause d'antériorité de devenir redondantes. Cependant, toutes les étiquettes pour lesquelles c'était possible sont devenues redondantes avant la production du présent document. Donc, l'ensemble des étiquettes redondantes et faisant l'objet d'une clause d'antériorité est maintenant permanent et immuable : de nouvelles entrées de l'un ou l'autre type NE DOIVENT PAS être ajoutées et les entrées existantes NE DOIVENT PAS être supprimées. Les processus de prise de décision sur quelles étiquettes étaient initialement l'objet d'une clause d'antériorité et quelles ont été rendues redondantes est décrit dans la [RFC4645]. Beaucoup des étiquettes faisant l'objet d'une clause d'antériorité sont maintenant déconseillées -- bien sûr, elles étaient déconseillées même avant la [RFC4646]. Par exemple, l'étiquette "art-lojban" était déconseillée en faveur de la sous étiquette de langage principal 'jbo'. Ces étiquettes pourraient avoir été rendues "redondantes" par l'enregistrement de certaines de leurs sous étiquettes comme des "variantes". Les sous étiquettes "variant-like" dans les enregistrements de clause d'antériorité NE DEVRONT PAS être enregistrées à l'avenir, même avec une signification similaire ou identique.

### 3.5 Procédure d'enregistrement des sous étiquettes

La procédure indiquée ici DOIT être utilisée par tous ceux qui veulent utiliser une sous étiquette non actuellement dans le registre IANA des sous étiquettes de langue ou qui souhaitent ajouter, modifier, mettre à jour, ou supprimer des informations dans des enregistrements existants comme permis par le présent document.

Seules les sous étiquettes du type " language" et "variant" vont être considérées pour un enregistrement indépendant de nouvelles sous étiquettes. Les sous étiquettes nécessaires pour la stabilité et les sous étiquettes nécessaires pour garder le registre synchronisé avec ISO 639, ISO 15924, ISO 3166, et UN M.49 dans les limites définies par le présent document utilisent aussi ce processus, comme décrit au paragraphe 3.3 et sous réserve des dispositions pour la stabilité comme décrit au paragraphe 3.4.

Les demandes d'enregistrement sont acceptées à l'égard des informations dans les champs "Comments", "Deprecated", "Description", "Prefix", "Valeur préférée", "Macrole language", ou "Suppress-Script" dans un enregistrement d'une sous

étiquette, comme décrit au paragraphe 3.4. Les changements à tous les autres champs dans le registre IANA NE sont PAS permis.

L'enregistrement d'une nouvelle sous étiquette ou la demande de modifications à une étiquette ou sous étiquette existante commence par le remplissage par le demandeur du formulaire d'enregistrement reproduit ci-dessous. Noter que chaque réponse n'est pas limitée en taille de sorte que la demande peut adéquatement décrire l'enregistrement. Les champs dans la section "Enregistrement demandé" doivent respecter les exigences du paragraphe 3.1 avant que l'enregistrement puisse être approuvé.

Formulaire d'enregistrement de sous étiquette de langage

1. Nom du demandeur :
2. Adresse de messagerie du demandeur :
3. Enregistrement demandé :
  - Type :
  - Sous étiquette :
  - Description :
  - Préfixe :
  - Valeur préférée :
  - Déconseillée :
  - Écriture supprimée :
  - Langage macrole :
  - Commentaires :
4. Signification prévue de la sous étiquette :
5. Référence à la description publiée du langage (livre ou article) :
6. Autres informations pertinentes :

Des exemples de formulaires d'enregistrement complétés se trouvent dans l'Appendice B. Une liste complète des formulaires d'enregistrement approuvés est en ligne à <http://www.iana.org> ; les lecteurs devraient noter que le registre des étiquettes de langues est maintenant obsolète et devraient à la place chercher le registre des sous étiquettes de langue.

Le formulaire d'enregistrement de sous étiquette DOIT être envoyé à <ietf-languages@iana.org>. Les demandes d'enregistrement subissent une période de révision de deux semaines avant d'être approuvées et soumises à l'IANA pour inclusion dans le registre. Si des modifications sont faites à la demande durant le cours du processus d'enregistrement (comme des corrections pour satisfaire aux exigences du paragraphe 3.1 ou pour rendre les champs "Description" uniques pour le type d'enregistrement donné) le formulaire modifié DOIT aussi être envoyé à <ietf-languages@iana.org> au moins une semaine avant la soumission à l'IANA.

La liste de langages ietf-le est une liste ouverte et peut être jointe en envoyant une demande à <ietf-languages-request@iana.org>. La liste peut être hébergée par l'IANA ou par tout tiers à la demande de l'IESG.

Avant de transmettre un enregistrement à l'IANA, le réviseur de sous étiquettes de langue DOIT s'assurer que toutes les exigences du présent document sont satisfaites. Cela inclut de s'assurer que les valeurs dans le champ "Subtag" correspondent en casse en accord avec la description du paragraphe 3.1.4 et que les champs "Description" sont uniques pour le type d'enregistrement donné comme décrit au paragraphe 3.1.5. Le réviseur DOIT aussi s'assurer qu'un enregistrement File-Date approprié est inclus dans la demande, pour aider l'IANA quand elle met à jour le registre (voir le paragraphe 5.1).

Certains champs dans le formulaire d'enregistrement ainsi que dans l'enregistrement du registre lui-même permettent l'utilisation de caractères non ASCII. Les demandes d'enregistrement DEVRAIENT utiliser le codage UTF-8 pour la cohérence et la clarté. Cependant, comme certains clients de messagerie ne prennent pas en charge ce codage, d'autres codages PEUVENT être utilisés pour la demande d'enregistrement. Le réviseur des sous étiquettes de langues est chargé de s'assurer que les caractères Unicode appropriés apparaissent à la fois dans le formulaire de demande archivé et dans l'enregistrement du registre. Dans le cas d'une erreur de transcription ou de codage par l'IANA, le réviseur de sous étiquettes de langue demandera que le registre soit réparé, en fournissant toutes les informations nécessaires pour aider l'IANA.

Les sous étiquettes de langage étendu (type "extlang") sont par définition toujours englobées par un autre langage. Tous les enregistrements de type "extlang" DOIVENT donc contenir un champ "Prefix" au moment de l'enregistrement. Ce champ "Prefix" ne peut jamais être altéré ou supprimé, et les demandes de le faire DOIVENT être rejetées.

Les sous étiquettes de variante sont généralement enregistrées pour être utilisées avec une gamme particulière d'étiquettes de langage, et les sous étiquettes de variante fondées sur la terminologie du langage auquel elles sont appliquées sont encouragées. Par exemple, la sous étiquette 'rozaĵ' (résien) est destinée à être utilisée avec des étiquettes de langage qui commencent par la sous étiquette de langage principal "sl" (slovène) car le résien est un dialecte de slovène. Donc, la sous

étiquette 'rozaj' serait appropriée dans des étiquettes comme "sl-Latn-rozaj" ou "sl-IT-rozaj". Cette information est mémorisée dans le champ "Prefix" dans le registre. Les demandes d'enregistrement de variante DEVRAIENT inclure au moins un champ "Prefix" dans le formulaire d'enregistrement.

Les demandes d'allocation d'un enregistrement additionnel d'un type donné avec une valeur de sous étiquette existante DOIVENT être rejetées. Par exemple, la sous étiquette de variante 'rozaj' existe déjà dans le registre, donc ajouter un second enregistrement de type "variant" à la sous étiquette 'rozaj' est interdit.

Le champ "Prefix" pour une sous étiquette de variante enregistrée donnée existe dans le registre IANA comme un guide d'utilisation. Des champs "Prefix" supplémentaires PEUVENT être ajoutés en remplissant un formulaire d'enregistrement supplémentaire. Sous cette forme, le champ "Toutes autres informations pertinentes :." DOIT indiquer qu'il est l'ajout d'un préfixe.

Les demandes d'ajout d'un champ "Prefix" à une sous étiquette de variante qui implique une signification différente DEVRAIT être rejetées. Par exemple, une demande d'ajouter le préfixe "de" à la sous étiquette '1994' afin que l'étiquette "de-1994" représente un dialecte germanique ou une forme orthographique devrait être rejetée. La sous étiquette '1994' représente une orthographe slovène particulière, et l'enregistrement supplémentaire changerait ou brouillerait la signification sémantique allouée à la sous étiquette. Une sous étiquette distincte DEVRAIT être proposée à la place.

Les demandes d'ajout d'un "Prefix" à une sous étiquette de variante qui n'a pas de champ "Prefix" actuel DOIVENT être rejetées. Les variantes sont enregistrées sans préfixe parce que elles sont potentiellement utiles avec de nombreux ou même tous les langages. Ajouter un ou plusieurs champs "Prefix" serait potentiellement dommageable pour l'utilisation de la variante, car cela réduit considérablement la portée de la sous étiquette (ce qui n'est pas permis selon les règles de stabilité (paragraphe 3.4) par opposition à l'élargissement de portée de sous étiquette, qui est ce que fait normalement l'ajout d'un "Prefix". Un exemple d'une telle variante "no-prefix" est la sous étiquette 'fonipa', qui représente l'alphabet phonétique international, un schéma qui peut être utilisé pour transcrire de nombreux langages.

Les champs "Description" fournis dans la demande DOIVENT contenir au moins une description écrite ou transcrite en écriture latine ; la demande PEUT aussi inclure des champs "Description" supplémentaires dans n'importe quelle écriture ou langage. Le champ "Description" est utilisé à des fins d'identification et ne représente pas nécessairement le nom natif réel du langage ou de sa variante. Il n'a pas non plus besoin d'être dans un langage particulier, mais DEVRAIT être à la fois convenable et suffisant pour identifier l'élément dans l'enregistrement. Le réviseur de sous étiquettes de langue va vérifier et éditer tous les champs "Description" proposés pour s'assurer de leur unicité et prévenir des collisions avec les champs "Description" dans d'autres enregistrements de même type. Si cela arrive dans une demande d'enregistrement indépendante, le réviseur de sous étiquettes de langue DOIT resoumettre l'enregistrement à la liste <ietf-languages@iana.org>, la traitant comme une modification d'une demande en discussion, comme décrit au paragraphe 3.5, sauf si le seul objet de la demande est d'introduire un champ "Description" dupliqué, auquel cas la demande DEVRA être rejetée.

Il n'est pas garanti que le champ "Description" soit stable. Des corrections ou précision d'intention sont des exemples de changements possibles. Des tentatives de fournir des traductions ou transcriptions des entrées du registre (qui, par définition, ne fournissent pas de nouvelles informations) ont peu de chances d'être approuvées.

Peu après l'achèvement de la période de révision de deux semaines, le réviseur de sous étiquettes de langue DOIT effectuer une des actions suivantes :

- o Explicitement accepter la demande et transmettre le formulaire contenant l'enregistrement à insérer ou modifier à <iana@iana.org> en accord avec la procédure décrite au paragraphe 3.3.
- o Explicitement rejeter la demande parce que des objections significatives ont été soulevées sur la liste ou à cause de problèmes avec les contraintes du présent document (qui DOIVENT être explicitement citées).
- o Étendre la période de révision en accordant une augmentation supplémentaire de deux semaines pour permettre plus de discussion. Après chaque incrément de deux semaines, le réviseur de sous étiquettes de langue DOIT indiquer sur la liste si l'enregistrement a été accepté, rejeté, ou étendu.

Noter que le réviseur de sous étiquettes de langue PEUT soulever des objections sur la liste si il ou elle le désire. La chose importante est que l'objection DOIT être rendue publique.

Parfois la demande a besoin d'être modifiée par suite de la discussion durant la période de révision ou à cause des exigences du présent document. Le demandeur, le réviseur des sous étiquettes de langue, ou d'autres PEUVENT soumettre une version modifiée du formulaire d'enregistrement complété, qui va être considéré au lieu de la demande d'origine avec l'approbation explicite du demandeur. De tels changements ne redémarrent pas la période de discussion de deux semaines, bien qu'une demande contenant l'enregistrement final soumis à l'IANA DOIVE apparaître sur la liste au moins une semaine avant que le réviseur de sous étiquettes de langue transmette l'enregistrement à l'IANA. Le demandeur PEUT modifier une demande rejetée avec des informations plus appropriées ou supplémentaires et la soumettre à nouveau ; cela démarre une nouvelle période de commentaires de deux semaines.

Les enregistrements initiés du fait des dispositions des paragraphes 3.3 ou 3.4 NE DEVRONT PAS être rejetés ensemble (car ils doivent finalement apparaître dans le registre) et DEVRAIENT être complétés aussi rapidement que possible. Le processus de révision permet aux membres de la liste de commenter sur les informations spécifiques du formulaire et l'enregistrement qu'il contient et donc d'aider à s'assurer qu'il est correct et cohérent. Le réviseur de sous étiquettes de langue PEUT rejeter une version spécifique du formulaire, mais il DOIT proposer un remplacement convenable, étendant la période de révision comme décrit ci-dessus, jusqu'à ce que le formulaire soit dans un format permettant l'approbation du réviseur et trouve un consensus général sur la liste.

Les décisions prises par le réviseur de sous étiquettes de langue PEUVENT faire l'objet d'un appel à l'IESG [RFC2028] selon les mêmes règles que les autres décisions de l'IETF [RFC2026]. Cela inclut la décision d'étendre la période de révision ou la défaillance à annoncer une décision d'une manière claire et en temps utile.

Les enregistrements approuvés apparaissent dans le registre des sous étiquettes de langue. Les formulaires d'enregistrement approuvés sont disponibles en ligne à <http://www.iana.org>.

Les mises à jour ou changements aux enregistrements existants suivent la même procédure que les nouveaux enregistrements. Le réviseur des sous étiquettes de langue décide si il y a un consensus pour mettre à jour l'enregistrement suite à la période de revue de deux semaines ; normalement, les objections du demandeur original vont avoir un poids supplémentaire dans la formation d'un tel consensus.

Les enregistrements sont permanents et stables. Une fois enregistrées, les sous étiquettes ne vont pas être supprimées du registre et vont rester une façon valide de spécifier un langage ou variante spécifique.

Note : L'objet de la section "Référence à une description publiée" dans le formulaire d'enregistrement est d'aider à vérifier si un langage est enregistré ou à quel langage ou variante de langage une sous étiquette particulière se réfère. Dans la plupart des cas, la référence à une grammaire ou dictionnaire d'autorité de ce langage va être utile ; dans les cas où il n'existe pas de tels travaux, d'autres travaux bien connus qui décrivent ce langage ou dans ce langage PEUT être appropriés. Le réviseur des sous étiquettes de langue décide ce qui constitue une "assez bonne" référence. Cette exigence n'est pas destinée à exclure des langages ou dialectes particuliers à cause de la taille de la population de locuteurs ou du manque d'orthographe normalisée. Les langages de minorités vont être considérés également sur leurs propres mérites.

### 3.6. Possibilités d'enregistrement

Les possibilités d'enregistrement de sous étiquettes ou d'informations sur des sous étiquettes incluent :

- o Les sous étiquettes de langage principal pour les langages qui ne figurent pas dans ISO 639 qui ne sont pas des variantes de langages mentionnés ou enregistrés PEUVENT être enregistrées. Au moment de la rédaction du présent document, il n'y avait pas d'exemple de cette forme de sous étiquette. Avant de tenter d'enregistrer une sous étiquette de langage, il DOIT y avoir une tentative d'enregistrer le langage dans la norme ISO 639. Des sous étiquettes NE DOIVENT PAS être enregistrées pour les langages définis par des codes qui existent dans ISO 639-1, ISO 639-2, ou ISO 639-3, celles qui sont prises en considération par les autorités d'enregistrement de ISO 639, ou dont l'enregistrement auprès de ces autorités n'a jamais été tenté. Si ISO 639 a précédemment rejeté l'enregistrement d'un langage, il est raisonnable de supposer qu'il doit y avoir des preuves supplémentaires, très impératives du besoin avant qu'il soit enregistré comme sous étiquette de langage principal dans le registre IANA (dans la mesure où il est très peu probable que soient enregistrées des sous étiquettes de ce type).
- o Les dialectes ou autres divisions ou variations au sein d'un langage, son orthographe, système d'écriture, usage régional ou historique, translittération ou autres transformations, ou variation distinctives PEUVENT être enregistrées comme sous étiquettes de variante. Un exemple est la sous étiquette 'rozaj' (le dialecte résien de Slovénie).
- o L'ajout ou le maintien de champs (généralement de nature informative) dans les enregistrements d'étiquette ou sous étiquette comme décrit au paragraphe 3.1 est permis. De tels changements sont soumis aux dispositions de stabilité du paragraphe 3.4. Cela inclut les champs "Description", "Comments", "Deprecated", et "Valeur préférée" pour les champs obsolètes ou supprimés, ou l'ajout des champs "Suppress-Script" ou "Macrole langage" des sous étiquettes principales de langage, ainsi que les autres changements permis par le présent document, comme l'ajout d'un champ "Prefix" approprié à une sous étiquette de variante.
- o L'ajout d'enregistrements et de changements de valeurs relatives à un champ nécessaires pour refléter des allocations faites par ISO 639, ISO 15924, ISO 3166-1, et UN M.49 comme décrit au paragraphe 3.4 est permis.

Les sous étiquettes proposées à l'enregistrement qui causeraient que tout ou partie d'une étiquette faisant l'objet d'une clause d'antériorité devienne redondante mais dont la signification entre en conflit avec, ou altère, la signification de l'étiquette faisant l'objet d'une clause d'antériorité DOIVENT être rejetées.

Le présent document laisse au processus d'enregistrement décrit au paragraphe 3.5 la décision de quelles sous étiquettes ou changements à des sous étiquettes sont appropriés (ou pas).

Note : les sous étiquettes de langage principal de quatre caractères sont réservées pour permettre la possibilité de codes alpha4 dans de futurs ajouts à la famille de normes ISO 639.

ISO 639 définit une autorité d'enregistrement pour les ajouts et changements à la liste des langages de ISO 639. Cette agence est le Centre international d'information sur la terminologie (Infoterm) Aichholzgasse 6/12, AT-1120 Vienne, Autriche ; téléphone: +43 1 26 75 35 Ext. 312 Fax : +43 1 216 32 72.

ISO 639-2 définit une autorité d'enregistrement pour les ajouts et changements à la liste des langages de ISO 639-2. Cette agence est :

Library of Congress  
Network Development and MARC Standards Office  
Washington, DC 20540, USA  
téléphone : +1 202 707 6237 Fax: +1 202 707 0115  
URL : <http://www.loc.gov/standards/iso639-2>

ISO 639-3 définit une autorité d'enregistrement pour les ajouts et changements à la liste des langages de ISO 639-3. Cette agence est :

SIL International  
ISO 639-3 Registrar  
7500 W. Camp Wisdom Rd.  
Dallas, TX 75236, USA  
téléphone : +1 972 708 7400, ext. 2293  
Fax: +1 972 708 7546  
mél : [iso639-3@sil.org](mailto:iso639-3@sil.org)  
URL: <http://www.sil.org/iso639-3>

ISO 639-5 définit une autorité d'enregistrement pour les ajouts et changements à la liste des langages de ISO 639-5. Cette agence est la même que pour ISO 639-2 et est :

Library of Congress  
Network Development and MARC Standards Office  
Washington, DC 20540, USA  
téléphone : +1 202 707 6237  
Fax : +1 202 707 0115  
URL : <http://www.loc.gov/standards/iso639-5>

L'agence de maintenance pour ISO 3166-1 (codes de pays) est :

Agence de maintenance ISO 3166  
c/o Organisation Internationale de normalisation  
Case postale 56  
CH-1211 Genève 20  
téléphone : +41 22 749 72 33 Fax: +41 22 749 73 49  
URL : <http://www.iso.org/iso/en/prods-services/iso3166ma/index.html>

L'autorité d'enregistrement pour ISO 15924 (codes d'écriture) est :

Unicode Consortium  
Box 391476  
Mountain View, CA 94039-1476, USA  
URL : <http://www.unicode.org/iso15924>

La division Statistiques du Secrétariat des Nations Unies tient la norme des codes de pays ou zones pour utilisation statistique et peut être jointe à :

Statistical Services Branch  
Statistics Division  
United Nations, Room DC2-1620  
New York, NY 10017, USA  
Fax : +1-212-963-0623  
mél : [statistics@un.org](mailto:statistics@un.org)

URL : <http://unstats.un.org/unsd/methods/m49/m49alpha.htm>

### 3.7 Extensions et registre des extensions

Les sous étiquettes d'extension sont celles introduites par des sous étiquettes d'un seul caractère ("singletons") autres que 'x'. Elles sont réservées pour la génération d'identifiants qui contiennent un composant de langage et sont compatibles avec les applications qui comprennent les étiquettes de langage.

La structure et la forme des extensions sont définies par le présent document afin que des mises en œuvre puissent être créées qui soient compatibles avec les applications qui pourraient être créées en utilisant des singletons à l'avenir. De plus, définir un mécanisme pour maintenir des singletons va donner de la stabilité du présent document en réduisant le probable besoin de futures révisions ou mises à jour.

Les sous étiquettes d'un seul caractère sont allouées par l'IANA en utilisant la politique de "revue de l'IETF" définie par la [RFC5226]. Cette politique exige le développement d'une RFC, qui DEVRA définir le nom, l'objet, les processus, et les procédures pour maintenir les sous étiquettes. L'autorité de maintenance ou d'enregistrement, incluant le nom, l'adresse de messagerie de contact, la liste de diffusion de discussion, et l'URL de localisation du registre, DOIVENT être indiqués clairement dans la RFC. La RFC DOIT spécifier ou inclure chacun des éléments suivants :

- o La spécification DOIT faire référence à la version ou révision spécifique du document qui gouverne sa création et DOIT faire référence à ce paragraphe de ce document.
- o La spécification et toutes les sous étiquettes définies par la spécification DOIVENT suivre l'ABNF et les autres règles pour la formation des étiquettes et sous étiquettes comme défini dans le présent document. En particulier, elle DOIT spécifier que la casse n'est pas significative et que les sous étiquettes NE DOIVENT PAS excéder huit caractères.
- o La spécification DOIT spécifier une représentation canonique.
- o La spécification de sous étiquettes valides DOIT être disponible gratuitement sur l'Internet.
- o La spécification DOIT être dans le domaine public ou disponible via une licence sans redevance acceptable pour l'IETF et spécifiée dans la RFC.
- o La spécification DOIT être munie d'un numéro de version, et chaque version de la spécification DOIT être numérotée, datée, et stable.
- o La spécification DOIT être stable. C'est-à-dire, les sous étiquettes d'extension, une fois définies par une spécification, NE DOIVENT PAS être rétractées ou changer de signification de façon substantielle.
- o La spécification DOIT inclure, dans une section séparée, le formulaire d'enregistrement reproduit dans ce paragraphe (ci-dessous) à utiliser pour l'enregistrement de extension lors de sa publication comme RFC.
- o L'IANA DOIT être informée des changements des informations de contact et de l'URL pour la spécification.

L'IANA tiendra un registre des sous étiquettes allouées d'un seul caractère (singleton). Ce registre DOIT utiliser le format "record-jar" décrit par l'ABNF au paragraphe 3.1.1. Lors de la publication d'une extension comme RFC, l'autorité de maintenance définie dans la RFC DOIT transmettre ce formulaire d'enregistrement à <iesg@ietf.org>, qui DOIT transmettre la demande à <iana@iana.org>. L'autorité de maintenance de l'extension DOIT veiller à la précision de l'enregistrement en envoyant une pleine copie de mise à jour de l'enregistrement à <iana@iana.org> avec pour ligne d'objet "LANGUAGE TAG EXTENSION UPDATE" chaque fois que le contenu change. Seuls les champs "Comments", "Contact\_Email", "Mailing\_List", et "URL" PEUVENT être modifiés dans ces mises à jour.

Manquer à maintenir cet enregistrement, à maintenir le registre correspondant, ou à satisfaire les autres conditions imposées par ce paragraphe du présent document PEUT faire l'objet d'un appel à l'IESG [RFC2028] selon les mêmes règles que les autres décisions de l'IETF (voir la [RFC2026]) et PEUT résulter en ce que l'autorité pour maintenir l'extension soit retirée ou réallouée par l'IESG.

```
%%  
Identifiant :  
Description :  
Commentaires :  
Ajout :  
RFC :  
Autorité :  
Adresse de messagerie de contact :  
Liste de diffusion :  
URL :  
%%
```

**Figure 6 : Format des enregistrements dans le registre des extensions d'étiquettes de langue**

"Identifiant" contient la sous étiquette d'un seul caractère (singleton) allouée à l'extension. Le projet Internet soumis pour définir l'extension DEVRAIT spécifier quelle lettre ou chiffre utiliser, mais l'IESG PEUT changer l'allocation quand elle approuve la RFC.

"Description" contient le nom et la description de l'extension. "Commentaires" est un champ FACULTATIF et PEUT contenir une description plus large de l'extension.

"Ajout" contient la date de publication de la RFC de l'extension dans le format "full-date" spécifié dans la [RFC3339]. Par exemple : 2004-06-28 représente le 28 juin 2004, dans le calendrier grégorien.

"RFC" contient le numéro de RFC alloué à cette extension.

"Autorité" contient le nom de l'autorité de maintenance pour l'extension.

"Adresse de messagerie de contact" contient l'adresse de messagerie utilisée pour contacter l'autorité de maintenance.

"Liste de diffusion" contient l'URL ou l'adresse de messagerie de l'abonnement à la liste de diffusion utilisée par l'autorité de maintenance.

"URL" contient l'URL du registre pour cette extension.

La détermination de si un projet Internet satisfait les conditions ci-dessus et la décision d'accorder ou retirer une telle autorité appartient seulement à l'IESG et est soumise au processus normal de révision et d'appel associé au traitement des RFC.

Les auteurs d'extension sont avertis que beaucoup de processeurs (y compris les mieux formés) vont ignorer toute relation particulière ou signification inhérente à l'ordre des sous étiquettes d'extension. Les auteurs d'extension DEVRAIENT éviter des relations de sous étiquette ou des mécanismes de canonisation qui interfèrent avec la confrontation ou avec des restrictions de longueur qui existent parfois dans des protocoles courants où l'extension est utilisée. En particulier, des applications PEUVENT tronquer les sous étiquettes en faisant une confrontation ou en les faisant tenir dans des longueurs limitées, donc il est RECOMMANDÉ que les informations les plus significatives soient dans les sous étiquettes les plus significatives (les plus à gauche) et que la spécification traite en douceur les sous étiquettes tronquées.

Quand une étiquette de langage est à utiliser dans un protocole spécifique connu, il est RECOMMANDÉ que l'étiquette de langage ne contienne pas d'extension non prise en charge par ce protocole. De plus, on notera que certains protocoles PEUVENT imposer des limites supérieures à la longueur des chaînes utilisées pour mémoriser ou transporter l'étiquette de langage.

### **3.8 Mise à jour du registre des sous étiquettes de langage**

Après l'adoption du présent document, le registre IANA des sous étiquettes de langue avait besoin d'une mise à jour afin qu'il contienne l'ensemble complet des sous étiquettes valides dans une étiquette de langage. La [RFC5645] décrit le processus utilisé pour créer cette mise à jour.

Les enregistrements qui sont en cours selon les règles définies dans la [RFC4646] quand le présent document sera adopté DEVRONT être complétés selon les règles contenues dans le présent document.

### **3.9 Applicabilité du registre des sous étiquettes de langage**

Le registre des sous étiquettes de langue est la source des éléments de données utilisés pour construire les étiquettes de langage, suivant les règles décrites dans le présent document. Les étiquettes de langage sont conçues pour indiquer les attributs linguistiques des divers contenus, incluant non seulement du texte mais aussi la plupart des formats de support, comme la vidéo ou l'audio. Elles forment aussi la base du langage et de la négociation locale dans divers protocoles et API.

Le registre est donc applicable à de nombreuses applications qui ont besoin d'une forme d'identification de langage, avec les limitations suivantes :

- o Il n'est pas destiné à être la seule source de données dans la création d'une interface d'utilisateur de choix de langage. Par exemple, le registre ne contient pas de traduction pour les descriptions de sous étiquette ou pour les étiquettes composées de sous étiquettes. Les sources pour des données localisées sur la base du registre sont généralement disponibles, notamment dans [CLDR]. Le registre n'indique pas non plus quelles combinaisons de sous étiquettes sont particulièrement utiles ou pertinentes.

- o Il ne fournit pas d'information indiquant les relations entre les différents langages, comme il pourrait être utilisé dans une interface d'utilisateur pour choisir hiérarchiquement, régionalement, ou sur d'autres modèles organisationnels, les étiquettes de langage.
- o Il ne fournit pas d'information sur des chevauchements potentiels entre différentes étiquettes de langage, car la notion de ce qui constitue un langage n'est pas précise : plusieurs étiquettes de langage différentes pourraient être des choix raisonnables pour le même élément de contenu.
- o Il ne contient pas d'information sur les choix de repli appropriés quand on effectue la négociation de langage. Un bon langage de repli pourrait être linguistiquement sans relation avec le langage spécifié. Le fait qu'un langage soit souvent utilisé comme langage de repli pour un autre est généralement le résultat de facteurs externes, comme la géographie, l'histoire, ou la culture -- facteurs qui pourraient ne pas s'appliquer dans tous les cas. Par exemple, la plupart des gens qui utilisent le breton (un langage celtique utilisé dans l'ouest de la France) vont probablement préférer être servis en français (une langue romane) si le breton n'est pas disponible.

## 4. Formation et traitement des étiquettes de langage

Cette Section traite de comment utiliser les informations du registre avec la syntaxe d'étiquette pour choisir, former, et traiter les étiquettes de langage.

### 4.1 Choix de l'étiquette de langage

Le principe directeur pour la formation des étiquettes de langage est de "étiqueter sagement le contenu". Parfois il y a un choix entre plusieurs étiquettes possibles pour le même contenu. Le choix de l'étiquette à utiliser dépend du contenu et de l'application en question, et un peu de jugement pourrait être nécessaire dans le choix d'une étiquette.

L'interopérabilité est mieux servie quand la même étiquette de langage est utilisée de façon cohérente pour représenter le même langage. Si une application a des exigences qui rendent les présentes règles inapplicables, alors cette application risque d'endommager l'interopérabilité. Il est fortement RECOMMANDÉ que les utilisateurs ne définissent pas leurs propres règles pour le choix des étiquettes de langage.

Les normes, protocoles, et applications qui font référence au présent document de façon normative mais appliquent des règles différentes de celles données dans ce paragraphe DOIVENT spécifier comment le choix d'étiquette de langage diffère des lignes directrices données ici.

Pour s'assurer d'une rétro compatibilité cohérente, le présent document contient plusieurs dispositions pour tenir compte de l'instabilité potentielle des normes utilisées pour définir les sous étiquettes qui constituent les étiquettes de langage.

Ces dispositions signifient qu'aucune étiquette de langage valide ne peut devenir invalide, ni qu'une étiquette de langage a une portée plus réduite à l'avenir (elle peut avoir une portée plus large). L'étiquette de langage la plus appropriée pour une application ou élément de contenu donné pourrait évoluer dans le temps, mais une fois appliquée, l'étiquette elle-même ne peut pas devenir invalide ou avoir sa signification complètement changée.

Une sous étiquette DEVRAIT seulement être utilisée quand elle ajoute des informations distinctives utiles à l'étiquette. Les sous étiquettes étrangères interfèrent avec la signification, la compréhension, et le traitement des étiquettes de langage. En particulier, les utilisateurs et les mises en œuvre DEVRAIENT suivre les champs "Prefix" et "Suppress-Script" dans le registre (défini au paragraphe 3.1) : ces champs donnent des directives sur quand des sous étiquettes spécifiques supplémentaires DEVRAIENT être utilisées ou évitées dans une étiquette de langage.

Le choix des sous étiquettes utilisées pour former une étiquette de langage DEVRAIT suivre ces lignes directrices :

1. Utiliser des étiquettes aussi précises que possible, mais pas plus spécifiques que justifié. Éviter d'utiliser des sous étiquettes qui ne sont pas importantes pour distinguer le contenu d'une application.
  - \* Par exemple, 'de' pourrait suffire pour étiqueter un message écrit en allemand, tandis que "de-CH-1996" est probablement d'une précision inutile pour une telle tâche.
  - \* Noter que certaines séquences de sous étiquettes pourraient ne pas représenter le langage qu'un utilisateur occasionnel peut attendre. Par exemple, le langage suisse allemand (Schweizerdeutsch) est représenté par "gsw-CH" et non par "de-CH". Cette dernière étiquette représente l'allemand ('de') comme utilisé en Suisse ('CH'), aussi appelé haut allemand de Suisse (Schweizer Hochdeutsch). Les deux sont des langages réels, et les distinguer pourrait être important pour une application.

2. La sous étiquette de script NE DEVRAIT PAS être utilisée pour former des étiquettes de langage sauf si l'écriture ajoute des informations distinctives à l'étiquette. Les sous étiquettes de script ont d'abord été formellement définies dans la [RFC4646]. Leur utilisation peut affecter la confrontation et l'identification de sous étiquettes pour les mises en œuvre de la [RFC1766] ou de la [RFC3066] (qui sont rendues obsolètes par le présent document) car ces sous étiquettes apparaissent entre les sous étiquettes de langage principal et de région. Certaines applications peuvent tirer avantage de l'utilisation des sous étiquettes de script dans les étiquettes de langage, pour autant que l'utilisation est cohérente pour un contexte donné. Les sous étiquettes de script ne sont jamais appropriées pour un contenu non écrit (comme des enregistrements audio). Le champ "Suppress-Script" dans l'enregistrement de langage principal ou étendu dans le registre indique des sous étiquettes de script qui n'ajoutent pas d'informations distinctives pour la plupart des applications ; ce champ définit quand les utilisateurs NE DEVRAIENT PAS inclure une sous étiquette de script avec une sous étiquette de langage principal particulière. Par exemple, si une mise en œuvre choisit un contenu en utilisant le filtrage de base [RFC4647] (décrit à l'origine au paragraphe 14.4 de la [RFC2616]) et si l'utilisateur a demandé la gamme de langages "en-US", le contenu étiqueté "en-Latn-US" ne va pas correspondre à la demande et donc ne va pas être choisi. Donc, il est important de savoir quand les sous étiquettes de script vont être utilisées et quand elles ne devraient pas l'être.

Par exemple:

- \* La sous étiquette 'Latn' ne devrait pas être utilisée avec le langage principal 'en' parce que presque tous les documents anglais sont écrits en écriture latine et cela n'ajoute pas d'informations distinctives. Cependant, si un document était écrit en anglais mélangeant l'écriture latine avec une autre écriture comme le braille ('Brai') alors il pourrait être approprié de choisir d'indiquer les deux écritures pour aider au choix de contenu, comme l'application d'une feuille de style.
  - \* Quand on étiquette un contenu qui n'est pas écrit (comme un enregistrement de parole humaine) la sous étiquette de script ne devrait pas être utilisée, même si le langage est couramment écrit dans plusieurs écritures. Donc, les sous titres d'un film pourraient utiliser l'étiquette "uz-Arab" (ouzbek, écriture arabe) mais la piste audio pour le même langage serait étiquetée simplement "uz". (L'étiquette "uz-Zxxx" pourrait aussi être utilisée lorsque le contenu n'est pas écrit, car la sous étiquette 'Zxxx' représente le "code pour documents non écrits".)
3. Si une étiquette ou sous étiquette a un champ "Valeur préférée" dans son entrée de registre, alors la valeur de ce champ DEVRAIT être utilisée pour former l'étiquette de langage de préférence à l'étiquette ou sous étiquette dans laquelle la valeur préférée apparaît.
    - \* Par exemple, utiliser 'jbo' pour Lojban de préférence à l'étiquette faisant l'objet d'une clause d'antériorité "art-lojban".
  4. Utiliser des sous étiquettes ou séquences de sous étiquettes pour les langages individuels de préférence aux sous étiquettes pour les collections de langages. Une "collection de langages" est un groupe de langages qui descendent d'un ancêtre commun, sont parlés dans la même zone géographique, ou sont autrement en relations. Certaines collections de langages ont des codes alloués par [ISO639-5] (et certains de ces codes [ISO639-5] sont aussi définis comme des collections dans [ISO639-2]). Ces codes sont inclus comme des sous étiquettes de langage principal dans le registre. Les sous étiquettes pour une collection de langages dans le registre ont un champ "Scope" avec une valeur de "collection". Une sous étiquette pour une collection de langages est toujours préférée à des solutions de remplacement moins spécifiques comme 'mul' et 'und' (voir ci-dessous) et une sous étiquette représentant une collection de langages PEUT être utilisée quand des informations de langage plus spécifiques ne sont pas disponibles. Cependant, la plupart des utilisateurs et des mises en œuvre ne savent pas qu'il y a une relation entre la collection et ses langages individuels. De plus, la relation entre les langages individuels dans la collection n'est pas bien définie ; en particulier, les langages ne sont généralement pas mutuellement intelligibles. Comme les sous étiquettes sont différentes, une demande pour la collection va normalement seulement produire des éléments étiquetés avec la sous étiquette de la collection, et pas les éléments étiquetés avec des sous étiquettes pour les langages individuels contenus dans la collection.
    - \* Par exemple, les collections sont interprétées de façon inclusive, de sorte que la sous étiquette 'gem' (langages germaniques) pourrait, mais NE DEVRAIT PAS, être utilisée avec un contenu qui serait mieux étiqueté avec "en" (anglais) "de" (allemand) ou "gsw" (suisse allemand, alémanique). Bien que 'gem' collecte tous ces langages (et d'autres) la plupart des mises en œuvre ne vont pas confronter 'gem' aux langages individuels ; donc, utiliser la sous étiquette ne va pas produire le résultat désiré.
  5. [ISO639-2] a défini plusieurs codes inclus dans le registre des sous étiquettes qui exigent une attention supplémentaire quand on choisit les étiquettes de langage. Dans la plupart de ces cas, où omettre l'étiquette de langage est permis, cette omission est préférable à l'utilisation de ces codes. Les étiquettes de langage NE DEVRAIENT PAS incorporer ces sous étiquettes comme préfixe, sauf si les informations supplémentaires apportent de la valeur à l'application.
    - \* La sous étiquette de langage principal 'mul' (Multiple) identifie un contenu dans plusieurs langages. Cette sous étiquette NE DEVRAIT PAS être utilisée quand une liste de langages ou d'étiquettes individuelles pour chaque élément de contenu peut être utilisée à la place. Par exemple, l'en-tête "Content-Language" de la [RFC3282] permet d'utiliser une liste des langages à utiliser, pas juste une seule étiquette de langage.
    - \* La sous étiquette de langage principal 'und' (indéterminé) identifie un contenu linguistique dont le langage n'est pas déterminé. Cette sous étiquette NE DEVRAIT PAS être utilisée sauf si une étiquette de langage est exigée et si les

informations de langage ne sont pas disponibles ou ne peuvent pas être déterminées. Omettre l'étiquette de langage (lorsque c'est permis) est préféré. La sous étiquette 'und' pourrait être utile pour les protocoles qui exigent qu'une étiquette de langage soit fournie ou lorsque une sous étiquette de langage principal est exigée (comme dans in "und-Latn"). La sous étiquette 'und' PEUT aussi être utile quand on confronte les étiquettes de langage dans certaines situations.

- \* La sous étiquette de langage principal 'zxx' (Non-linguistique, non applicable) identifie un contenu pour lequel une classification de langage est inappropriée ou ne s'applique pas. Des exemples pourraient inclure de la musique instrumentale ou électronique, des enregistrements de sons consistant en sons non verbaux, du matériel audiovisuel sans narration, dialogue, titres imprimés, ou sous titres, des fichiers de données lisibles par la machine consistant en langages machine ou codes de caractère, ou du code source de programmation.
  - \* La sous étiquette de langage principal 'mis' (non codé) identifie un contenu dont le langage est connu mais n'a actuellement pas de sous étiquette correspondante. Cette sous étiquette NE DEVRAIT PAS être utilisée. Parce que l'ajout d'autres codes à l'avenir peut rendre son application invalide, il est de nature instable et donc incompatible avec les objectifs de stabilité du BCP 47. Il est toujours préférable d'utiliser d'autres sous étiquettes : soit 'und', soit (avec accord préalable) des sous étiquettes d'utilisation privée.
6. Utiliser des sous étiquettes de variante avec modération et dans l'ordre correct. La plupart des sous étiquettes de variante ont un ou plusieurs champs "Prefix" dans le registre qui expriment la liste des sous étiquettes avec lesquelles ils sont appropriés. Des variantes DEVRAIENT seulement être utilisées avec des sous étiquettes qui apparaissent dans un de ces champs "Prefix". Si une variante mentionne une seconde variante dans un de ses champs "Prefix", la première variante DEVRAIT apparaître directement après la seconde variante dans toute étiquette de langage où les deux se produisent. Les variantes d'usage général (celles sans champ "Prefix" du tout) DEVRAIENT apparaître après toutes les autres sous étiquettes de variante. On ordonne toutes les variantes restantes en plaçant d'abord la sous étiquette la plus significative. Si aucune des sous étiquettes n'est plus significative ou si aucune relation ne peut être déterminée, on les range dans l'ordre alphabétique. Parce que les variantes sont très spécialisées, en utiliser beaucoup ensemble rend généralement l'étiquette si étroite que cela noie la précision supplémentaire ajoutée. Mettre les sous étiquettes dans un autre ordre interfère avec l'interopérabilité, ainsi qu'avec l'interprétation globale de l'étiquette.

Par exemple :

- \* l'étiquette "en-scotland-fonipa" (anglais, dialecte écossais, transcription phonétique IPA) est correctement ordonnée parce que 'scotland' a un "Prefix" de "en", tandis que 'fonipa' n'a pas de champ "Prefix".
  - \* l'étiquette "sl-IT-rozaj-biske-1994" est correctement ordonnée : 'rozaj' mentionne "sl" comme seul "Prefix", 'biske' mentionne "sl-rozaj" comme seul "Prefix". La sous étiquette '1994' a plusieurs préfixes, y compris "sl-rozaj". Cependant, il suit à la fois 'rozaj' et 'biske' parce que un de ses champs "Prefix" est "sl-rozaj-biske".
7. L'étiquette faisant l'objet d'une clause d'antériorité "i-default" (langage par défaut) était à l'origine enregistrée en accord avec la [RFC1766] pour satisfaire les besoins de la [RFC2277]. Elle n'est pas utilisée pour indiquer un langage spécifique, mais plutôt pour identifier une condition ou contenu utilisé lorsque les préférences de langage de l'utilisateur ne peuvent pas être établies. Elle NE DEVRAIT PAS être utilisée sauf comme moyen d'étiqueter le contenu par défaut pour des applications ou protocoles qui exigent qu'un contenu de langage par défaut soit étiqueté avec cette étiquette spécifique. Elle PEUT aussi être utilisée par une application ou protocole pour identifier quand le contenu de langage par défaut est retourné.

#### 4.1.1 Étiquetage des langages inclus

Certains enregistrements de langage principal dans le registre ont un champ "Macrole language" (paragraphe 3.1.10) qui contient une transposition de chaque "langage englobé" à son langage macrole. La transposition "Macrole language" ne définit pas quelle est la relation entre le langage englobé et son langage macrole, ni comment les langages englobés par le même langage macrole sont en relation les uns avec les autres. Deux différents langages englobés par le même langage macrole peuvent différer l'un de l'autre plus, disons, que le français et l'espagnol.

Quelques langage macrole spécifiques, comme le chinois ('zh') et l'arabe ('ar'), sont traités différemment, voir le paragraphe 4.1.2.

La sous étiquette de langage englobé la plus spécifique DEVRAIT être utilisée pour former l'étiquette de langage, bien que la sous étiquette de langage principal du langage macrole ou la sous étiquette du langage englobé PUISSE être utilisée. Cela signifie, par exemple, d'étiqueter le cree courant avec 'crk' plutôt que 'cr' (cree), et ainsi de suite.

La portée de chaque sous étiquette de langage macrole inclut, par définition, tous ses langages englobés. Comme la relation entre les langages englobés varie, les utilisateurs ne peut pas supposer que la sous étiquette de langage macrole signifie un langage englobé particulier, ni qu'une paire de langages englobés donnée sont mutuellement intelligibles ou autrement interchangeables.

Les applications PEUVENT utiliser des informations de langage macrole pour améliorer la confrontation ou la négociation de langages. Par exemple, l'information que 'sr' (serbe) et 'hr' (croate) partagent un langage macrole exprime une relation plus étroite entre ces langues que entre, disons, 'sr' (serbe) et 'ma' (macédonien). Cependant, cette relation n'est ni garantie ni exclusive. Par exemple, le roumain ('ro') et le moldave ('mo') ne partagent pas de langage macrole, mais sont en relation bien plus étroite l'un avec l'autre que le cantonais ('yue') et le wu ('wu'), qui eux partagent un langage macrole.

#### 4.1.2 Utilisation de sous étiquettes de langage étendu

Pour s'accommoder des formes d'étiquette de langage utilisées avant l'adoption du présent document, les étiquettes de langage fournissent un mécanisme spécial de compatibilité : la sous étiquette de langage étendu. Les langages choisis ont reçu des sous étiquettes de langage principal et étendu. Cela inclut des langages macrole, comme le malais ('ms') et l'ouzbek ('uz') qui ont une variété dominante spécifique généralement synonyme avec le langage macrole. D'autres langages, comme les langages macroles chinois ('zh') et arabe ('ar') et les divers langages de signes ('sgn'), ont traditionnellement utilisé leur sous étiquette de langage principal, éventuellement couplée avec diverses sous étiquettes de région ou au titre d'une étiquette faisant l'objet d'une clause d'antériorité enregistrée, pour indiquer le langage.

Avec l'adoption du présent document, des sous étiquettes spécifiques de la norme ISO 639-3 sont devenues disponibles pour identifier les langages contenus dans ces diverses familles ou groupements de langages. Cela présente un choix d'étiquettes de langage où précédemment aucune n'existait :

- o Chaque sous étiquette de langage englobé DEVRAIT être utilisée comme sous étiquette de langage principal. Par exemple, un document en chinois mandarin va être étiqueté "cmn" (la sous étiquette pour le chinois mandarin) de préférence à "zh" (chinois).
- o Si la compatibilité est désirée ou nécessaire, la sous étiquette englobée PEUT être utilisée comme une sous étiquette de langage étendu. Par exemple, un document en chinois mandarin pourrait être étiqueté "zh-cmn" à la place de "cmn" ou "zh".
- o Le langage macrole ou la sous étiquette de préfixe PEUT quand même être utilisée pour former l'étiquette à la place de la sous étiquette de langage englobé la plus spécifique. C'est-à-dire, des étiquettes comme "zh-HK" ou "sgn-RU" sont toujours valides.

Le chinois ('zh') fournit une utile illustration de cela. Dans le passé, divers contenus ont utilisé des étiquettes commençant par la sous étiquette 'zh', avec une signification spécifique de l'application associée à des codes de région, des séquences d'utilisation privée, ou des valeurs enregistrées faisant l'objet d'une clause d'antériorité. C'est parce que historiquement seule la sous étiquette de langage macrole 'zh' était disponible pour former des étiquettes de langage. Cependant, les langages englobés par la sous étiquette de chinois 'zh' sont, principalement, non mutuellement intelligibles quand ils sont parlés, et les formes écrites de ces langues montrent aussi une large variation de forme et d'usage.

Pour assurer la compatibilité, les langages chinois englobés par la sous étiquette 'zh' sont dans le registre à la fois comme sous étiquettes de langage principal et comme sous étiquettes de langage étendu. Par exemple, le code ISO 639-3 pour le cantonais est 'yue'. Un contenu en cantonais pourrait historiquement avoir utilisé une étiquette comme "zh-HK" (car le cantonais est couramment parlé à Hong Kong) bien que cette étiquette signifie en fait n'importe quel type de chinois utilisé à Hong Kong. Avec la disponibilité de codes de ISO 639-3 dans le registre, le contenu en cantonais peut être directement étiqueté en utilisant la sous étiquette 'yue'. Le contenu peut l'utiliser comme sous étiquette de langage principal, comme dans l'étiquette "yue-HK" (cantonais, Hong Kong). Ou il peut utiliser une sous étiquette de langage étendu avec 'zh', comme dans l'étiquette "zh-yue-Hant" (chinois, cantonais, écriture traditionnelle).

Comme noté précédemment, les applications peuvent choisir d'utiliser la sous étiquette de langage macrole pour former l'étiquette à la place d'utiliser la sous étiquette de langage englobé la plus spécifique. Par exemple, une application avec de grandes quantités de données utilisant déjà des étiquettes avec la sous étiquette 'zh' (chinois) pourrait continuer d'utiliser cette sous étiquette plus générale même pour de nouvelles données, bien que le contenu pourrait être étiqueté plus précisément avec 'cmn' (mandarin), 'yue' (cantonais), 'wu' (wu), et ainsi de suite. De même, une application qui utilise déjà des étiquettes qui commencent par la sous étiquette 'ar' (arabe) pourrait continuer d'utiliser cette sous étiquette plus générale même pour de nouvelles données, qui pourraient être étiquetées plus précisément avec 'arb' (arabe standard).

Dans certains cas, les langages englobés avaient des étiquettes enregistrées pour eux durant l'époque de la RFC 3066. Ces étiquettes faisant l'objet d'une clause d'antériorité non déjà déconseillées ou rendues redondantes ont été déconseillées dans le registre à l'adoption du présent document. Comme valeurs faisant l'objet d'une clause d'antériorité, elles restent d'utilisation valide, et certains contenus ou applications pourraient les utiliser. Comme avec d'autres étiquettes faisant l'objet d'une clause d'antériorité, comme les mises en œuvre pourraient n'être pas capables d'associer les étiquettes faisant l'objet d'une clause d'antériorité avec la sous étiquette de langage englobée équivalente qui est recommandée par le présent document, les mises en œuvre sont encouragées à canoniser les étiquettes pour les besoins de comparaison. Des exemples incluent les étiquettes "zh-hakka" (hakka) et "zh-guoyu" (chinois mandarin ou standard).

Les langages de signes partagent un mode de communication plutôt qu'un héritage linguistique. De nombreux langages des signes se sont développés indépendamment, et la sous étiquette 'sgn' indique seulement la présence d'un langage de signes. Un certain nombre de langages de signes avaient aussi des étiquettes faisant l'objet d'une clause d'antériorité enregistrée pour eux durant la période de la RFC 3066. Par exemple, l'étiquette faisant l'objet d'une clause d'antériorité "sgn-US" était enregistrée pour représenter spécifiquement le langage américain des signes, sans référence aux États Unis. Elle est toujours valide, mais déconseillée : un document en langage américain des signes peut être étiqueté soit "ase" soit "sgn-ase" (la sous étiquette 'ase' est pour le langage appelé "langage américain des signes").

#### 4.2 Signification de l'étiquette de langage

La signification de l'étiquette de langage se rapporte à la signification des sous étiquettes qu'elle contient. Chaque sous étiquette, à son tour, implique une certaine gamme d'attentes qu'on pourrait avoir pour le contenu, bien que ce ne soit pas garanti. Par exemple, l'utilisation d'une sous étiquette de script comme 'Arab' (écriture arabe) ne signifie pas que le contenu soit seulement de caractères arabes. Elle signifie que le langage impliqué est de l'écriture arabe prédominante. Donc, une étiquette de langage et ses sous étiquettes peuvent englober une très large gamme de variations et rester quand même appropriée dans chaque instance particulière.

La validité d'une étiquette n'est pas le seul facteur qui détermine son utilité. Bien que chaque étiquette valide ait une signification, elle ne pourrait pas représenter un usage de langage du monde réel. Cela est inévitable dans un système dans lequel les sous étiquettes peuvent être librement combinées. Par exemple, des étiquettes comme "ar-Cyrl-CO" (arabe, écriture cyrillique, comme utilisé en Colombie) ou "tlh-Kore-AQ-fonipa" (klingon, écriture coréenne, comme utilisé en Antarctique, transcription phonétique IPA) sont toutes deux valides et ont peu de chances de représenter une combinaison utile d'attributs de langage.

La signification d'une étiquette donnée ne dépend pas du contexte dans lequel elle apparaît. La relation entre la signification d'une étiquette et les objets d'information auxquels cette étiquette est appliquée peut cependant varier.

- o Pour un seul objet d'information, les étiquettes de langage associées pourraient être interprétées comme l'ensemble de langages nécessaires pour une compréhension complète de l'objet complet. Exemple : les documents de texte en clair.
- o Pour une agrégation des objets d'information, les étiquettes de langage associées pourraient être prises comme l'ensemble de langages utilisés à l'intérieur des composants de cette agrégation. Exemples : des mémorisations et bibliothèques de documents.
- o Pour les objets d'information dont l'objet est de fournir des solutions de remplacement, les étiquettes de langage associées pourraient être regardées comme une indication que le contenu est fourni dans plusieurs langages et qu'on doit inspecter chacune des alternatives afin de trouver le ou les langages. Dans ce cas, la présence de plusieurs étiquettes pourrait ne pas signifier qu'on doit être multilingues pour avoir une compréhension complète du document. Exemple : MIME multipart/alternative [RFC2046].
- o Pour les langages de balisage, comme HTML et XML, les informations de langage peuvent être ajoutées à chaque partie du document identifié par la structure de balisage (incluant le document complet lui-même). Par exemple, on pourrait écrire `<span lang="fr">C'est la vie.</span>` à l'intérieur d'un document en allemand ; l'utilisateur germanophone pourrait alors accéder à un dictionnaire français-allemand pour trouver ce que veut dire la section marquée. Si l'utilisateur écoute le document sur une interface de synthèse vocale, cette formation pourrait être utilisée pour signaler au synthétiseur d'appliquer de façon appropriée les règles de prononciation de texte à prononcer en français pour ce morceau de texte, au lieu d'appliquer les règles inappropriées de l'allemand.
- o Pour les langages de balisage et les formats de document qui permettent d'identifier l'audience, une étiquette de langage pourrait indiquer la ou les audiences appropriées pour ce document. Par exemple, le même document HTML décrit à l'alinéa précédent pourrait avoir un en-tête HTTP "Content-Language: de" pour indiquer que l'audience prévue pour le fichier est allemande (même si trois mots apparaissent et sont identifiés comme du français en son sein).
- o Pour les systèmes et API, les étiquettes de langage forment la base de la plupart des mises en œuvre d'identifiants locaux. Par exemple, voir le projet de répertoire commun de données locales (CLDR, *Common Locale Data Repository*) de Unicode (voir [UTS35]).

Les étiquettes de langage sont en relation quand elles contiennent une séquence similaire de sous étiquettes. Par exemple, si une étiquette de langage B contient l'étiquette de langage A comme préfixe, alors B est normalement "plus près" ou "plus spécifique" que A. Donc, "zh-Hant-TW" est plus spécifique que "zh-Hant".

Cette relation n'est pas garantie dans tous les cas : précisément, il N'EST PAS garanti que les langages qui commencent par la même séquence de sous étiquettes soient mutuellement intelligibles, mais cela pourrait être. Par exemple, l'étiquette "az" partage un préfixe avec "az-Latn" (azéri écrit en utilisant l'écriture latine) et "az-Cyrl" (azéri écrit en utilisant l'écriture cyrillique). Une personne maîtrisant une des écritures pourrait n'être pas capable de lire l'autre, même si le contenu linguistique (par exemple, ce qui serait entendu si les deux textes étaient lus à haute voix) pourrait être identique. Le contenu étiqueté "az" est très probablement écrit dans une seule écriture et donc pourrait n'être pas intelligible à un lecteur familier d'une autre écriture.

De même, toutes les sous étiquettes ne spécifient pas une distinction réelle dans le langage. Par exemple, les étiquettes "en-US" et "en-CA" signifie, en gros, anglais avec des caractéristiques généralement vues comme des caractéristiques respectivement des États Unis et du Canada. Cela n'implique pas qu'une frontière dialectale significative existe entre un point choisi arbitrairement dans les États Unis et un point choisi arbitrairement au Canada. Ni qu'une sous étiquette de région particulière implique que des distinctions linguistiques n'existent pas au sein de cette région.

### 4.3 Listes de langages

Dans certaines applications, un seul élément de contenu pourrait être mieux associé à plus d'une étiquette de langage. Des exemples d'un tel usage incluent :

- o Des éléments de contenu qui contiennent plusieurs variantes distinctes. Ceci est souvent utilisé pour indiquer une audience appropriée pour un élément de contenu donné quand plusieurs choix pourraient être appropriés. Des exemples de cela pourraient inclure :
  - \* Des méta données sur l'audience appropriée pour le titre d'un film. Par exemple, un DVD pourrait étiqueter ses pistes audio individuelles 'de' (allemand) 'fr' (français), et 'es' (espagnol) mais le titre global mentionnerait "de, fr, es" comme audience globale.
  - \* Un dictionnaire Français/Anglais, Anglais/Français étiqueté à la fois "en" et "fr" pour spécifier qu'il s'applique également au français et à l'anglais.
  - \* Une traduction côte à côte ou interlinéaire d'un document, comme c'est fait couramment pour les ouvrages classiques en Latin ou Grec.
- o Des éléments de contenu qui contiennent un seul langage mais exigent plusieurs niveaux de spécificité. Par exemple, une bibliothèque pourrait souhaiter classer un ouvrage particulier à la fois comme norvégien ('no') et comme nynorsk ('nn') pour les audiences capables d'apprécier la distinction ou qui ont besoin de choisir le contenu de façon plus précise.

### 4.4 Considérations de longueur

Il n'y a pas de limite supérieure définie pour la taille des étiquettes de langage. Bien qu'historiquement la plupart des étiquettes de langage aient consisté en sous étiquettes de langage et de région avec une longueur totale combinée de jusqu'à six caractères, de plus grandes étiquettes ont toujours été possibles et leur utilisation est en fait apparue.

Ni la syntaxe d'étiquette de langage ni d'autres exigences dans le présent document n'imposent une limite supérieure fixe au nombre de sous étiquettes dans une étiquette de langage (et donc un limite supérieure à la taille d'une étiquette). La syntaxe de l'étiquette de langage suggère, selon les spécificités du langage, que plus de sous étiquettes (et donc une étiquette plus longue) sont parfois nécessaires pour identifier complètement le langage pour certaines applications ; donc, il est possible d'envisager des séquences longues ou complexes de sous étiquettes.

#### 4.4.1 Travail avec des mémoires tampon, de taille limitée

Certaines applications et protocoles sont forcés d'allouer des tailles fixes de mémoire tampon ou de limiter autrement la longueur d'une étiquette de langage. Une mise en œuvre ou spécification conforme PEUT refuser de prendre en charge la mémorisation des étiquettes de langage qui excèdent une longueur spécifiée. Une telle limitation DEVRAIT être clairement documentée, et une telle documentation DEVRAIT inclure ce qui arrive aux étiquettes plus longues (par exemple, si une valeur d'erreur est générée ou si l'étiquette de langage est tronquée). Un protocole qui permet que les étiquettes soient tronquées à une limite arbitraire, sans donner d'indication sur ce qu'est cette limite, peut causer des dommages en changeant de façon substantielle la signification des étiquettes.

En pratique, la plupart des étiquettes de langage n'exigent pas plus de quelques sous étiquettes et ne vont pas approcher les limitations de mémoire tampon raisonnablement dimensionnée ; voir le paragraphe 4.1.

Certaines spécifications ou protocoles ont des limites sur la longueur d'étiquette mais n'ont pas une limite fixe de longueur. Par exemple, la [RFC2231] n'a pas une limite explicite de longueur : la longueur disponible pour l'étiquette de langage est contrainte par la longueur des autres composants d'en-tête (comme le nom du jeu de caractères) couplé avec la limite de 76 caractères de la [RFC2047]. Donc, la "limite" pourrait être 50 caractères ou plus, mais elle pourrait être assez petite.

Les considérations pour la définition d'une limite de mémoire tampon sont :

- Les mises en œuvre NE DEVRAIENT PAS tronquer les étiquettes de langage sauf si la signification de l'étiquette est délibérément changée, ou sauf si l'étiquette ne tient pas dans une taille de limite de mémoire tampon spécifiée par un protocole pour la mémorisation ou la transmission.
- Les mises en œuvre DEVRAIENT avertir l'utilisateur quand une étiquette est tronquée car la troncature change la signification de l'étiquette.
- Les mises en œuvre de protocoles ou spécifications qui ont des contraintes d'espace mais n'ont pas de limite fixe DEVRAIENT utiliser l'étiquette la plus longue possible de préférence à la troncature.
- Les protocoles ou spécifications qui spécifient des tailles de mémoire tampon limitées pour les étiquettes de langage DOIVENT permettre des étiquettes de langage d'au moins 35 caractères. Noter que la [RFC4646] recommande une taille minimum de champ de 42 caractères parce que elle incluait trois éléments de la production "extlang". Deux d'entre eux sont maintenant réservés en permanence, de sorte qu'une sous étiquette de langage principal enregistrée de la longueur maximum de 8 caractères est maintenant plus longue que la plus longue combinaison de langage "extlang". Les protocoles ou spécifications qui utilisent couramment des sous étiquettes d'extensions ou d'utilisation privée pourraient souhaiter réserver ou recommander une taille "minimum de mémoire tampon" plus longue.

L'illustration suivante montre comment la recommandation de 35 caractères a été déduite :

langage = 8 ; plus longue valeur enregistrée permise plus longue que principal+extlang qui exige 7 caractères  
 script = 5 ; si pas supprimé : voir le paragraphe 4.1  
 region = 4 ; le code numérique de région UN M.49 ; les codes ISO 3166-1 exigent 3  
 variant1 = 9 ; exige "langage" comme préfixe  
 variant2 = 9 ; très rare, car elle a besoin de "language-variant1" comme préfixe  
 total = 35 caractères

**Figure 7 : Déduction des limites sur la longueur d'étiquette**

#### 4.4.2 Troncature des étiquettes de langage

La troncature d'une étiquette de langage altère la signification de l'étiquette, et donc DEVRAIT être évitée. Cependant, la troncature des étiquettes de langage est parfois nécessaire à cause d'une taille limitée de mémoire tampon. Une telle troncature NE DOIT PAS permettre qu'une sous étiquette soit coupée en son milieu, ou la formation d'étiquettes invalides (par exemple, une qui se terminerait sur le caractère "-").

Cela signifie que les applications ou protocoles qui tronquent les étiquettes DOIVENT le faire en retirant progressivement les sous étiquettes avec leur "-" précédant à partir du côté droit de l'étiquette de langage jusqu'à ce que l'étiquette soit assez courte pour la mémoire tampon. Si l'étiquette résultante se termine par une sous étiquette d'un seul caractère, cette sous étiquette et son "-" précédant DOIVENT aussi être supprimés. Par exemple :

Étiquette à tronquer : zh-Latn-CN-variant1-a-extend1-x-wadegile-private1  
 1. zh-Latn-CN-variant1-a-extend1-x-wadegile  
 2. zh-Latn-CN-variant1-a-extend1  
 3. zh-Latn-CN-variant1  
 4. zh-Latn-CN  
 5. zh-Latn  
 6. zh

**Figure 8 : Exemple de troncature d'étiquette**

#### 4.5 Canonisation des étiquettes de langage

Comme une étiquette de langage particulière peut être utilisée par de nombreux processus, les étiquettes de langage DEVRAIENT toujours être créées ou générées en forme canonique.

Une étiquette de langage est en "forme canonique" quand l'étiquette est bien formée selon les règles des paragraphes 2.1 et 2.2 et qu'elle a été canonisée en appliquant chacune des étapes suivantes dans cet ordre, en utilisant le données provenant du registre de l'IANA (voir le paragraphe 3.1) :

1. Les séquences d'extension sont ordonnées dans l'ordre ASCII insensible à la casse par sous étiquette singleton.
  - \* Par exemple, la séquence de sous étiquettes '-a-babble' vient avant '-b-warble'.
2. Les étiquettes redondantes ou faisant l'objet d'une clause d'antériorité sont remplacées par leur "Valeur préférée", si il y en a une.
  - \* Le corps de champ de la "Valeur préférée" pour les étiquettes redondantes ou faisant l'objet d'une clause d'antériorité est une "gamme de langage étendue" [RFC4647] et pourrait consister en plus d'une sous étiquette.

- \* Les champs "Valeur préférée" dans le registre donnent des transpositions d'étiquettes déconseillées en équivalants modernes. Beaucoup d'elles ont été créées avant l'adoption du présent document (comme la transposition de "nyn" en "nn" ou "i-klingon" en "tlh"). D'autres sont le résultat d'enregistrements ou ajouts ultérieurs au registre comme permis ou exigé par le présent document (par exemple, "zh-hakka" a été déconseillé en faveur du code ISO 639-3 'hak' quand le présent document a été adopté).
3. Les sous étiquettes sont remplacées par leur "Valeur préférée", si il y en a une. Pour les langues étendues, la sous étiquette de langage principal d'origine est aussi remplacée si il y a une sous étiquette de langage principal dans la "Valeur préférée".
    - \* Le corps de champ de la "Valeur préférée" pour les langues étendues est une "gamme étendue de langage" et se transpose normalement en une sous étiquette de langage principal. Par exemple, la séquence de sous étiquettes "zh-hak" (chinois, hakka) est remplacé par la sous étiquette 'hak' (hakka).
    - \* La plupart des sous étiquettes "non extlang" sont soit des sous étiquettes de région où le nom ou désignation de pays a changé, soit des corrections d'écriture à ISO 639-1.

La forme canonique ne contient pas de sous étiquette "extlang". Il y a une forme de remplacement de "extlang" qui conserve ou réinstalle les sous étiquettes "extlang". Cette forme peut être utile dans des environnements où la présence de la sous étiquette "Prefix" est considérée comme avantageuse pour la confrontation ou le choix (voir le paragraphe 4.1.2).

Une étiquette de langage est en "forme extlang" quand l'étiquette est bien formée selon les règles des paragraphes 2.1 et 2.2 et qu'elle a été traitée en appliquant à chacune les deux étapes suivantes dans l'ordre indiqué, en utilisant les données du registre IANA :

- 1 L'étiquette de langage est d'abord transformée en sa forme canonique, comme décrit ci-dessus.
2. Si l'étiquette de langage commence par une sous étiquette de langage principal qui est aussi une sous étiquette "extlang", alors l'étiquette de langage est précédée du "Prefix" de "extlang".
  - \* Par exemple, "hak-CN" (hakka, Chine) a la sous étiquette de langage principal 'hak', qui à son tour a un enregistrement "extlang" avec un "Prefix" 'zh' (chinois). La forme "extlang" est "zh-hak-CN" (chinois, hakka, Chine).
  - \* Noter que l'étape 2 (précéder d'un préfixe) peut restaurer une sous étiquette qui a été retirée à l'étape 1 (de canonisation).

Exemple : l'étiquette de langage "en-a-aaa-b-ccc-bbb-x-xyz" est en forme canonique, tandis que "en-b-ccc-bbb-a-aaa-X-xyz" est bien formée et potentiellement valide (les extensions 'a' et 'b' ne sont pas définies à la publication du présent document) mais pas en forme canonique (les extensions ne sont pas en ordre alphabétique).

Exemple : bien que l'étiquette "en-BU" (anglais comme utilisé en Birmanie) conserve sa validité, l'étiquette de langage "en-BU" n'est pas en forme canonique parce que la sous étiquette 'BU' a une transposition canonique en 'MM' (Myanmar).

La canonisation des étiquettes de langage n'a pas d'implication sur l'utilisation de lettres majuscules ou minuscules quand on traite ou compare des sous étiquettes (et comme décrit au paragraphe 2.1). Toutes les comparaisons DOIVENT être effectuées de manière insensible à la casse.

Quand on effectue la canonisation d'étiquettes de langage, les processeurs PEUVENT régulariser la casse des sous étiquettes (ce processus est FACULTATIF) suivant la casse utilisée dans le registre (voir le paragraphe 2.1.1).

Si plus d'une variante apparaît au sein d'une étiquette, les processeurs PEUVENT réordonner les variantes pour obtenir un meilleur comportement de confrontation ou une présentation plus cohérente. Réordonner les variantes DEVRAIT suivre les recommandations sur l'ordre des variantes du paragraphe 4.1.

Si le champ "Deprecated" apparaît dans un enregistrement du registre sans un champ "Valeur préférée" qui l'accompagne, cette étiquette ou sous étiquette est alors déconseillée sans remplacement. Ces valeurs sont canoniques quand elles apparaissent dans une étiquette de langage. Cependant, les étiquettes qui incluent ces valeurs NE DEVRAIENT PAS être choisies par les utilisateurs ou générées par les mises en œuvre.

Une extension DOIT définir les relations qui existent entre les diverses sous étiquettes dans l'extension et donc PEUT définir un autre schéma de canonisation pour les sous étiquettes de l'extension. Les extensions PEUVENT définir comment l'ordre des sous étiquettes de l'extension est interprété. Par exemple, une extension pourrait définir que ses sous étiquettes sont en ordre canonique quand elles sont placées dans l'ordre de l'ASCII : c'est-à-dire, "en-a-aaa-bbb-ccc" à la place de "en-a-ccc-bbb-aaa". Une autre extension pourrait définir que l'ordre des sous étiquettes influence leur signification (de sorte que "en-b-ccc-bbb-aaa" ait une valeur différente de "en-b-aaa-bbb-ccc"). Cependant, les spécifications d'extension DEVRAIENT être conçues de telle façon qu'elles soient tolérantes au traitement normal décrit au paragraphe 3.7.

#### 4.6 Considérations sur les sous étiquettes d'utilisation privée

Les sous étiquettes d'utilisation privée, comme toutes les autres sous étiquettes, DOIVENT se conformer aux contraintes de format et de contenu de l'ABNF. Les sous étiquettes d'utilisation privée n'ont pas de signification en dehors de l'accord privé entre les parties qui ont l'intention d'utiliser ou échanger les étiquettes de langage qui les emploient. Les mêmes sous étiquettes PEUVENT être utilisées avec une signification différente dans un autre accord privé. Elles NE DEVRAIENT PAS être utilisées lorsque des solutions de remplacement existent et NE DEVRAIENT PAS être utilisées dans un contenu ou des protocoles destinés à une utilisation générale.

Les sous étiquettes d'utilisation privée sont simplement sans objet pour un échange d'informations sans accord préalable. La valeur et la signification des étiquettes d'utilisation privée et des sous étiquettes utilisées au sein d'une telle étiquette de langage ne sont pas définies dans le présent document.

Les séquences d'utilisation privée introduites par le singleton 'x' sont complètement opaques aux utilisateurs ou mises en œuvre en dehors de l'accord d'utilisation privée. Donc, en plus des séquences de sous étiquettes d'utilisation privée introduites par la sous étiquette de singleton 'x', le registre des sous étiquettes de langue fournit les sous étiquettes d'utilisation privée de langage, écriture, et région déduits des codes d'utilisation privée alloués par les normes sous-jacentes. Ces sous étiquettes sont valides pour être utilisées à former des étiquettes de langage ; elles sont RECOMMANDÉES sur les séquences de sous étiquettes de singleton d'utilisation privée 'x' parce que elles portent plus d'informations via leur liaison à la structure inhérente à l'étiquette de langage.

Par exemple, les sous étiquettes de région 'AA', 'ZZ', et celles dans les gammes 'QM'-'QZ' et 'XA'-'XZ' (dérivées des codes d'utilisation privée ISO 3166-1) peuvent être utilisées pour former une étiquette de langage. Une étiquette comme "zh-Hans-XQ" porte une grande quantité d'informations publiques, échangeables sur le matériel de langage (c'est-à-dire du chinois dans l'écriture chinoise simplifiée et convient pour une certaine région géographique 'XQ'). Bien que la région géographique précise ne soit pas connue en dehors de l'accord privé, l'étiquette porte beaucoup plus d'informations qu'une étiquette opaque comme "x-somelang" ou même "zh-Hans-x-xq" (où la signification de la sous étiquette 'xq' est entièrement opaque).

Cependant, dans certains cas le contenu étiqueté avec des sous étiquettes d'utilisation privée peut interagir avec d'autres systèmes d'une manière différente et éventuellement non convenable comparée aux étiquettes qui utilisent des sous étiquettes opaques, définies en privé, de sorte que le choix de la meilleure approche dépend parfois du domaine particulier en question.

### 5. Considérations relatives à l'IANA

Cette section traite des processus et exigences nécessaires pour que l'IANA tienne les registres de sous étiquettes et d'extensions définis par le présent document et en accord avec les exigences de la [RFC5226].

L'impact sur la maintenance par l'IANA des deux registres définis par le présent document va être une légère augmentation de la fréquence de nouvelles entrées ou mises à jour. Il est aussi demandé à l'IANA de créer une nouvelle liste de diffusion (décrite au paragraphe 5.1) pour annoncer les changements et mises à jour du registre.

#### 5.1 Registre des sous étiquettes de langages

L'IANA a mis à jour le registre en utilisant les instructions et le contenu fournis par le document d'accompagnement [RFC5645]. Les critères et processus pour choisir l'ensemble mis à jour d'enregistrements sont décrits dans ce document. L'ensemble mis à jour des enregistrements n'a pas d'impact sur l'IANA, car le travail de création va être effectué en externe.

Les futurs travaux sur le registre des sous étiquettes de langue incluent les activités suivantes :

- o Insérer ou remplacer tous les enregistrements. Ces enregistrements sont préformatés pour l'IANA par le réviseur de sous étiquettes de langue, comme décrit au paragraphe 3.3.
- o Archiver et rendre publiquement disponibles les formulaires d'enregistrement.
- o Annoncer chaque version mise à jour du registre sur la liste de diffusion "ietf-languages-announcements@iana.org".

Chaque formulaire d'enregistrement envoyé à l'IANA contient un seul enregistrement à incorporer dans le registre. Le formulaire sera envoyé à <iana@iana.org> par le réviseur de sous étiquettes de langue. Il aura une ligne "subject" indiquant si le formulaire inclus représente une insertion d'un nouvel enregistrement (indiqué par le mot "INSERT" dans la ligne "subject") ou le remplacement d'un enregistrement existant (indiqué par le mot "MODIFY" dans la ligne "subject"). Un enregistrement ne peut jamais être supprimé du registre.

L'IANA va extraire l'enregistrement du formulaire et placer l'enregistrement inséré ou modifié dans la section appropriée du registre des sous étiquettes de langue, en groupant les enregistrements par champ "Type". Les enregistrements insérés peuvent être placés n'importe où au sein de la section appropriée ; il n'est pas garanti que les enregistrements du registre soient placés dans un ordre particulier sauf qu'ils vont toujours être groupés par "Type". Les enregistrements modifiés écrasent l'enregistrement qu'ils remplacent.

Chaque fois qu'une entrée est créée ou modifiée dans le registre, l'enregistrement "File-Date" au début du registre est mis à jour pour refléter la plus récente date de modification. Le format de date DEVRA être le format "full-date" de la [RFC3339]. La date DEVRA être celle à laquelle cette version du registre a été publiée par l'IANA. Il DEVRA y avoir au plus une version du registre publiée par jour. Un enregistrement "File-Date" est aussi inclus dans chaque demande à l'IANA d'insérer ou modifier des enregistrements, indiquant la date d'acceptation des enregistrements dans la demande.

Le fichier mis à jour du registre DOIT utiliser le codage de caractères UTF-8, et IANA DOIT vérifier que le fichier du registre a le codage approprié. Des caractères non ASCII peuvent être envoyés à l'IANA en attachant le formulaire d'enregistrement au message électronique ou en utilisant divers codages dans le corps de message (UTF-8 est recommandé). L'IANA vérifiera tous les caractères non clairs ou corrompus avec le réviseur de sous étiquettes de langue avant de publier la mise à jour du registre.

L'IANA archivera aussi et rendra publiquement disponible à partir de <http://www.iana.org> chaque formulaire d'enregistrement. Noter que plusieurs enregistrements peut appartenir au même enregistrement dans le registre.

Les développeurs qui dépendent du registre des sous étiquettes de langue aimeraient parfois être informés des changements dans le registre afin de pouvoir mettre à jour leurs mises en œuvre. Quand un changement est fait au registre des sous étiquettes de langue, l'IANA enverra un message d'annonce à <[ietf-languages-announcements@iana.org](mailto:ietf-languages-announcements@iana.org)> (une liste d'auto abonnement à laquelle seule l'IANA peut envoyer).

## 5.2 Registre des extensions

Le registre des extensions d'étiquettes de langue peut contenir au plus 35 enregistrements, et donc les changements à ce registre sont supposés être très peu fréquents.

Le travail futur de l'IANA sur le registre des extensions d'étiquettes de langue est limité à deux cas. D'abord, l'IESG PEUT demander que de nouveaux enregistrements soient insérés dans ce registre de temps en temps. Ces demandes DOIVENT inclure l'enregistrement à insérer dans le format exact décrit au paragraphe 3.7. De plus, il PEUT y avoir des demandes occasionnelles de l'autorité de maintenance pour une extension spécifique pour mettre à jour les informations ou l'URL de contact dans l'enregistrement. Ces demandes DOIVENT inclure l'enregistrement complet à jour. L'IANA n'est pas responsable de la validation des informations fournies, seulement qu'elles sont correctement formatées. L'IANA DEVRAIT prendre des mesures raisonnables pour s'assurer que la demande vient de l'autorité de maintenance désignée par l'enregistrement présent dans le registre.

## 6. Considérations sur la sécurité

Les étiquettes de langage utilisées dans la négociation de contenu, comme toutes les autres informations échangées sur l'Internet, pourraient être une source de problèmes parce que elles pourraient être utilisées pour déduire la nationalité de l'expéditeur, et donc identifier de potentielles cibles de surveillance.

Il y a un cas particulier du problème général que tout ce qui est envoyé est visible au receveur et éventuellement aussi à des tiers. Il est utile d'être conscient que de tels problèmes peuvent exister dans certains cas.

L'évaluation de l'ampleur exacte de la menace, et de toutes les contre-mesures possibles, est laissée à chaque protocole d'application (voir le BCP 72 [RFC3552] pour les directives sur les bonnes pratiques sur les menaces et défenses sur la sécurité).

L'étiquette de langage associée à un élément d'information particulier n'a pas de conséquence pour déterminer si le contenu pourrait contenir des homographes possibles. Le fait qu'un texte soit étiqueté comme étant dans un langage ou utilise une sous étiquette d'écriture particulière ne fournit aucune assurance qu'il ne contient pas de caractères provenant d'écritures autres que celle ou celles associées à, ou spécifiées par cette étiquette de langage.

Comme il n'y a pas de limite au nombre de sous étiquettes de variantes, d'utilisation privée, et d'extension, et par conséquent pas de limite à la longueur possible d'une étiquette, les mises en œuvre doivent se garder contre les attaques de

débordement de mémoire tampon. Voir au paragraphe 4.4 les détails de la troncature d'étiquette de langage, qui peut se produire par suite des défenses contre le débordement de mémoire tampon.

Pour prévenir les attaques de déni de service, les applications NE DEVRAIENT PAS dépendre de ce que le registre des sous étiquettes de langue ou le registre des extensions d'étiquettes de langue soient toujours accessibles. De plus, bien que la spécification de sous étiquettes valides pour une extension (voir le paragraphe 3.7) DOIVE être disponible sur l'Internet, les mises en œuvre NE DEVRAIENT PAS dépendre mécaniquement de ce que ces sources soient toujours accessibles.

Les registres spécifiés dans le présent document ne conviennent pas pour un accès fréquent ou en temps réel, ou pour la restitution de tout le contenu du registre. La plupart des applications n'ont pas besoin du tout des données du registre. Pour d'autres, être capable de valider ou canoniser les étiquettes de langage à une date particulière du registre va être suffisant, car le contenu du registre change seulement occasionnellement. Les changements sont annoncés à <ietf-languages-announcements@iana.org>. Cette liste de diffusion est destinée aux organisations et individus intéressés, pas à des abonnements pour déclencher des mises à jour automatiques de logiciel. La taille du registre ne le rend pas utilisable pour des mises à jour automatiques de logiciel. Il est fortement conseillé aux mises en œuvre qui envisagent d'intégrer le registre des sous étiquettes de langue dans un schéma de mise à jour automatique de distribuer seulement des différences convenablement codées, et seulement via leur propre infrastructure -- pas directement à partir de l'IANA.

Les changements, ou leur absence, peuvent aussi être facilement détectés en regardant l'enregistrement "File-Date" au début du registre, ou en utilisant les caractéristiques du protocole utilisé pour le téléchargement, sans avoir à télécharger le registre complet. Au moment de la publication du présent document, l'IANA rend le registre des étiquettes de langues disponible sur HTTP 1.1. La façon appropriée de mettre à jour une copie locale du registre des sous étiquettes de langue en utilisant HTTP 1.1 est d'utiliser un GET conditionnel [RFC2616].

## 7. Considérations de jeux de caractères

La syntaxe dans le présent document exige que les étiquettes de langage utilisent seulement les caractères A-Z, a-z, 0-9, et HYPHEN-MINUS (*tiret*) qui sont présents dans la plupart des jeux de caractères, de sorte que la composition des étiquettes de langage ne devrait pas avoir de problème de jeu de caractères.

Le rendu du texte fondé sur l'étiquette de langage n'est pas traité ici. Historiquement, des traitements se sont appuyés sur l'utilisation d'informations de jeu de caractères/codage (ou autres informations externes) afin de déduire comment une chaîne spécifique de caractères devrait être rendue. Notamment, cela s'applique aux variantes spécifiques de langage et de culture des idéogrammes Han comme utilisés en japonais, chinois, et coréen, où l'utilisation, par exemple, d'un codage de caractère japonais comme EUC-JP implique que le texte lui-même est en japonais. Quand des étiquettes de langage sont appliquées à des étendues de texte, les moteurs de rendu pourraient être capables d'utiliser cette information pour mieux choisir les fontes ou faire d'autres choix de rendu, en particulier lorsque les langages avec des traditions d'écriture distinctes utilisent les mêmes caractères.

## 8. Changements par rapport à la RFC 4646

Le principal objectif de la présente révision de la RFC 4646 était d'incorporer deux nouvelles parties de la norme ISO 639 (ISO 639-3 et ISO 639-5) et leur ensemble de codes de langage dans le registre IANA des sous étiquettes de langage. Cela permet l'identification de beaucoup plus de langages et de collections de langages que précédemment. Les changements spécifiques dans le présent document pour atteindre ces buts sont :

- o Définition de l'incorporation des codes ISO 639-3 et ISO 639-5 à utiliser comme sous étiquettes de langage principal et étendu. Il réserve aussi de façon permanente et déconseille l'utilisation de sous étiquettes "extlang" supplémentaires. Les changements nécessaires pour réaliser cela étaient :
  - \* Modifier les commentaires ABNF.
  - \* Mettre à jour les diverses sections d'exigences d'enregistrement et de stabilité pour référencer ISO 639-3 et ISO 639-5 en plus de ISO 639-1 et ISO 639-2.
  - \* Éditer le texte pour éliminer les références aux sous étiquettes de langage étendu qui ne sont plus utilisées.
  - \* Expliquer les changements dans la section sur les sous étiquettes de langage étendu.
- o Changer l'ABNF relatif aux étiquettes faisant l'objet d'une clause d'antériorité. La liste des étiquettes irrégulière est maintenant établie. Les étiquettes faisant l'objet d'une clause d'antériorité bien formées sont maintenant décrites par la production "langtag", et la production "grandfathered" a par suite été supprimée. La description des deux types d'étiquettes faisant l'objet d'une clause d'antériorité a été ajoutée au paragraphe 2.2.8.
- o Ajout d'un alinéa sur les "collections" au paragraphe 4.1.

- o Changement des règles de majuscules pour les champs "Étiquette" au paragraphe 3.1.
- o Partage du paragraphe 3.1 en sous paragraphes.
- o Modification du paragraphe 3.5 pour permettre d'ajouter, modifier ou supprimer les champs "Suppress-Script" via le processus d'enregistrement. C'était un erratum de la RFC 4646.
- o Modification des exemples qui utilisait le code de région "CS" (anciennement Serbie et Monténégro) pour utiliser "RS" (Serbie) à la place.
- o Modification des règles de création et maintien d'enregistrement des champs "Description" pour empêcher les doublés, y compris les doublés inversés.
- o Suppression dans ce paragraphe de la longue description des raisons de la création de la RFC 4646, qui a aussi causé la suppression de la référence au schéma XML.
- o Modification du texte du paragraphe 2.1 pour mettre l'accent sur le fait que les étiquettes de langage ne sont pas sensibles à la casse.
- o Remplacement de l'exemple "fr-Latn-CA" au paragraphe 2.1 par "sr-Latn-RS" et "az-Arab-IR" parce que "fr-Latn-CA" ne respecte pas le "Suppress-Script" sur "Latn" avec "fr".
- o Changement des exigences de bonne formation pour rendre facultative la vérification de répétition de singleton (elle est exigée pour la vérification de validité) au paragraphe 2.2.9.
- o Changement du texte du paragraphe 2.2.9 qui se réfère à la vérification de protection des droits acquis pour noter que la liste est maintenant incluse dans l'ABNF.
- o Modification et ajout de texte au paragraphe 3.2. La description de tâche a été placée en premier. Une note a été ajoutée pour rendre clair que le réviseur de sous étiquettes de langue peut déléguer diverses tâches non critiques, y compris de modération de liste de diffusion. Finalement, du texte a été ajouté pour préciser le processus de nomination et préciser que les décisions et performances du réviseur peuvent faire l'objet d'un appel.
- o Ajout de texte au paragraphe 3.5 pour préciser que la liste `ietf-languages@iana.org` list est gérée par une personne désignée par l'IESG.
- o Ajout de texte au paragraphe 3.1.5 pour préciser que la première description dans un enregistrement "langage" correspond au nom de référence pour le langage dans ISO 639-3.
- o Modification du paragraphe 2.2.9 pour définir les classes de conformité relatives aux étiquettes spécifiques (anciennement appelées des mises en œuvre "bien formées" et "valides"). Des notes ont été ajoutées sur la suppression de "extlang" de l'ABNF de la RFC 4646, permettant d'être bien formé en utilisant cette ancienne définition. La référence à la bonne formation de la RFC 3066 a aussi été ajoutée.
- o Ajout de texte à la fin du paragraphe 3.1.2 notant que de futures versions de ce document pourraient ajouter de nouveaux types de champs au format de registre et recommandant que les mises en œuvre ignorent tous les champs non reconnus.
- o Ajout de texte sur ce que signifie l'absence d'un champ "Suppress-Script" dans un enregistrement au paragraphe 3.1.9.
- o Ajout de texte permettant la correction d'erreurs d'orthographe et typographiques au paragraphe 3.1.5.
- o Ajout de texte au paragraphe 3.1.8 pour interdire les conflits de champ "Préfixe" (comme des références circulaires de préfixes).
- o Modification du texte au paragraphe 3.5 pour exiger que le réviseur de sous étiquette annonce sa décision (ou extension) suite à la période de deux semaines. Précisé aussi que toute décision ou manque à décider peut faire l'objet d'un appel.
- o Modification du texte au paragraphe 4.1 pour inclure les principes directeurs (jusque là anecdotiques) du choix d'étiquette, et précision de la non utilisation de sous étiquettes de script dans les applications non écrites.
- o Interdiction de plusieurs utilisations de la même variante dans une étiquette (c'est-à-dire, "de-1901-1901"). Précédemment, c'était seulement une recommandation ("DEVRAIT").

- o Suppression du langage inapproprié de la [RFC2119] de l'illustration du paragraphe 4.4.1.
- o Remplacement de l'exemple déconseillant "zh-guoyu" avec "zh-hakka"->"hak" au paragraphe 4.5, notant que c'est ce document qui a causé le changement.
- o Remplacement de l'alinéa du paragraphe 4.1 qui traite de "mul"/"und" pour inclure les sous étiquettes 'zxx' et 'mis', ainsi que l'étiquette "i-default". Une référence normative à la RFC 2277 a été ajoutée.
- o Ajout de texte au paragraphe 3.5 pour préciser que toute modification d'une demande d'enregistrement doit être envoyée à la liste <ietf-languages@iana.org> avant soumission à l'IANA.
- o Changement de l'ABNF pour le format record-jar utilisant la production LWSP pour utiliser une production d'espace de saut à la ligne similaire à obs-FWS dans la [RFC5234]. Cela empêche effectivement les lignes blanches involontaires dans un champ.
- o Texte précisé et révisé aux paragraphes 3.3, 3.5, et 5.1 pour préciser que le réviseur de sous étiquettes de langage envoie les formulaires d'enregistrement complets à l'IANA, que l'IANA extrait l'enregistrement du formulaire, et que le formulaire doit aussi être archivé séparément du registre.
- o Ajout de texte à la Section 5 exigeant de l'IANA qu'elle envoie une annonce à la liste ietf-languages-announcements chaque fois que le registre est mis à jour.
- o Modification du registre pour utiliser l'UTF-8 comme codage de caractères. Cela englobe aussi des instructions supplémentaires à l'IANA et au réviseur des sous étiquettes de langage dans le processus d'enregistrement.
- o Modification des règles du paragraphe 2.2.4 afin que les codes "exceptionnellement réservés" de ISO 3166-1 autres que "UK" soient inclus dans le registre. En particulier, cela permet que le code "EU" (Union Européenne) soit utilisé pour former des étiquettes de langage ou (plus couramment) pour les applications qui utilisent le registre pour des codes de région à référencer dans cette sous étiquette.
- o Modification de la section des considérations relatives à l'IANA (Section 5) pour retirer le langage normatif non utile de la [RFC2119].

## 9. Références

### 9.1 Références normatives

- [ISO15924] Organisation Internationale de normalisation, norme ISO 15924:2004. "Information et documentation -- Codes pour la représentation des noms de scripts", janvier 2004.
- [ISO3166-1] Organisation Internationale de normalisation, norme ISO 3166-1:2006. "Codes pour la représentation des noms de pays et leurs subdivisions -- Partie 1: Codes de pays", novembre 2006.
- [ISO639-1] Organisation Internationale de normalisation, norme ISO 639-1:2002. "Codes pour la représentation des noms de langages -- Partie 1 : Code Alpha-2", juillet 2002.
- [ISO639-2] Organisation Internationale de normalisation, norme ISO 639-2:1998. "Codes pour la représentation des noms de langages -- Partie 2 : Code Alpha-3", octobre 1998.
- [ISO639-3] Organisation Internationale de normalisation, norme ISO 639-3:2007. "Codes pour la représentation des noms de langages -- Partie 3 : Code Alpha-3 pour la couverture complète des langages", février 2007.
- [ISO639-5] Organisation Internationale de normalisation, norme ISO 639-5:2008. "Codes pour la représentation des noms de langages -- Partie 5 : Code Alpha-3 pour les familles et groupes de langages", mai 2008.
- [ISO646] Organisation Internationale de normalisation, norme ISO/CEI 646:1991, "Technologie de l'information -- Jeu de caractères codés sur 7 bits pour l'échange d'informations", 1991.
- [RFC2026] S. Bradner, "Le processus de [normalisation de l'Internet](#) -- Révision 3", ([BCP0009](#)) octobre 1996. (*Remplace [RFC1602](#), [RFC1871](#)*) (*MàJ par [RFC3667](#), [3668](#), [3932](#), [3979](#), [3978](#), [5378](#), [6410](#), [8179](#)*)

- [RFC2119] S. Bradner, "[Mots clés à utiliser](#) dans les RFC pour indiquer les niveaux d'exigence", BCP 14, mars 1997. (MàJ par [RFC8174](#))
- [RFC2227] J. Mogul, P. Leach, "Simple [mesure du nombre d'accès](#) et limitation d'usage pour HTTP", octobre 1997. (P.S.)
- [RFC3339] G. Klyne, C. Newman, "[La date et l'heure sur l'Internet](#) : horodatages", juillet 2002. (P.S.)
- [RFC4647] A. Phillips, M. Davis, "[La correspondance des étiquettes de langues](#)", [BCP0047](#), septembre 2006. (Remplace [RFC3066](#))
- [RFC5226] T. Narten et H. Alvestrand, "Lignes directrices pour la rédaction d'une section Considérations relatives à l'IANA dans les RFC", BCP 26, mai 2008. (Remplace [RFC2434](#) ; remplacée par [RFC8126](#))
- [RFC5234] D. Crocker, éd., P. Overell, "[BNF augmenté pour les spécifications de syntaxe](#) : ABNF", janvier 2008. ([STD0068](#))
- [SpecialCasing] The Unicode Consortium, "Unicode Character Database, Special Casing Properties", mars 2008, <<http://unicode.org/Public/UNIDATA/SpecialCasing.txt>>.
- [UAX14] Freitag, A., "Unicode Standard Annex #14: Line Breaking Properties", août 2006, <<http://www.unicode.org/reports/tr14/>>.
- [UN\_M.49] Nations Unies, Division statistiques, "Codes standard de pays ou zones pour utilisation statistique", révision 4 (publication des Nations Unies, n° 98.XVII.9, juin 1999.
- [Unicode] Unicode Consortium, "The Unicode Standard, Version 5.0", Boston, MA, Addison-Wesley, 2003. ISBN 0-321-49081-0, janvier 2007.

## 9.2 Références pour information

- [CLDR] "The Common Locale Data Repository Project", <<http://cldr.unicode.org>>.
- [iso639.prin] ISO 639 Joint Advisory Committee, "Working principles for ISO 639 maintenance", mars 2000, <[http://www.loc.gov/standards/iso639-2/iso639jac\\_n3r.html](http://www.loc.gov/standards/iso639-2/iso639jac_n3r.html)>.
- [record-jar] Raymond, E., "The Art of Unix Programming", 2003, <urn:isbn:0-13-142901-9>.
- [RFC1766] H. Alvestrand, "Étiquettes pour l'identification des langues", mars 1995. (Obsolète, voir [RFC3066](#), [3282](#)) (P.S.)
- [RFC2028] R. Hovey, S. Bradner, "Les [organisations impliquées dans le processus](#) de normalisation de l'IETF", octobre 1996. (MàJ par [RFC3668](#), [RFC3979](#)) ([BCP0011](#))
- [RFC2046] N. Freed et N. Borenstein, "[Extensions de messagerie Internet](#) multi-objets (MIME) Partie 2 : Types de support", novembre 1996. (D. S., MàJ par [2646](#), [3798](#), [5147](#), [6657](#), [8098](#))
- [RFC2047] K. Moore, "MIME ([Extensions de messagerie Internet](#) multi-objets) Partie trois : extensions d'en-tête de message pour texte non ASCII", novembre 1996. (MàJ par [RFC2184](#), [RFC2231](#)) (D.S.)
- [RFC2231] N. Freed, K. Moore, "Extensions MIME [Valeur de paramètre et Mot codé](#) : jeux de caractères, langages, et continuations", novembre 1997. (P.S.)
- [RFC2616] R. Fielding et autres, "[Protocole de transfert hypertexte](#) -- HTTP/1.1", juin 1999. (D.S., MàJ par [2817](#), [6585](#))
- [RFC2781] P. Hoffman et F. Yergeau, "[UTF-16](#), un codage de la norme ISO 10646", février 2000.
- [RFC3066] H. Alvestrand, "Étiquettes pour l'identification des langues", BCP 47, janvier 2001. (Obsolète, voir la [RFC4646](#).)
- [RFC3282] H. Alvestrand, "[En-têtes de langage de contenu](#)", mai 2002. (D.S.)
- [RFC3552] E. Rescorla, B. Korver, "Lignes directrices pour la rédaction d'une section de considérations sur la sécurité dans les RFC", juillet 2003. ([BCP0072](#))

- [RFC3629] F. Yergeau, "[UTF-8, un format de transformation](#) de la norme ISO 10646", STD 63, novembre 2003.
- [RFC4645] D. Ewell, "Registre de sous étiquette de langage initial", septembre 2006. (*Information*)
- [RFC4646] A. Phillips, M. Davis, "[Étiquettes d'identification des langues](#)", [BCP0047](#) septembre 2006. (*Remplacée par RFC5646*)
- [RFC5645] D. Ewell, "Mise à jour du registre des sous étiquettes de langages", septembre 2009. (*Information*)
- [UTS35] Davis, M., "Unicode Technical Standard #35: Locale Data Markup Language (LDML)", décembre 2007, <<http://www.unicode.org/reports/tr35/>>.

## Appendice A. Exemples d'étiquettes de langages (information)

Sous étiquette de langage simple :

de (allemand)

fr (français)

ja (japonais)

i-enochian (exemple d'étiquette faisant l'objet d'une clause d'antériorité)

Sous étiquette de langage plus sous étiquette de script :

zh-Hant (chinois écrit en utilisant l'écriture traditionnelle chinoise)

zh-Hans (chinois écrit en utilisant l'écriture chinoise simplifiée)

sr-Cyrl (serbe écrit en utilisant l'écriture cyrillique)

sr-Latn (serbe écrit en utilisant l'écriture latine)

Sous étiquette de langage étendu et sa contrepartie de sous étiquette de langage principal :

zh-cmn-Hans-CN (chinois, mandarin, écriture simplifiée, comme utilisé en Chine)

cmn-Hans-CN (mandarin chinois, écriture simplifiée, comme utilisé en Chine)

zh-yue-HK (chinois, cantonais, comme utilisé à Hong Kong SAR)

yue-HK (cantonais chinois, comme utilisé à Hong Kong SAR)

Région d'écriture de langage :

zh-Hans-CN (chinois écrit en utilisant l'écriture simplifiée comme utilisé en Chine continentale)

sr-Latn-RS (serbe écrit en utilisant l'écriture latine comme utilisée en Serbie)

Variante de langage :

sl-rozaj (dialecte résien du slovène)

sl-rozaj-biske (dialecte San Giorgio du résien dialecte du slovène)

sl-nedis (dialecte nadiza du slovène)

Variante régionale de langage :

de-CH-1901 (allemand comme utilisé en Suisse en utilisant la variante 1901 [d'orthographe])

sl-IT-nedis (slovène comme utilisé en Italie, dialecte Nadiza)

Variante régionale d'écriture de langage :

hy-Latn-IT-arevela (arménien oriental écrit en écriture latine, comme utilisé en Italie)

Région de langage :

de-DE (allemand d'Allemagne)

en-US (anglais comme utilisé aux États Unis)

es-419 (espagnol approprié pour la région Amérique Latine et Caraïbes utilisant le code de région des UN)

Sous étiquettes d'utilisation privée :

de-CH-x-phonebk

az-Arab-x-AZE-derbend

Valeurs du registre d'utilisation privée :

x-n'importequoi (utilisation privée utilisant le singleton 'x')

qaa-Qaaa-QM-x-southern (toutes des étiquettes privées)

de-Qaaa (allemand, avec une écriture privée)

sr-Latn-QM (serbe, écriture latine, région privée)  
sr-Qaaa-RS (serbe, écriture privée, pour la Serbie)

Étiquettes qui utilisent des extensions (SEULEMENT pour exemple -- les extensions DOIVENT être définies par une révision ou mise à jour du présent document, ou par une RFC) :

en-US-u-islamcal  
zh-CN-a-myext-x-private  
en-a-myext-b-another

Étiquettes invalides :

de-419-DE (deux étiquettes de région)  
a-DE (utilise une sous étiquette d'un seul caractère en première position ; noter que quelques étiquettes faisant l'objet d'une clause d'antériorité qui commencent par "i-" sont valides)  
ar-a-aaa-b-bbb-a-ccc (deux extensions avec le même préfixe d'une seule lettre)

## Appendice B. Exemples de formulaire d'enregistrement

Formulaire d'enregistrement de sous étiquette de langage

1. Nom du demandeur : Han Steenwijk
2. Adresse de demandeur : han.steenwijk @ unipd.it
3. Enregistrement demandé :
  - Type : variante
  - Sous étiquette : biske
  - Description : Dialecte San Giorgio de résien
  - Description : Dialecte Bila de résien
  - Préfixe : sl-rozaj
  - Commentaires : Le dialecte de San Giorgio/Bila est un des quatre dialectes locaux majeurs de résien.
4. Signification prévue de la sous étiquette : La variété locale de résien est parlée à San Giorgio/Bila.
5. Référence à une description publiée du langage (livre ou article) : Jan I.N. Baudouin de Courtenay - Opyt fonetiki rez'janskich govorov, Varsava - Peterburg: Vende - Kozancikov, 1875.

Formulaire d'enregistrement de sous étiquette de langage

1. Nom du demandeur : Jaska Zedlik
2. Adresse de demandeur : jz53 @ zedlik.com
3. Enregistrement demandé :
  - Type : variante
  - Sous étiquette : tarask
  - Description : Biélorusse en orthographe taraskievica
  - Préfixe : be
  - Commentaires : La sous étiquette représente l'orthographe biélorusse de Branislau Taraskievic telle que publiée dans "Bielaruski klasycny pravapis" de Juras Buslakou, Vincuk Viacorka, Zmicier Sanko, et Zmicier Sauka (Vilnia-Miensk 2005).
4. Signification prévue de la sous étiquette : La sous étiquette est destinée à représenter l'orthographe biélorusse comme publiée dans "Bielaruski klasycny pravapis" de Juras Buslakou, Vincuk Viacorka, Zmicier Sanko, et Zmicier Sauka (Vilnia-Miensk 2005).
5. Référence à une description publiée du langage (livre ou article) : Taraskievic, Branislau. Bielaruskaja gramatyka dla skol. Vilnia: Vyd. "Bielaruskaha kamitetu", 1929, 5<sup>ème</sup> édition. Buslakou, Juras; Viacorka, Vincuk; Sanko, Zmicier; Sauka, Zmicier. "Bielaruski klasycny pravapis". Vilnia-Miensk, 2005.
6. Autres informations pertinentes : Le biélorusse en orthographe Taraskievica est devenue de large utilisation, en particulier sur le segment des biélorussophones de l'Internet, mais à côté de cela des livres et des journaux sont aussi publiés qui utilisent cette orthographe du biélorusse.

## Appendice C. Remerciements

Toute liste de contributeurs est forcément incomplète ; prière de regarder ce qui suit comme seulement une sélection parmi le groupe des personnes qui ont contribué à faire de ce document ce qu'il est aujourd'hui.

Les contributeurs aux RFC 4646, RFC 4647, RFC 3066, et RFC 1766, les prédécesseurs du présent document, ont fait d'énormes contributions directement ou indirectement au présent document et sont généralement responsables du succès des étiquettes de langage.

Les personnes suivantes ont contribué au présent document : Stephane Bortzmeyer, Karen Broome, Peter Constable, John Cowan, Martin Duerst, Frank Ellerman, Doug Ewell, Deborah Garside, Marion Gunn, Alfred Hoenes, Kent Karlsson, Chris Newman, Randy Presuhn, Stephen Silver, Shawn Steele, et beaucoup, beaucoup d'autres.

Des remerciements particuliers sont dus à Harald Tveit Alvestrand, qui a produit les RFC 1766 et 3066, et sans qui le présent document n'aurait pas été possible.

Un merci particulier à Michael Everson, qui a servi de réviseur des étiquettes de langage pendant presque toute la période de la RFC 1766/RFC 3066, ainsi que de réviseur des sous étiquettes de langage depuis l'adoption de la RFC 4646.

Un merci particulier aussi à Doug Ewell, pour sa production du premier registre complet des sous étiquettes, pour son travail de soutien et de maintien des nouveaux enregistrements, et son édition soigneuse des [RFC4645] et [RFC5645].

## Adresse des auteurs

Addison Phillips (éditeur)  
Lab126  
mél : [addison@inter-locale.com](mailto:addison@inter-locale.com)  
URI : <http://www.inter-locale.com>

Mark Davis (éditeur)  
Google  
mél : [markdavis@google.com](mailto:markdavis@google.com)