

言語資源メタデータデータベース SHACHI
データ登録マニュアル

080122 版

【はじめに】

- 2008 年 1 月 22 日より, SHACHI の仕様が改訂になりました. 新しく設置された項目は, **coverage**, **spatial** のみです. また, メタ項目名, 並び順が変更になっていますが, 登録内容(メタ項目の定義)はほとんど変更ありませんのでご安心下さい.
- 登録の際は, この登録マニュアル, および, SHACHI 公式ホームページ <http://shachi.org/> の [SHACHI の概要]→左欄 SHACHI の仕様を参照して下さい.
- データ登録は, <http://shachi.org/edit/> の New registration (言語資源登録フォーマット) を利用して下さい.
- 入力フォーム左端の項目をメタデータ項目(メタ項目)と呼びます.
- 言語資源登録フォーマットは,
 - ①選択欄 (下矢印ボタンで選択項目(=値)が表示される)
 - ②記入欄 (自由に書き込みができる)の 2 種類があります. 基本的には, 両方を埋めて下さい.
- 登録手順
 - 1. 選択欄の該当する項目(=値)を選択
選択欄の項目に該当するものがあれば選択する.
 - 2. 記入欄 に詳細な情報を記載
言語資源の公式ページ, 既存のカタログなどから詳細な情報を記入.
 - アルファベット, 数字, 記号(括弧, 中点など)は半角で入力する.
日本語入力の場合, カタカナは全角を使用.
 - 記入事項が複数ある場合は, **+ボタン**を押し, 行を増やして入力
(旧入力フォームの縦棒“|”(改行)と同じ意味です.)
例) 旧 Spanish|English
新 + Spanish
+ English
※上記+は, +ボタンをさしています.
- SHACHI では, 多様な言語資源の中でも, コーパス, 辞書, シソーラス, 語彙リスト (glossary=lexicon)を収録対象としています. 論文1本, 電子図書1冊などは, 対象外です.
- 辞書・シソーラス・語彙リスト (コーパス)を登録の際には, カラーのメタ項目は入力不要です.
- カラー項目以外は入力必須. 但し, "relation.utilization"に関しては入力不要です. (完全自動

収集欄です.)

■ **Title** : 言語資源(言語資源)の公式な名称

言語資源名の補完機能が搭載されました。タイトルの一部を打ち込むと、一致するものが下に表示されます。

■ **Title. alternative** : 言語資源の別名、サブネーム、短縮名、翻訳名（公式な英語名に対するフランス語名など）。別名の表記が複数ある場合は、+ボタンで行数を増やし、あるだけ記載。

例) フランス語を収録したコーパスで、フランス語名と英語名がある場合は

公式ページの一番トップに書いて在る方を“Title”に入力。

例) **BNC Corpus**

British National Corpus

公式ページの一番トップに書いて在る方を“Title”に入力。

言語資源名の補完機能が搭載されました。タイトルの一部を打ち込むと、一致するものが下に表示されます。

■ **creator** : 言語資源の内容に主たる責任を持つ人や、組織などのエンティティ（「通常その名前を識別に用いる」とされているが、名前＝エンティティではないので注意）。

実際に言語資源を作った人、研究機関。URL やメールアドレスがあれば、それも記載。

研究所が作成したものであれば、プロジェクトリーダーは研究者であると考えられるので researcher を選択。企業による作成であれば develop の選択を原則とする。

■ **subject** : 言語資源の内容に含まれるトピック(主題)。トピックを示すキーワードやフレーズ。言語資源のタイトルのもう1つ上の概念にあたるような、メタ的なタイトル、キーワードを探して記入。公式ホームページのトップにある、キーワード(主題)を拾う。タイトル自体が、一般的な名前の場合は、タイトルと同じものが入る場合もあります。

例) 新聞記事データベース、話し言葉コーパス、バイリンガルコーパス

例) 対話コーパス

■ **subject.linguisticField**: 言語資源の言語学的な分類です。その言語資源が何の目的のために収録されたかにもよる。判断できなければ、未選択でよい。たとえば、人間の言語認知のための研究で作成された言語資源であれば、cognitive science、言語処理技術の開発ために作成されたものであれば、computational linguistics とする。

[用語の説明]

Linguistic Field	言語分野
anthropological_linguistics	言語人類学
applied_linguistics	応用言語学(教育への応用など)
cognitive_science	認知科学
computational_linguistics	コンピューター言語学
discourse_analysis	談話分析
forensic_linguistics	法言語学
general_linguistics	一般言語学
historical_linguistics	歴史言語学
history_of_linguistics	言語学の歴史
language_acquisition	言語獲得
language_documentation	言語ドキュメンテーション
lexicography	辞書編纂
linguistics_and_literature	言語学と文学
linguistic_theories	言語理論
mathematical_linguistics	数理言語学
morphology	形態論
neurolinguistics	神経言語学
philosophy_of_language	言語の原理
phonetics	音声学
phonology	音韻論
pragmatics	語用論
psycholinguistics	心理言語学
semantics	意味論
sociolinguistics	社会言語学
syntax	統語論
text_and_corpus_linguistics	テキストとコーパス言語学
translating_and_interpreting	翻訳と通訳
typology	類型論
writing_systems	書記体系

- **subject.monoMultilingual**: 言語資源が単言語か多言語かを選択.
- **subject.resourceSubject**: コーパスか, 辞書か, シソーラスか, もしくは語彙リストかを記入.
 ※glossary : 旧 lexicon. glossary>dictionary
- **description**: その言語資源の説明を 1~5 文程度に要約. 公式ホームページのトップページに記載されていることが多い.
- **description.price**: 言語資源がすでに公開・配布されている場合, 有料であれば, その価格を記載. 無料であれば”フリー”と記載.
- **description.language**: 言語資源の仕様書(マニュアル)の言語.
- **publisher**: 言語資源の配布窓口. この言語資源を利用可能にすることに責任を持つエンティティ. 個人の場合もあれば, 組織やサービスの場合もある(「通常その名前」とされているが, creator と同様, 名前=エンティティではないので注意). メールアドレス, URL などがあれば, それらも記載.
- **contributor**: 言語資源の内容に協力, 貢献している人や組織. 言語資源構築にあたる資金を提供した期間など. メールアドレス, URL などがあれば, それらも記載.

選択欄で該当の項目(値)を選択。ただし、speaker と author に関しては、以下のピンク色の欄を用いる。

※以下ピンク蛍光ペンの欄は、上記 contributor の選択欄の中にある、speaker , author に関する情報(すなわち、旧 sender 情報)に関する登録です。登録する言語資源が、コーパスの際に、入力が必要です。場合によっては glossary も対象となります。

- (contributor.attribute.speakerAuthor) : 書き言葉の場合は、書いた人、話し言葉の場合は、話し手がどういう人であることを記入してください。
 - speaker: 言語資源が話し言葉を収録している場合に選択、かつ、自由記入欄に、その人がどういう人かを記載。
例) 英語学習者
 - author: 言語資源が書き言葉を収録している場合に選択、かつ、自由記入欄に、その書き手がどういう人かを記載。
例) 新聞記者・雑誌記者・エッセイスト
- (contributor.attribute.motherTongue) : 特別なコーパスでなければ、native を選択して下さい。
 - ◇ nonnative : コーパスによっては、ノンネイティブ話者の発話分析を目的に作られたものや、ネイティブとノンネイティブを混ぜているものなどがあります。その場合、*nonnative を選択します。そういった情報が提示されていれば、その人数割合、国籍情報などの詳細を記入欄の方へ記入して下さい。
例) “日本人がしゃべった英語”を分析などの目的で収録しているコーパスは、nonnative を選択して下さい。詳細情報は記入欄の方に記入して下さい。
- (contributor.attribute.intonation) : 言葉の発信者が標準語を話す人であるか、否かを記載。方言を収録しているような言語資源以外は、standard dialect(標準語)として下さい。この場合、選択項目を選択するだけでよいです。方言の場合は、どのような方言であるかを記載して下さい。
 - ◇ standard dialect : 標準語
 - ◇ dialect : 方言、なまり
- (contributor.attribute.level) : 言語資源の言語の精度を示す。訓練を受けた人の言語が収録されている場合は、その職業を記載。特に訓練を受けていない一般の人であれば、“一般”とする。また、英語学習者など、特別な立場の人の言語である場合は、“英語学習者”などと、具体的に記入する。
 - ◇ professional : 例えば、話し言葉を収録した言語資源の場合

合、話者がアナウンサーであれば、professional とする。書き言葉を収録した場合、記者などが書いた言語資源を収録している場合は professional とする。

◇ **amateur** : 一般の人の話し言葉、書き言葉を収録している場合は、amateur を選択。

- **(contributor.attribute.age)**: どういう年齢の人たちの言葉を収録したものであるかを記載。

例) 20 代, 13~15 歳

子供の発話を収録している場合には、通常、その説明があります。したがって、特に子供の発話を収録したという情報が無い限りは、大人を選択してください。中学生以上を大人とすることにします。12 歳以上大人。

◇ **adult**: 通常は adult となります。

◇ **child**: 幼児言語発達など、あえて子供の言語を収録している場合は child を選択して下さい。

- **(contributor.attribute.gender)**: 女性の発話か、男性の発話かといった記載が無い場合は、通常は、mix となります。もし、収録の男女別人数などの情報がある場合には、自由記入欄(選択欄の左)にその割合、人数などを記述して下さい。

◇ **male** : 発信者が男性のみから成る言語資源

◇ **female** : 発信者が女性のみから成る言語資源

◇ **male & female** : 発信者が男性・女性で構成される言語資源男女割合などの情報が無い場合で、明らかに男女混合で構成されていると考えられる場合は、*male & female を選択。

- **date.created** : 言語資源が制作された年月日。yy/mm/dd で記載。

例) 1999-2003, 03/07/22, 2005 Dec.

- **date.issued** : 言語資源が配布された年月日。yy/mm/dd で記載。

例) 2007, 02/08/11

- **date.modified** : 言語資源が改訂された年月日

- **type**: 言語資源のタイプをさします。今回の収集では、Text(テキストデータ)、Sound(音声データ)、Image(動画・静止画)のいずれか、もしくはその組み合わせが入ります。言語資源、もしくは、コーパスといった場合、その多くは text だと思いますが、音声のみからなる言語資源・コーパスもあります。例えば、ある会議の音声を記録として録音していたものを、そのまま(音声)コーパスとして、公開されているものも存在します。

- **type.discourseType** : ダブリン・コアに基づく分類方法です。当てはまるものがあれば選択しどれにも当てはまらなければ、無理に選択しなくて構いません。

Discourse Type	談話タイプ
drama	演劇
formulaic_discourse	定型の談話
interactive_discourse	双方向の談話
language_play	言語 芝居
oratory	演説
narrative	物語
procedural_discourse	手続きの談話
report	報告
singing	唱歌
unintelligible_speech	不明瞭な発話

[用語の詳細説明]

drama (演劇) : 二人以上で演じるために制作された創作劇や寸劇の談話(会話)。

formulaic_discourse : (定型の談話) : 儀式や慣習的形式に則った談話。

例) 祈り, 祝福, 呪文, 結婚式の誓約の言葉など定型の文言。

interactive_discourse : (双方向の談話) : 二人以上で交わされる言葉のやりとり。

例) 会話, インタビュー, 対話, 挨拶や暇乞いの言葉など。

language_play : (言語芝居) : 言語構造の機能に着目し, それを創造的に駆使させた戯曲や言葉の表現法。

例) 言葉を媒介にした芸術, ジョーク, 秘語, 逆さ語など。

oratory(演説) : 人前での談話, 規則・慣例に従った弁論。

例) 訓話, 講演, 政治演説, 招詞。

narrative(物語) : 一人での語り。

例) 史実や口伝話, 身の上話, 神話, 民話, 寓話, ユーモラスな物語話し聞かせ。

procedural discourse : (手続きの談話) : 方法, 手順, 決められた順序がある状況での説明や描写。

例) 調理法や使用説明書, 計画の説明。

report(報告) : ある出来事や状況に関する事実説明。

例) ニュース報道, 随筆, 評論。

singing(唱歌) : メロディに合わせてたり, 抑揚をつけたりした一連の言葉や明晰な音声。

例) 詠唱, 歌曲, コーラスの歌詞。

unintelligible_speech : (不明瞭な発話) : 普通の言語に訳せない発話。

例) 宗教的儀式で発されるわけの分からない発話や歌詞の中の特に意味を持たない言葉(ファ・ラ・ラ)。

■ **type.linguisticType**: 言語資源の性質, 特徴.

- ◇ **lexicon (語彙リスト)**: 語彙品目を体系的に収録したもの. 語彙品目は, 定義, 関連項目の記述, 他の語彙品目との意味的関連性の表示を含むことがある.

例) ワードリスト, シソーラス, ワードネット, フレームネット, 様々な辞書(バイリ

ンガルや多言語辞書, 用語辞典, 固有名詞辞典の他に, 成句辞書や世界歌

曲リストなど.

- ◇ **primary_text (一次テキスト)**: 言語資源に含まれる言語素材そのものが研究対象となるもの.

例) 行間翻訳付きテキスト, 〈録音テープの〉書き起こし, ビデオ撮りした儀式や物語の読み聞かせ, 研究対象となる言葉の構造・利用法を例示した話し言葉や書き言葉を記録したものなど.

- ◇ **language_description (言葉の記述)**: 言語構造を体系的に文書化することで, 言語やアスペクトを記述したもの.

例) 形式文法, 記述文法, フィールドノート, 音韻記述.

■ **type.purpose**: 言語資源を構築した目的.

例) 名古屋大学同時通訳データベースの場合, “音声機械翻訳技術の開発”.

- ◇ **lexicography**
- ◇ **analysis**
- ◇ **developing technologies**
- ◇ **education**

■ **type.style**: 言語には, 大きく分けて, 新聞や雑誌などのように, 人間が書いた言語(書き言葉)と, 間が講演や対話などのように, 話した言語(話し言葉)の2種類に大別されます.

まず, 選択欄で, 言語資源が書き言葉か, 話し言葉かを選択します. また, それらがどういうデータであるかを, 記入欄に入力します.

例) 学会講演, 模擬講演, 新聞記事, 英和辞書, 特許文書など.

■ **type.form**: 収録されている言語資源の文が定型文か非定型文かを記載. 特記事項があれば記入欄に記入.

- ◇ **fixed form**: 例えば, 収録されている言語資源が5単語以内の文, 疑問文, 単文など, 限定した文を収集している場合

また, 旅行対話, 機械と人間の対話,

例) 機械: 「お引き出しですか? 振込みですか?」

人間: 「振込み」

など, しゃべることが極めて限られているようなものを収録している場合.

法令文も含む. (書き方がある程度決められた性質のものであるため.)

◇ **unfixed form**: 非定型文. 新聞データベース, 同時通訳データベースなど, データに含まれる文が様々であるため.

■ **type.sentence**: 収録されている言語資源の文の長さに関する情報を記載する.

- ◇ **short sentences**: 短文ばかりを集めた言語資源 (医学用語リストなど)
- ◇ **long sentences**: 長文ばかりを集めた言語資源
- ◇ **mix**: 短文, 長文が入り混じっている言語資源

■ **type.annotation**: 言語資源に何らかのタグが付与されていれば **annotated** を選択. プレーン (アノテーションが無い) 言語資源であれば, **plain** を選択.

■ **type.annotationSample**: アノテーションされているタグセット, および, タグ情報を記載.

現在, 約 10 行強を貼って頂いていると思いますが, 20 行に変更します. さらに 1 行あけ, The rest is omitted. (以下省略) と書いてください. その情報の URL が分かる場合は, 必ず, これまでと同様, 記載しておいてください.

例)

=====

(サンプル 20 行ぐらい)

(1 行あける)

The rest is omitted.

(1 行あける)

ULR

=====

■ **type.sample**: 言語資源のサンプルが掲載されていたら, それをペーストする. (20 行ぐらい)
タグ情報などが分かるものがよい. 著作権に注意する. 約 20 行を貼り付け. さらに 1 行あけ, The rest is omitted. (以下省略) 記載. その情報の URL が分かる場合は, 記載.

例)

=====

(サンプル 20 行ぐらい)

(1 行あける)

The rest is omitted.

(1 行あける)

ULR

=====

■ **format. extent**: データのサイズもしくは収録時間, 話者人数, 単語数など. できるだけ詳細に記載. Sound のコードも反映してください.

例: coding: A-law, 8 bit, 8 kHz, no compression

既に登録済みの分に関しましては, 修正しなくて結構です. 但し, データクリー

ニング(リチェック)指示があった場合には、本項目も必ず、修正してください。

■ **format. medium** : 言語資源の搬送媒体. 例) CD-ROM, DVD, Excel ファイル

■ **format. encoding** : 文字コード(Character Set, character code)

文字コード【character code】を記載する箇所です。データベース構築において、重要な情報です。注意して探してください！主に、下記のコードがあります。

ASCII, JIS-xxx, GB-xxx, KS-xxx, Big5, CNS xxxx, ISO 2022 – ISO/IEC 2022, EUC-xx, Shift_JISxx, Unicode, Adobe-Japan1-0, UTF-8, CP869, DTD

※ 文字コード【character code】文字や記号をコンピュータで扱うために、文字や記号一つ一つに割り当てられた固有の数字のこと。欧米などで使われる英数字の文字コードは 1 バイトで表現されるが、これでは 256 文字しか表現できず、日本語の漢字などは収録できないため、日本や韓国、中国などではそれぞれ独自に 2 バイト(最大 65536 文字)のコード体系を定め、それぞれ自国内で使用している。1 バイトの文字コードとしては ASCII が世界標準となっているが、一部の大型コンピュータでは IBM 社の EBCDIC を使うものもある。日本語の文字コードは、JIS(日本工業規格)で標準化された JIS コードの他、主に UNIXなどで使われる EUC、Windows や Mac OSなどで使われるシフト JIS の 3 種類が利用されている。

<http://e-words.jp/w/E69687E5AD97E382B3E383BCE38389.html>

インターネットエクスプローラの表示→エンコード→その他 と辿っていくと、文字コードのリストが出てきますので参考にしてください。それらが文字コードの種類です。

■ **format. markup** : マークアップ言語／記述言語(markup language)

データベース構築において、重要な項目です。注意して探してください。

代表的なマークアップ言語 ※発見するコツ: ショートカットキーCtrl+Fで "ML"を検索

- SGML(Standard Generalized Markup Language)
- HTML(HyperText Markup Language) SGML から発展した
- TeX
- XML(eXtensible Markup Language)
- XHTML (Extensible HyperText Markup Language)

※ マークアップ言語【markup language】: 文書の一部を「タグ」と呼ばれる特別な文字列で囲うことにより、文章の構造(見出しやハイパーリンクなど)や、修飾情報(文字の大きさや組版の状態など)を、文章中に記述していく記述言語。マークアップ言語を使用して書かれた文書はテキストファイルになるため、テキストエディタを使って人間が普通に読むことが可能で、もちろん編集もできる。代表的なマークアップ言語としては、SGML(Standard Generalized Markup Language)、SGML から発展した HTML(HyperText Markup Language)、TeX などがある。ちなみに、XML(eXtensible

Markup Language)はタグを使用した囲み方は規定しているが、タグの意味は規定していない、メタ言語と呼ばれるものである。メタ言語はタグの意味を規定するスキーマ言語と組み合わせてマークアップ言語になる。

<http://e-words.jp/w/E3839EE383BCE382AFE382A2E38383E38397E8A880E8AA9E.html>

主なマークアップ言語／記述言語(markup language)

- **format.functionality** : 言語資源に付属しているツール。ツール名は無いが、こういうツールがあるという記載がある場合は、その部分を要約して下さい。また、よく漠然としている場合は、URL を記入しておいて下さい。

例) 構文解析ツール, 単語ソートツール, 音声視覚化ツールなど。

- **identifier** : 言語資源の公式ホームページの URI(Uniform Resource Identifier)/URL を記載。
- **source** : 言語資源が別の言語資源の全体もしくは一部から派生したり導かれたものであるときは、その元言語資源への参照。

例) 日経新聞のデータベースであれば、Source は”日経新聞”と記載。

- **language** : 言語資源が収録対象としている言語。(旧 Subject.language)

例) 「毎日新聞 2000 年収録データ」の場合、“日本語”

例) 日本の会社や、研究機関が構築した英仏辞書の場合、

+英語

+フランス語

- **relation** : 登録している言語資源が関連する言語資源への参照。

選択欄に関する下記の定義において、「言語資源」は登録中の言語資源との親、兄弟に当たる語資源を指します。

言語資源名の補完機能が搭載されました。タイトルの一部を打ち込むと、一致するものが下に表示されます。候補にあがってくる言語資源には、先頭に *SHACHI_ID* が付いています。*ID*は消さずに、そのままにしておいて下さい。

※サブコーパスが多い場合の Relation 記入方法

+マークをクリックして、1 件 1 行で入力します。最大 10 件まで入力して下さい。気力があれば、10 件以上でも全部入力してください。(言語資源名の補完機能をうまく利用してください。)

10 件目のタイトル直後に The rest is omitted(残りのサブコーパスの数)を入力しておてください。

例) 登録している言語資源の他に、27 のサブコーパスがある場合 (すなわち 28 人兄弟)10 件目最後の段 の言語資源名の後ろに The rest is omitted(17)

- ◇ **isVersionOf** : 言語資源はこの要素(登録している言語資源)が示す言語資源の 1 バージョン。
- ◇ **hasVersion** : 言語資源はこの要素(登録している言語資源)が示す言語資源をパー

ジョンとして持つ。

- ◇ **version** は同等のものが入るので、**isVersionOf= hasVersion** ですので、**hasVersion** に統一。
- ◇ **isReplaceBy** : 要素(登録している言語資源)が示す言語資源で置き換えられる。
- ◇ **replaces** : 要素(登録している言語資源)が示す言語資源を置き換える。
version が同じだけど、新しいものが出たので、前のものが死ぬ場合。
- ◇ **isRequiredBy** : 要素(登録している言語資源)が示す言語資源で必要とされる。
- ◇ **requires** : 要素(登録している言語資源)が示す言語資源を必要とする。
- ◇ **isPartOf**: 言語資源は要素(登録している言語資源)が示す言語資源の一部。
- ◇ **hasPart**: 要素(登録している言語資源)が示す言語資源をその一部として持つ。
- ◇ **isReferencedBy**: 要素(登録している言語資源)が示す言語資源から参照、引用されている。
- ◇ **references**: 要素(登録している言語資源)が示す言語資源を参照、引用する。
- ◇ **isFormatOf**: 言語資源は要素(登録している言語資源)が示す言語資源と同じだが、異なるフォーマットによる。
- ◇ **データ内容は同じなのだけど、フォーマット違いがある。**
- ◇ **hasFormat**: 同じ内容だが異なるフォーマットの言語資源を持つ。
- ◇ **conformsTo**: 言語資源が準拠している確たる標準を示す。
例えば、ある有名データベースと全く同じ基準(標準)で構築されている。
例)ある辞書と全く同じじゃないのニュージーランド版が出た。

- **relation.utilization** : 記入しないでください。完全自動登録の実験を行っている欄です。
- **coverage.temporal**: 時間的な対象 言語資源データの時代。yyyy/mm/dd で記載。講演など、現代に行われているものが収録されている場合は、**Date. created** の年代と同じになります。
- **coverage.spatial**: 言語資源の知的内容の空間的な性質。場所名や地理的座標で特定できる場所などを入力します。

例) スイスなど、1つの国の中で地域によって言語が異なる国が複数存在します。特にヨーロッパは、かつての国・帝国のなごりなどから、現在の国境と言語の境界は一致していません。そのような場合、言語資源を収録した場所(言語の話されている地域名)を本項目に入力することで、どこの、どういった言語であるかを特定することができます。

たとえば、愛知県の三河弁を収録している場合、“三河”と入力します。この欄は、国名や言語名を登録するのではなく、地域名を入力します。地域は、住所のような明確な地理名である必要はなく、尾張、三河のような地域名でも構いません。中国のような国土が広い場合、など、地域を特定することで、言語の種類も特定できるため、このメタ項目が有効に働きます。

- **rights (copy rights, access rights, license)**: 言語資源にアクセスできる人、もしくはセキュリティ要件。

Annotation.corpus:

+++++

【主な参考文献】

OLAC Metadata Set

<http://www.language-archives.org/OLAC/olacms.html>

KANZAKI Dublin Core: メタデータを記述するボキャブラリ

<http://www.kanzaki.com/docs/sw/dublin-core.html#reference>

社団法人 電子情報技術産業協会, ヒューマンインターフェース技術に関する調査報告書, pp. 81-83 平成 15 年 4 月.

【作業の円滑化】

■ショートカットキーの利用

Ctrl + C: 選択した項目をコピーする

Ctrl + X: 選択した項目を切り取る

Ctrl + V: 選択した項目を貼り付ける

Ctrl + Z: 操作を元に戻す

Ctrl + Y: 操作をやり直す

- 入力中のデータの保存を頻繁に実施
- 分からない問題は, 早めに問い合わせる

.....

日本語の言語資源の英語登録(日英翻訳登録)

■Title 等の記入方法

Title: 日本語タイトルの英訳

Title.alternative

+日本語タイトル(日本語表記)

+ローマ字に直したもの

■Creator, Contributor などに, 団体名が入る場合

例) 国立国語研究所の場合

The National Institute for Japanese Language (ホームページにたいてい書いてあります.)

どうしても英訳組織名がみつからなければ,

+Kokuritsu-kokugo-kenkyujo 日本語名称 (ローマ字読み)

※本言語資源データベースが一般公開されると, ユーザーは世界中の人が想定されます.

したがって, 英語の情報があることが望まれます.

メール件名の統一

メールでのお問い合わせは、件名を下記のように統一してください。

例) 件名: [SHACHI] タイトル入力の補完機能について