OLIF Guidelines for Formulating Canonical Forms

Susan McCormick
OLIF2 Consortium
October, 2001

Adapted from SAP Terminology Entry Conventions

CONTENTS

1 General

1.1 Which form is the canonical form?

The *canonical form* of a word or phrase is its full, unmarked base form[1]. This is often the form to which inflection is added. Guidelines for formulating canonical forms vary based on several factors, including the grammatical distinctions that hold within a language, and the lexicographic and/or orthographic conventions in place for the language.

1.1.1 The canonical form of a single-word string

Guidelines for formulating the canonical form of a single-word string often depend on part-of-speech designation:

- If the entry is a **noun**, the canonical form is usu. the singular form in the case that denotes the subject of the sentence.

- If the entry is a **verb**, the canonical form is usu. the infinitive form; there is no notation for the infinitive marker.

- If the entry is an **adjective** or **adverb**, the canonical form is usu. the positive form.

The canonical form of a single-word entry string can be formulated as a non-base, inflected form if the inflection is essential to distinguish the meaning of the word:

e.g.:
- Plural noun whose meaning is different from the singular (e.g., *checker* and *checkers*, *America* and *Americas*)
- Collective noun that occurs primarily with plural inflection (e.g., *goods*)

1.1.2 The canonical form of a multiple-word string

General guidelines for formulating the canonical form of a multiple-word string are:

---

[1] The base form that constitutes a canonical form must be a form that may actually occur in the language in isolation, i.e., root or stem forms that are not valid in isolation (e.g., *writ-* as a stem for *write*) are not admissible as canonical forms.

- Identify the head word within the string and use its canonical form when representing the string:

e.g.:

English multiple-word strings with noun heads in canonical form:

bill of material
cost center
maintenance request

- Adhere to standard lexicographical conventions for the given language for ordering the different words within the multiple-word string. (See *Language-specific guidelines.)*

- Omit filler or function words that are not essential to understanding the meaning of the multiple-word string:

e.g.:
Not: *the indicator switch*
But: *indicator switch*

1.2 General spelling and punctuation guidelines

General spelling and punctuation conventions must also be considered when formulating the canonical form:

- Spell the canonical form based on the declared standard for the given language, e.g., *Duden* for German, *Oxford English Dictionary* for UK English, *Webster's* for US English.

- Formulate the canonical form in lower case unless an upper-case spelling is required in the given language to identify or distinguish it:

  - Proper nouns (e.g., personal names, company names, product designations) represent a common set of canonical forms that may contain upper-case spellings. The orthographic conventions of the given language dictate which part of the proper noun should actually be formulated in upper case. Usually, one or more elements has an initial upper case:

    e.g.,
    European Commission   = Eng. proper noun
    Commission européenne = Fr. proper noun

  - Acronyms also often require upper-case handling; typically, the entire acronym appears in upper case:

e.g.,
>>> *BOM*
>>> *EU*
>>> *MIOS*

- The orthographic conventions of some languages require that words that function as particular parts of speech appear with an initial upper case character, e.g., nouns in German.

- Formulate multiple-word strings with a single space between words.

- Hyphenate words based on the standards in place for the given language either generally or specifically within your organisation.

- Use punctuation marks only if they are required as integral parts of the canonical form itself.

- Do not add notes, comments or other additional information within parentheses as part of the canonical form:

e.g.:
>>> Not: *follow-up action (for contract)*
>>> But: *follow-up action*

- Do not use commas to indicate grammatical structure:

e.g.:
>>> Not: *indicator, necessary*
>>> But: *necessary indicator*

2 Language-specific guidelines

2.1 English

- Formulate **nouns** in the singular, unless the meaning of the entry string requires a plural marking.

- Enter **verbs** in infinitive form except in special cases where the entry itself applies to a specific verb form. In the case of the infinitive, there is no notation for the infinitive marker *to*.

- Formulate **adjectives** and **adverbs** in the positive form unless the meaning of the entry string requires the comparative or superlative inflection.

- **Multiple-word strings:** Adhere to standard English conventions for ordering the different words within the multiple-word string.

  - **Noun multiple-word strings:** Enter in standard, unmarked sentence order; if the noun is modified by a descriptive adjective, the adjective should appear before the noun:

    e.g.:
    > *service agreement*
    > *binding agreement*
    > *reference of publication*
    > *contract for an indefinite period*
    > *outstanding maintenance request*

  - **Verb multiple-word strings:** The general rule is to enter verb multiple-word strings with the verb in first position; all other words in the multiple-word string follow the verb in their standard, unmarked sentence order:

    e.g.:
    > *adhere to contract*
    > *re-format document*
    > *calculate cost*

  - **Adjective multiple-word strings:** Enter in standard, unmarked sentence order; if the adjective is modified by an adverb, enter the adverb first:

    e.g.:
    > *necessary to reorganization*
    > *partially profitable*

- **Prepositional multiple-word strings:** Enter the prepositional phrase with the noun object in base form unless a non-base form is essential to distinguish the meaning of the phrase; do not include filler adjectives, determiners, or other modifiers unless they are meaning-distinguishing:

    e.g.:
    > *on receipt*
    > *in escrow*

- **When the multiple-word string is actually a clause:** If the multiple-word string contains both a subject and predicate, formulate it with the subject followed by the predicate. Filler or function words that are not essential to the meaning can be left out:

    e.g.:
    > *invoice paid upon receipt*
    > *network traffic monitored by central office*
    > *treasurer to clear all accounts*

- **Abbreviations:** A general rule of thumb is to formulate abbreviations without a period, although conventions in place in the user's organisation can override this.

2.2   German

- Follow standard conventions for upper- and lower-case spellings.

- Follow the standard German rule of an initial upper-case spelling with nouns borrowed from other languages.

- Use the German umlauts (Ä/ä, Ö/ö und Ü/ü) as required.

- **Compound nouns:** Compounds in German are generally written with all of the compound elements combined into a single word. In some cases, a hyphen is required as a separator between elements. Use the following guidelines to determine when to hyphenate a compound German noun:

  - Formulate a compound composed of two or three words without hyphens:

    e.g.:
    Bestellvorschlagskennzeichen

  - If a compound noun consists of multiple compounds, use a hyphen to separate the compounds if the entire compound noun is otherwise no longer clearly parsable. This does not apply if the two compounds are connected with the compounding –s:

    e.g.:
    Zahlungsträger-Druckprogramm
    Personalveränderungsmitteilungsblatt

  - Use a hyphen to separate compound elements when the compound noun consists of an acronym or abbreviation, and a noun (the ordering of elements is not relevant):

    e.g.:
    ABAP-Report
    Perioden-LIFO
    Bestell-Nr.

  - Formulate the compound noun without a hyphen if it is composed of a shortened form and a noun:

    e.g.:
    Dispokredit

- If the compound noun is composed of the word *nicht* and a noun that is neither an acronym nor an abbreviation, formulate it without a hyphen:

  e.g.:
  > *Nichtlagermaterial*
  > *Nicht-SAP-Anwendung*

- If the head of the compound noun is a borrowed word, and the preceding descriptive element is German, formulate the compound without a hyphen:

  e.g.:
  > *Standardanwendungssoftware*

- Use hyphens to separate the compound elements of a noun that consists of two or more borrowed elements preceding the head of the compound. If the word is a proper noun or trademark, formulate it as it appears in the language of origin:

  e.g.:
  > *Batch-Input-Mappe*
  > *Screen Painter*

- **Verbs:**

  - Enter verbs in lower case, in infinitive form, and without the infinitive marker *zu*:

    e.g.:
    > *bemerken*   not: *zu bemerken*

  - Avoid using any overt notation for marking separable prefixes:

    e.g.:
    > *aufrufen*   not: *auf rufen, rufen auf, auf:rufen*, etc.

- **Adjectives:** Formulate adjectives, including participial adjectives, in uninflected form:

  e.g.:
  > *rechtsbündig*
  > *empfangend*

▪ **Multiple-word strings:**

- Remember not to use commas or other punctuation to indicate grammatical structure.

- Formulate the head word in its canonical form and enter all other words with their appropriate inflection.

- Enter the multiple-word string in standard, unmarked sentence order.

  - **Noun multiple-word strings:** Enter in standard, unmarked sentence order; if the noun is modified by a descriptive adjective, the adjective should precede the noun and should be formulated with a **strong adjective ending**, unless other elements in the multiple-word string preclude this:

    e.g.:
    geschäftsführender Direktor
    Auswahl an
    Wirkung auf

  - **Verb multiple-word strings:** The general rule is to enter verb multiple-word strings with the verb in first position; all other words in the multiple-word string follow the verb in their standard, unmarked sentence order:

    e.g.:
    kaufen auf Abruf

  - **Adjective multiple-word strings:** Enter in standard, unmarked sentence order; if the adjective is modified by an adverb, enter the adverb first:

    e.g.:
    unbedingt nötig
    leicht erträglich

  - **Prepositional multiple-word strings:** Enter the prepositional phrase with the noun object in base form unless a non-base form is essential to distinguish the meaning of the phrase; do not include filler adjectives, determiners, or other modifiers unless they are meaning-distinguishing:

    e.g.:
    nach Meinung
    um Uhr

- **When the multiple-word string is actually a clause**

  - If the multiple-word string contains both a subject and predicate, formulate it with the subject followed by the predicate. Filler words that are not essential to the meaning can be left out:

    e.g..:
    Anhänger schwenkt aus.
    Quittung wird ausgedruckt.

  - If the inflected verb in the clause is *sein*, it can be omitted:

    e.g..:
    Maschine nicht betriebsbereit
    Aufzug in Reparatur

2.3 French

- Follow standard French spelling conventions.

- Adhere to standard conventions for upper and lower case; in general, initial characters are lower-case unless standard French orthography requires an upper case:

  e.g.:
  > *conversion en euro*
  > *devise de transaction*
  > *Institut monétaire européen*

- Use the accepted accents and diacritic marks if possible:

  e.g.:
  > *nº article*
  > *intérêt*

- **Compound nouns:** Generally, compounds in French are formulated without hyphens to connect the compound elements. Adhere to this rule as well for special cases such as the following:

  - Formulate compounds composed of an acronym or abbreviation without a hyphen:

    e.g.:
    > *procédure FIFO*
    > *analyse ABC*

  - If the compound consists of a borrowed element with the head in French, formulate the compound without a hyphen:

    e.g.:
    > *dossier batch input*

  - Do not use hyphens for compound nouns that consist of elements borrowed from English:

    e.g.:
    > *batch input*
    > *capital lease*
    > *joint venture*

- **Verbs:** Enter verbs in infinitive form with no infinitive marker:

    e.g.:
    ordonnancer
    comptabiliser
    générer

- **Adjectives:** Formulate adjectives, including participial adjectives, in uninflected form:

    e.g.:
    compatible
    fiscal
    engagé

- **Multiple-word strings:**

  - Remember not to use commas or other punctuation to indicate grammatical structure.

  - Formulate the head word in its canonical form and enter all other words with their appropriate inflection.

  - Enter the multiple-word string in standard, unmarked sentence order.

    - **Noun multiple-word strings:** Enter in standard, unmarked sentence order; if the noun is modified by a descriptive adjective, the adjective should generally follow the noun:

        e.g.:
        fiche du point de vente
        entité organisationnelle
        amortissement exceptionnel fiscal

    - **Verb multiple-word strings:** The general rule is to enter verb multiple-word strings with the verb in first position; all other words in the multiple-word string follow the verb in their standard, unmarked sentence order:

        e.g.:
        indiquer niveau de danger

    - **Adjective multiple-word strings:** Enter in standard, unmarked sentence order; if the adjective is modified by an adverb, enter the adverb first:

e.g.:
inhérent à

- **Prepositional multiple-word strings:** Enter the prepositional phrase with the noun object in base form unless a non-base form is essential to distinguish the meaning of the phrase; do not include filler adjectives, determiners, or other modifiers unless they are meaning-distinguishing:

  e.g.:
  à côté
  en cas de besoin

▪ **When the multiple-word string is actually a clause**

- If the multiple-word string contains both a subject and predicate, formulate it with the subject followed by the predicate. Filler and function words that are not essential to the meaning can be left out:

  e.g.:
  article déjà mis à la disposition du fournisseur

2.4 Spanish

- Follow standard Spanish spelling conventions.

- Use the accepted accents and the Spanish *ñ* for both upper- and lower-case spellings

- **Compound nouns:** Generally, compounds in Spanish are formulated without hyphens to connect the compound elements. Adhere to this rule as well for special cases such as the following:

    - Do not use hyphens to write compounds containing acronyms or abbreviations:

        e.g.:
        > *análisis ABC*
        > *procedimiento FIFO*
        > *Nº pedido*
        > *proveedor SC*

    - Do not use hyphens for compound nouns that contain elements borrowed from English:

        e.g.:
        > *juego de datos batch input*
        > *procedimiento multitasking*
        > *sistema stand alone*
        > *buffer directory*
        > *batch input*

- **Verbs:** Enter verbs in infinitive form with no infinitive marker:

    e.g.:
    > *programar*
    > *contabilizar*

- **Adjectives:** Formulate adjectives, including participial adjectives, in uninflected form:

e.g.:
> *compatible*
> *amortizable*

- **Multiple-word strings:**

  - Remember not to use commas or other punctuation to indicate grammatical structure.

  - Formulate the head word in its canonical form and enter all other words with their appropriate inflection.

  - Enter the multiple-word string in standard, unmarked sentence order.

    - **Verb multiple-word strings:** The general rule is to enter verb multiple-word strings with the verb in first position; all other words in the multiple-word string follow the verb in their standard, unmarked sentence order:

      e.g.:
      > *comprar mediante órdenes de entrega*
      > *compensar diferencias de inventario*

- **When the multiple-word string is actually a clause**

  - If the multiple-word string contains both a subject and predicate, formulate it with the subject followed by the predicate. Filler words that are not essential to the meaning can be left out.

2.5 Portuguese

- Follow standard Portuguese spelling conventions; maintain consistency with the standard that you are using, either Continental or Brazilian.

- Use the accepted accents and diacritic marks for both upper- and lower-case spellings, where possible.

- Formulate words that are borrowed from English with the English spelling unless they have been assimilated into the language and accordingly have a valid native spelling.

- Formulate canonical forms in lower case unless the entry string contains an acronym, abbreviation, or proper noun.

- **Compound nouns:** Generally, compounds in Portuguese are formulated without hyphens to connect the compound elements. Adhere to this rule as well for special cases such as the following:

  - Formulate compounds composed of the head and two or more modifiers with forward slashes to separate the modifiers:

    e.g.:
    comparação planejado/real
    escala preço/quantidade
    código de débito/crédito

  - Do not use hyphens or slashes between compound elements if the compound contains an acronym or abbreviation:

    e.g.:
    avaliação FIFO
    análise ABC
    elemento PEP

  - Write compounds containing shortened forms without hyphens and with prepositional connectors, where indicated:

    e.g.:
    registro info para compras
    tipo de info

- Do not use hyphens for compound nouns that contain elements borrowed from English:

  e.g.:
  > pasta batch input
  > programa include

- **Verbs:**

  - Formulate Portuguese verbs in infinitive form:

    e.g.:
    > *exibir*
    > *eliminar*

  - Do not enter non-infinitive verb forms, including participles, unless the form's meaning or usage has to be distinguished from that of the infinitive

- **Adjectives:** Enter adjectives in the masculine singular form:

  e.g.:
  > *compatível*
  > *fiscal*
  > *obrigatório*

- **Multiple-word strings:**

  - Remember not to use commas or other punctuation to indicate grammatical structure.

  - Formulate the head word in its canonical form and enter all other words with their appropriate inflection.

  - Enter the multiple-word string in standard, unmarked sentence order.

    - **Verb multiple-word strings:** The general rule is to enter verb multiple-word strings with the verb in first position; all other words in the multiple-word string follow the verb in their standard, unmarked sentence order.

- **When the multiple-word string is actually a clause**

  - If the multiple-word string contains both a subject and predicate, formulate it with the subject followed by the predicate. Filler words that are not essential to the meaning can be left out.

2.6 Danish

- Follow standard Danish spelling conventions.

- Adhere to standard conventions for upper and lower case; in general, nouns are entered with initial lower case and in the singular:

  > e.g.:
  > > *materiale*
  > > *produktionsniveau*
  > > *produktion*

- Use the accepted special characters if possible:

  > e.g.:
  > > *kendetegnsværdi*
  > > *købe*

- **Compound Nouns:** Compounds in Danish are generally written with the compound elements combined into a single word. In some cases, a hyphen or other marker is required as a separator between elements. Use the following guidelines to determine when to use a separating character in a compound Danish noun:

  - Formulate compounds composed of the head and two or more alternate modifiers with forward slashes to separate the modifiers and a hyphen to link the modifiers to the head:

    > e.g.:
    > > *minimum/maksimum-analyse*
    > > *pris/ydelses-forhold*
    > > *plan/faktisk-sammenligning*

  - Use a hyphen to separate compound elements when the compound noun consists of an acronym or abbreviation, and a noun (the ordering of elements is not relevant):

    > e.g.:
    > > *FIFO-princip*
    > > *ABC-analyse*
    > > *PHM-ordre*

- Do not use hyphens for compound nouns that consist wholly of elements borrowed from English:

  e.g.:
  > *buffer directory*
  > *batch input*
  > *capital lease*
  > *joint venture*

- Use hyphens to separate the compound elements of a noun in which one or more of the elements is borrowed:

  e.g.:
  > *breakpoint-tekst*
  > *koncern-template*

(If the borrowed elements have been assimilated into Danish, however, the compound is written without hyphens, e.g., *kommunikationsserver, requestparameter*.)

- **Verbs:**

  - Formulate Danish verbs in infinitive form:

    e.g.:
    > *tidsplanlægge*
    > *bogføre*

  - Do not enter non-infinitive verb forms, including participles, unless the form's meaning or usage has to be distinguished from that of the infinitive.

- **Adjectives:** Enter adjectives in the positive singular form:

  e.g.:
  > *kompatibel*
  > *skattemæssig*
  > *momspligtig*

- **Multiple-word strings:**

  - Remember not to use commas or other punctuation to indicate grammatical structure.

- Formulate the head word in its canonical form and enter all other words with their appropriate inflection.

- Enter the multiple-word string in standard, unmarked sentence order.

- **Verb multiple-word strings:** The general rule is to enter verb multiple-word strings with the verb in first position; all other words in the multiple-word string follow the verb in their standard, unmarked sentence order:

  e.g.:
  købe på afbetaling

▪ **When the multiple-word string is actually a clause**

- If the multiple-word string contains both a subject and predicate, formulate it with the subject followed by the predicate. Filler and function words that are not essential to the meaning can be left out:

  e.g.:
  floden tørrer ud.
  kvittering printet

- If the inflected verb in the clause is *være*, it can be omitted:

  e.g.:
  record afvist
  elevator under reparation