Person Names Guide

Please read through this guide while you are working on /Miscellaneous / Person Name Formats.
There is a new section describing the changes from v43.

Version 44 Changes

New Fields

nativeSpaceReplacement

If your language normally uses spaces between words, skip to Defaults for formality and length!

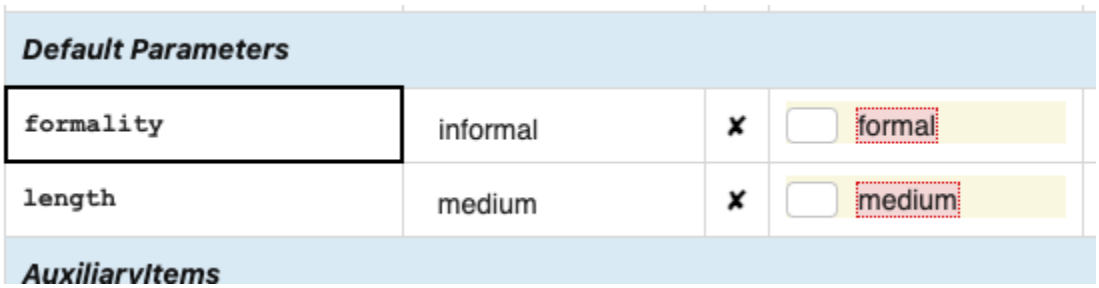
This and *foreignSpaceReplacement* are only for languages that don't use spaces between words.

AuxiliaryItems			
`foreignSpaceReplacement`		✓	· ☆
`nativeSpaceReplacement`		✓	❮EMPTY❯ ☆

Most languages use spaces between name parts, such as given and surname. For them both of the SpaceReplacement items above will be spaces. If your language doesn't use spaces between parts of names, like {surname} and {given}, you'll have ❮EMPTY❯ as the nativeSpaceReplacement. When you compose your patterns ***you will still put spaces*** between name parts (with one exception, #4 below).

1. Remember that if foreign names are not in your script, they will be handled by a different locale, so you don't need to worry about this — it does not affect your locale.
2. If foreign names *(in your script!)* customarily have a particular separator between them, put that separator in the foreignSpaceReplacement item, as above with •. This can also be a space, if that is used.
3. If foreign names *(in your script!)* also ***never*** have spaces in them, put ❮EMPTY❯ in the foreignSpaceReplacement item.
4. If there are some fields, like {title} that always attach to a previous or following field — *whether or not the name is native or foreign* — then, *and only then,* omit the space between it and what it attaches to.
    a. For example, you would have "{given} {surname}{title}", with a space between the given and surname, but none between the surname and title.

Defaults for formality and length

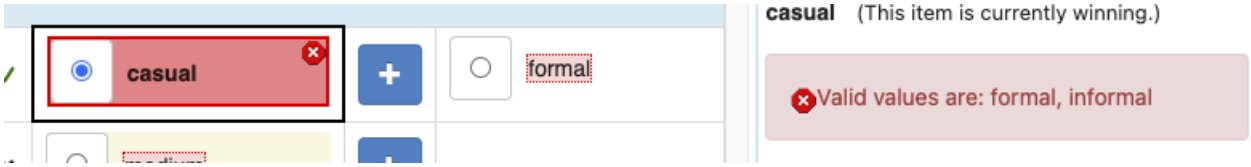


Different locales will often have different defaults for the formality and length, so you can set those with these fields. What does that mean? Programmers will often not know what level of formality to use for different languages, so they will set the name formatting to the default level. That is often the one that users will see.

For the ***formality*** default, look at the formal vs informal examples in the Person Names Report. Ask yourself, which would be the most appropriate form to use on computers for most users and scenarios? (This might be informal for English, but formal for your language.) Set the field to the value appropriate for your locale.

The ***length*** default works the same way: look at the report and pick the most appropriate form for customary usage in most scenarios for most users in your locale.

If you pick an invalid value, the error message will show all the valid options (or you can look below in this guide). You must not translate the values.



New modifiers

Certain locales should make use of three new modifiers:

-retain

> This is needed in languages that preserve punctuation when forming initials. For example, normally {given}=Anne-Marie is converted into initials with {given-initialCaps} as "A. M.". However, if your language preserves the -, you will want to express this as {given-initialCaps-retain}. In that case, the result is "A.-M.". (The periods are added by the pattern-initialSequence.)

-genitive, -vocative

> **If your language does not have grammatical case (noun/adjective inflections), skip to Revisiting Votes!**

The person name formatting does not itself support grammatical inflection. The name fields will be supplied by the data source, often a database, or contacts app, etc. For many languages, both the **referring** and **addressing** forms will be in the same, neutral grammatical case (typically nominative). However, if the data source is rich enough, it may be able to supply inflected name forms.

In that case, typically the **referring** forms will be in the *neutral* case, and the **addressing** forms will be in the *vocative* case. Some modifiers have been added to facilitate this, so that you can have a pattern like: {given-vocative} {surname-vocative}.

However, some **parts of the formatted name** may be in different grammatical cases, so the cases may not be consistent across the *whole* name:

English Pattern	**Examples**	**Latvian Pattern**	**Examples**
{given} {surname}	John Smith	{given} {surname}	Kārlis Ozoliņš
{title} {surname}	Mr Smith	{surname} {title}	**Ozoliņa** kungs

Notice that the surname in Latvian needs to change to the *genitive* case with that pattern:

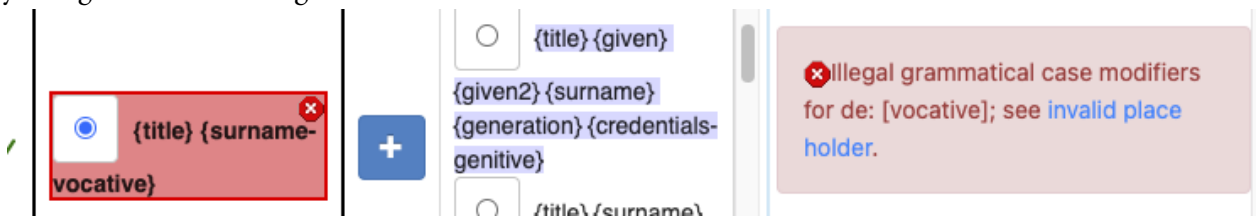
Ozoliņš ➡ **Ozoliņa**

That can be accomplished by changing the pattern to be {surname-genitive} {title}. In this case the {surname} should only be genitive if followed by the {title}. So if the name record has no title it will appear incorrectly. In such a case, you'll need an alternative pattern with no title.

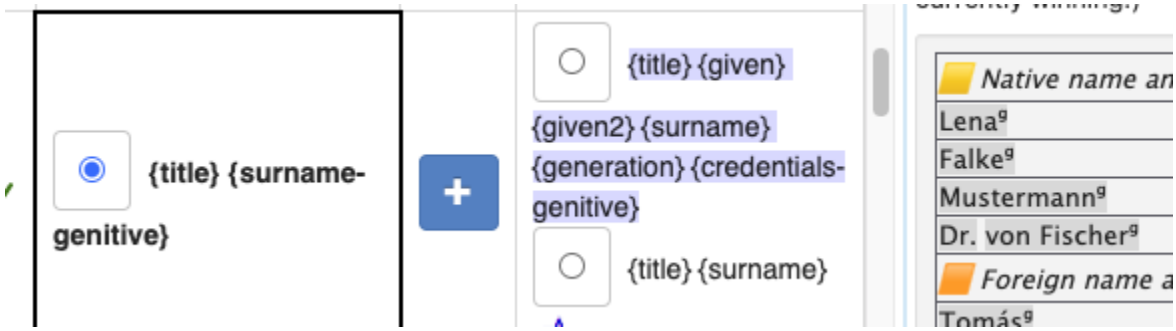
> ***Please let your coordinator know if you need*** *an alternative pattern for a particular pattern,* ***or*** ***submit a form*****.**

Because we don't want to ask you to supply sample names in different grammatical cases, in the examples the grammatical forms will be indicated by superscript letters, such as Ozoliņš[g]. Those superscripts will not show up for end-users: they will either see the correct grammatical form (where supported) or the neutral form (where not supported).

If you use a grammatical case from {*vocative, genitive*} that isn't present for your language, you'll get an error message:



In the above case, the language doesn't have a *vocative*. It does have a *genitive*, so changing that to -genitive will work:



We have only introduced the forms that have been requested, but we can add additional ones that are needed for referring or addressing.

> ***Please let your coordinator know*** *if you need one or more additional grammatical case modifiers for your language,* ***or*** ***submit a form*****.**

Remember that this is not full inflection support: the **whole** names are still restricted in this release to being either **referring** or **addressing**.

Revisiting Votes

Some locales had patterns missing spaces, insufficient variation between requested lengths, and unsuitable sample names. Where it was fairly clear that this was happening, the items have been reverted to Provisional, and will show up in your Dashboard. Here are some of those kinds of problems, and what you should do to address them.

1. Missing spaces **(Skip to #2 if your language uses spaces between words!)**
    - In general, spaces should be used between all elements of a pattern even if your language does not normally use spaces; the separators actually used in formatting will be determined by the values of the nativeSpaceReplacement and foreignSpaceReplacement fields. The only exception is for certain fields that should always be attached to another, as described under nativeSpaceReplacement and foreignSpaceReplacement in this Guide.

- However, some locales that typically do not use spaces have omitted spaces that should be present in name patterns; an example from CLDR v43 summary charts is Japanese (see last column), see also in the CLDR v43 Verification Charts for Japanese.

2. Insufficient variation among patterns for different lengths and usages.
    - An example from the CLDR v43 Verification Charts is Georgian, in which you can see that for both NativeSamples and ForeignSamples, the Main and Sorting charts show no variation for different lengths, usages, or formalities; compare the chart for English. You can also see from the summary chart that there is little variation.
    - This is usually caused by voting for the patterns inherited from root; in Survey Tool the winning value will be marked in pink or blue, and if you click on that item, text in the right-side Info Panel will say something like "This item is inherited from the root locale" or "This item is inherited from another field in the Root locale".
    - To fix this, you will need to enter name patterns that actually reflect the way names are formatted in your locale for different lengths, usages, and formalities. See the section on Name Patterns in this Guide.
3. Foreign sample names that don't appear to be foreign
    - An example from the CLDR v43 Verification Charts is Icelandic, in which you can see in the "ForeignSamples: Main" chart that all examples of the foreignFull name actually show an Icelandic name, not a foreign name. You can also see this in the summary chart.
    - The foreign sample names are supposed to help vetters understand examples of how foreign names (when displayed in the same script as your locale) should be formatted in your locale (using the submitted name patterns), which may be different from how native names are formatted. The foreign names should be names clearly associated with a different language (and usually a different region). See the section on Foreign Names in this Guide.
    - There are also cases of this with non-Latin scripts: see Hebrew full foreign names.
4. Native sample names that don't appear to be native
    - An example from the CLDR v43 Verification Charts is Estonian, in which you can see in the "NativeSamples: Main" chart that all examples of the nativeFull name actually show the nativeFull example from English, not a native name. You can also see this in the summary chart.
    - The native sample names are supposed to help vetters understand examples of how native names should be formatted (using the submitted name patterns). See the section on Native Names in this Guide.

    - There are also cases of this with non-Latin scripts. See Amharic (pick the source & target to see it transliterated to English) for example.

You should also review your language's Person Names Report, and compare it to another language's (such as French or English) to see if there are other problems that we didn't catch.

Review Report

As a reminder, be sure to review the Person Name Formatting Report when you are done!

Person Name Guide, Version 43

Why do we need formatting for people's names?

CLDR has added formatting for Person names, such as John Smith or 宮崎駿. These use patterns to show how a name record (for example, from a database) should be formatted for your language. Such a name record will have fields for the parts of people's names, such as a **given** field with a value of "Maria", and a **surname** field value of "Schmidt".

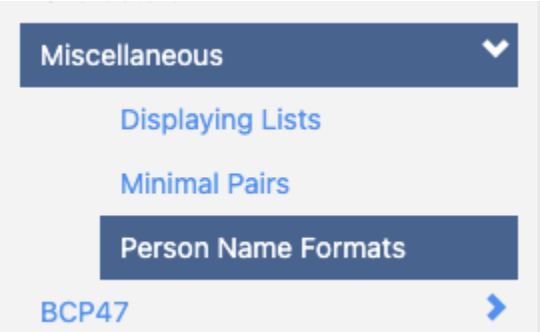
There is a wide variety in the way that people's names work in different languages.

- People may have a different number of names, depending on their culture--they might have only one name ("Zendaya"), two ("Albert Einstein"), or three or more.
- People may have multiple words in a particular name field, eg "Mary Beth" as a given name, or "van Berg" as a surname.
- Some languages, such as Spanish, have two surnames (where each can be composed of multiple words).
- The ordering of name fields can be different across languages, as well as the spacing (or lack thereof) and punctuation.
- Name formatting needs to be adapted to different circumstances, such as a need to be presented shorter or longer; formal or informal context; or when talking about someone, or talking to someone, or as a monogram (JFK).

The CLDR functionality targets formatting names for typical usage on computers (e.g. contact names, etc.), rather than special circumstances, such addressing royalty. However, the structure may be enhanced in the future when it becomes clear that additional features are needed for some languages.

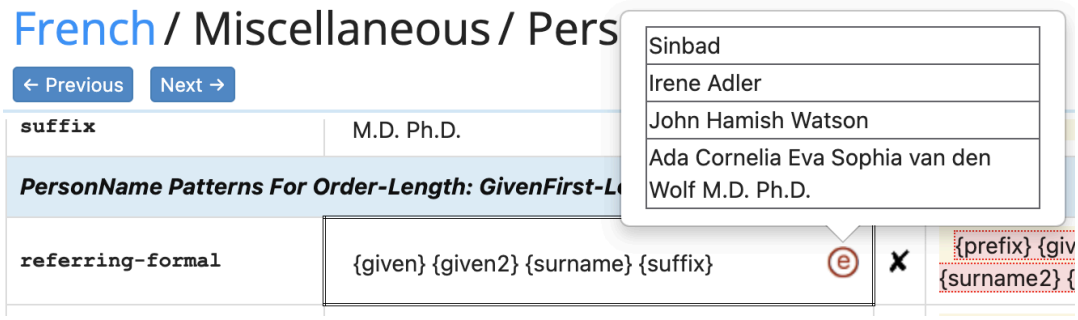
Starting out

To start, you'll go to the new formats once you have logged in and picked your locale.



Note that Person Name data can only be entered at the Modern level. If your locale's Coverage level is set to Moderate or Basic, they will not show in your dashboard.

Have the **Info Panel** open as you work. Remember that hovering over ⓔ in the English column will show you English examples of a pattern in use, while hovering over a Winning or Other pattern will show examples in your language, as shown below.



Here are the types of items you'll see as you go down the page. *Make sure you complete each section before you start the next.* (Of course, you can go back and make adjustments later.)

As you read this, it is important to remember that the Person Name patterns should be designed to work for names in your language, but also for *foreign* names that are displayed to people speaking your language. There are more details in Foreign Names below.

NameOrder For Locales[1]

In the first section is a setting for the default order of name parts, depending on the locale.

Note that any changes to the current settings for NameOrder For Locales must be approved by the TC, because it can have potentially disruptive consequences. Please file a forum request for the new value you want.

[1] https://unicode.org/reports/tr35/#Unicode_language_identifier

You set NameOrder for Locales to indicate

a) Whether your language normally has a surname-first or a given name-first.
b) How your language handles names from *other* languages

If your language is given-name first, then add your language's code to the **givenFirst** field (separated by spaces). Otherwise add it to the **surnameFirst** field.

> **Note:** If you don't know your language's code, you can look it up in the Language column of Locale Coverage.

Next, you need to indicate how your language handles *other* languages, *especially those using a different name order than your language*. For example, it is common in some languages to display Chinese, Japanese, and Korean names in their native order (**surnameFirst**), but typically names from almost all other languages are displayed as **givenFirst**, regardless of that language's native name order. So such a language would have:

- "zh ja ko" as the list of **surnameFirst** languages.
- All other languages in the list of **givenFirst** languages

There is a special code for "all other languages", which is "und". So you don't have to list all other languages individually!

- So if your language is **givenFirst** (say French = fr), you would normally have "und fr" in the **givenFirst** field, and some subset (perhaps empty) of **surnameFirst** languages in the **surnameFirst** field.
- If your language is **surnameFirst**, then you would put "und" in whichever field is the normal behavior in your language for displaying *most other* languages (English, Russian, Greek, Spanish, etc). That might be either **givenFirst** or **surnameFirst**. In the other field, put the explicit languages whose names behave that way.

Style guides, Wikipedia articles, and other popular media in your language mentioning people from other countries are great resources to determine common usage for languages you are unsure about. For example, to see how French names are handled in Korean you could look up "Catherine Deneuve" 카트린 드뇌브 - 위키백과, 우리 모두의 백과사전, and find that it is **givenFirst**.

For ***NameOrder*** locale lists, you should use ❮EMPTY❯ if you want to indicate that there are no locales with that feature in your language. For example, in Turkmen in version 43, the given name is always first, no matter what the name's locale is. In that case, the surnameFirst field will be ❮EMPTY❯.

Some people might have been over-zealous in the locales they added to the ***surnameFirst*** list. We know that the following languages use surname first *for their own language*:

Code	English Name
hu	Hungarian
ja	Japanese
km	Khmer

ko	Korean
mn	Mongolian
si	Sinhala
ta	Tamil
te	Telugu
vi	Vietnamese
yue	Cantonese
zh	Chinese

However, that does not necessarily mean that *your* language should use ***surnameFirst*** for all of them. Your language might use ***surnameFirst*** for all of these, some of these, *or* none of these. You need to look at what people in *your* language would expect to see.

> For example, if the Hungarian composer Béla Bartók would be written in the order "Bartók Béla" in *your* language, and similarly for other Hungarians, you would put 'hu' into ***surnameFirst*** — otherwise you wouldn't.

Wikipedia can be a useful resource for looking up sample names, and switching to your language. See NameOrder For Locales for more information.

Default Parameters

TBD: The v44 text needs to be integrated below.

Please see ***Defaults for formality and length*** section in the Version 44 Changes above.

AuxiliaryItems

Next you will find two or three fields. The ***foreignSpaceReplacement*** and ***nativeSpaceReplacement*** are only used for languages that don't normally have spaces between words. If that's your language, see Foreign Names. Otherwise ignore that.

The next one, ***pattern-initial***, is a pattern used to display an initial in your language, while the last, ***pattern-initialSequence*****,** is a pattern used to "glue" together multiple initials for multiword fields. For example, let's see what happens with the given name "Mary Beth" in English.

Action	**Result**
1. Split into words	"Mary" and "Beth"
2. Each converted to initials	"M" and "B"
3. The **pattern-initial** is applied to each	"M." and "B."

4. Finally recombined with **pattern-initialSequence**	"M. B."

What you will see in the Survey tool examples will use the above process. Note that a name record can contain an explicit pattern for a particular name that could be different and will override what results in the above process. Moreover, in some languages (such as Greek) the default computation for initials can be different. So don't worry if what your language does for initials is different from what you see.

Sample Name Fields

You will next find a series of names. The purpose of these is to see how your name patterns will work with a variety of names. Each name has an ID, such as "GivenSurnameOnly".

Here are the meanings of these name fields:

Name Field Code	Example	Description
title	"Dr.", "-san"	A title or honorific, such as **Mr** Johnson, **Herr Prof.** Schmidt, Takahashi-**san** *Typically an abbreviated form.*
given	"Richard"	Typically a given name (or names); may come before the family name depending on the language.
given-informal	"Rich"	Optional informal name or nickname
given2	"Edward"	A middle name (see below), sometimes consisting of (or containing) a patronymic or other form. For example, "Fyodor Mikhailovich Dostoevsky" has the patronymic "Mikhailovich" from his father's name: "Mikhail".
surname	"Smith"	Typically a family name, but sometimes a patronymic, matronymic, place name, etc. See other derivation for example.
surname2	"Iglesias"	Some languages use a second surname
generation	"Jr", "II"	Used in English and a few other languages for generational titles, such as John Smith **Jr** or John Smith **II.** Note that patronymics (or matronymics) such as Sergeyovich or those using affixes such as Arabic "bin"/"bint"/"ibn" are usually handled differently, typically as part of the given2 field. *Typically an abbreviated form.*

credentials	“Ph.D”, “MP”	A list of credentials or accreditations, such as John Smith, **DDS**, or Winston Churchill, **MP**. *Typically an abbreviated form.*

Notes:

1. Each of these fields can have multiple words. For example, a **title** can be “Frau Dr.” or a given can be “Mary Sue”.
2. The usage of all fields — especially **title, generation, credentials** and **surname2** — may vary considerably depending on the language.

What goes where?

Because name fields can have multiple words, it is not always obvious where to draw the line. Does “Mary Sue Ellen Smith” have “Mary Sue” as **given**, and “Ellen” as **given2** (aka middle name)? Or some other combination?

The key is that the **given2** is optional when formatting. So if “Mary Sue Ellen Smith” typically goes by “Mary” or “Mary Smith”, then “Mary” is the **given** and “Sue Ellen” is the **given2**. Conversely, if “Mary Sue Ellen Smith” would *never* go by “Mary”, but rather always be “Mary Sue”, then her **given** name would be the multiword “Mary Sue” and her given2 be “Ellen”.

Similarly, the choice between **surname** + **surname2** and a multi-word **surname** is determined by whether they are ever split in normal formatting.

Title, generation, and credentials

These are defined as follows:

title	A title or honorific, such as **Mr** Johnson, **Herr Prof.** Schmidt, Takahashi-**san.** This is not the same as credentials. For example, "Dr. Smith, MD" has both a title and a credential. While both “MD” and “Dr.” derive from the same source, they have different expressions and positions in formats, at least in some languages. It would be wrong to format these as "MD Smith Dr." (in English and other languages). *Typically an abbreviated form.*
generation	Used in English and a few other languages for generational titles, such as John Smith **Jr** or John Smith **II.** Note that patronymics (or matronymics) such as Sergeyovich or those using affixes such as Arabic “bin”/“bint”/“ibn” are usually handled differently, typically as part of the given2 field. *Typically an abbreviated form.*

credentials	A list of credentials or accreditations, such as John Smith, **DDS**, or Winston Churchill, **MP**. *Typically an abbreviated form.*

The positioning and surrounding characters for these fields may vary substantially from English.

Name Samples

The sample names are placed in two groups: native sample names and foreign sample names. Note that whenever we are talking about foreign names, it is foreign names *in your writing system*, such as "アルベルト・アインシュタイン" in Japanese, for the German name "Albert Einstein", or "Олексій" in Ukrainian for "Oleksiy".

> **Note:** foreign names *in a foreign script*, such as an appearance of "George Clooney" in the middle of Chinese text, is handled differently — don't worry about that case.

Your first priority should be to review *all* sample names, and adjust according to the descriptions below.

Native Names (in your script)

The native names should be names that are common in your locale. For the native names, you don't need to supply fields like **surname2** if they are never or very rarely used in your locale. Note that the native sample names should ***not*** be translated or transliterated, but instead be replaced by names that are common in your locale.

Don't just translate / transliterate the English names! Instead use names that are common in *your* language. However, avoid controversial (and infamous) people's names (e.g. Al Capone).

The nativeFull names have to be fully populated with fields, **however, if you never use {credentials} in your language, you have to remove that field from your patterns—you can mark that it isn't used in nativeFull by using ∅∅∅, but don't try to remove it from nativeFull or foreignFull.**

Foreign Names (in your script)

On the other hand, the foreign names should be foreign names, expressed in your writing system.

You can use the English names as the foreign names in your language. If your language is normally not in the Latin script, then you need to make sure that you translate / transliterate them into your language/script.

You may choose to use different foreign names from what the English has, but you need to choose values for all the fields. That is so that you can see how all the patterns would behave with names that have features that your language might not, such as **given2**, **generation**, **surname2**, or **surname-core** and **surname-prefix**.

The foreign name can come from a language that uses your writing system, but the **fullForeign** pattern needs to have ***all*** of the fields: so **surname2** as in Spanish, **surname-core** and **surname-prefix** as in Dutch, **generation** marker as in English, and so on.

> The foreignFull names have to be fully populated with fields, because databases will have those fields and you have to design the patterns to accommodate them.

Examples

The Examples show two sets of names produced by a pattern, plus a foreign name & script for comparison.

As in other examples in the Survey Tool, this shows what the pattern would look like with samples substituted into the pattern:

Header	**Description**
Native name and script:	Uses native sample names (in your script)
Foreign name and native script:	Uses foreign sample names (in your script)
Foreign name and script:	For comparison, shows a couple of foreign names in their own script, with corresponding patterns.

You can see that the spaces from the pattern show up in the result.

Name Patterns

Next are the Name Patterns. Each of these patterns will show examples in the Info Panel using the sample names you just added.

The syntax of the patterns is fairly straightforward. It's the desired output text, with the names of the fields as placeholders in *braces* (aka "curly brackets"), such as:

Sample Pattern	{given} {given2-initial} {surname}, {credentials}

Here we're specifying that the result should start with the "given" field, followed by the "given2" field transformed into an initial, followed by the "surname" field, with all of them separated by spaces. At the end is the "credentials" field, which is separated from the "surname" field by a comma and a space. This would produce output like this:

Sample Name	**Sample Result**
given=John, given2=Bob, surname=Smith, credentials=MD	John B. Smith, MD
given=John, given2=Bob, surname=Smith	John B. Smith
given=John, surname=Smith	John Smith
given=John	John

Whitespace and punctuation next to a field is automatically removed if that field is empty. See the Missing Fields section below if you want more information on how this works.

The field names in braces may also have *modifiers*, which are appended with a hyphen. "given2-initial" in the **Sample Pattern** above is a field name with a modifier. For a complete list of the supported modifiers, see the Modifiers section below.

The name records will follow the following rules.

- The given names are split into **given** and **given2**, to allow the **given2** (aka middle names) to be dropped for less formal or shorter formatting.
- If there is only a single name, like "Madonna", it will be in the **given** field. All of the other fields may be empty.
- If a name has text in the "surname2" field, it also has to have text in the "surname" field.

- A name can have a “surname-prefix” and “surname-core” fields instead of a plain “surname” field. This is only needed when the prefix needs to be omitted (e.g., turning “van der Poel” into “Poel”) or the prefix and core need to be separated or inverted (e.g., “Poel, van der”).

When developing name patterns for a particular language, keep in mind that they’ll also be used to format *foreign* names for your language. Remember that “foreign names” means “foreign names in your writing system”. (Foreign names in another writing system will be handled with other locales’ patterns).

This means you may have to account in the name patterns for some fields in the name record being populated that normally would be left empty in your language. For example, the “surname2” field from Spanish names[2] is normally omitted in English formatting. However, the priority is to get the patterns correct for native names, and then accommodate foreign name formatting as much as possible.

The patterns in your language may differ substantially from what is in English. Don’t assume that you just want to copy the English. In general:

- formal names should include both **given** and **surname**
- long formal names should include **given2** as well.
- most languages will not include the **surname2** (aka maternal family names or sometimes patronymics) unless it is also used in native names.

Your patterns could differ by what goes between placeholders (e.g. spaces or punctuation) but also may include a field that is not in English or discard a field that is in English. You can look at newspapers, magazines, and Wikipedia for examples of foreign names used in your language and script. For example, you can look at famous Spaniards in Wikipedia in your language to see how they appear, such as Pedro Almodóvar.

Code Values

The code value will have a series of ‘parameters’ separated by hyphens. Here is an explanation of those parameters. For English examples, hover over the ⓔ, as in the image above at the top of this page.

Order

Parameter	Description
`surnameFirst`	The surname precedes the given name. Other fields would be placed appropriately.
`givenFirst`	The given name precedes the surname. Other fields would be placed appropriately.

[2] It is customary in Spanish for people to have two surnames, the surname(s) of each parent. Traditionally, the first surname is the father's, and the second is the mother's. More information on Spanish naming customs.

`sorting/index`	This is the format that such names would have in a sorted list, such as the index in a book, or names in a telephone directory. For example, see Index. In English and many other languages with "givenFirst", this uses a "surnameFirst" form, but with punctuation to indicate a reversed order. Eg: Smith, John B., Jr. NOTE: your language may need you to use **surname-core** and **surname-prefix** for some of these patterns. For example, for English it is used for the formal cases to get formats like "Humboldt, Alexander von".

The patterns for "surnameFirst" and "givenFirst" may be the same, if the pattern doesn't contain a surname, such as in an informal-addressing name. Even if your language only uses one order (eg, just "givenFirst"), both are needed for use in foreign language names displayed in your language.

A sorting/index pattern may be the same as either of the others, although it will be more similar to the "surnameFirst", with some exceptions (like Icelandic where sorting is done by the given name).

Length

Parameter	Description
`long`	Usually includes all parts needed for a legal name or identification.
`medium`	Includes the fields in most typical usage on computers.
`short`	Names using a shorter format; typically where screen real-estate is limited, or where a longer form of the name has already been used.

Usage

The usage indicates if the formatted name is being used to address someone, refer to someone, or sort their name with other names.

Again, the format may be the same in some cases for some cultures where usage is different. For example: usage=sorting/index may have the same format as usage=addressing, if the language uses "surname" first.

Parameter	Description
`addressing`	Used when speaking "to" a person, such as at the start of an email addressed to that person, or in the user interface to address an application's user.
`referring`	Used when speaking "about" a person, such as a list of customers or musicians in a music application.
`monogram`	A highly abbreviated version of the name, *for use in computer interfaces*, ideally no wider than about two "em's" in width. Think of an 'em' as the width of an emoji, so no wider than ⬛⬜. For example, in English, the "long" monogram might be MED, the

	medium be MD, and the short be M. It's okay for the different lengths to be duplicates of each other. Monograms will not use the **initial** or **initialSequence** patterns.

The name format samples don't themselves handle grammatical inflections in this release. You can indicate a desired grammatical case: see New modifiers.

Monograms

Remember that monograms are meant for computer usage (not what you would have embroidered on clothing). Unless required in your language, you should avoid any literal characters in your patterns (such as '.'). However, whenever such characters are needed, then they should be included. If so, then you may see examples like M.-D in the examples and reports. That can happen if a field is missing. Take the following pattern used for Arabic, Farsi, and Pashto in version 43.

{given-monogram-allCaps}.{given2-monogram-allCaps}.{surname-monogram-allCaps}

When there is no 'given2' value (eg, no middle name), then {given2-monogram-allCaps} disappears. The specification and software will remove one of the dots since they overlap. So nothing needs to be done. That is, when given2 doesn't exist, the above pattern is equivalent to:

{given-monogram-allCaps}.{surname-monogram-allCaps}

Giving a result like: "M.D"

However, there may be circumstances where that doesn't work: check the examples for the pattern to make sure. For example, if the pattern is:

{given-monogram-allCaps}.{given2-monogram-allCaps}:{surname-monogram-allCaps}

Then when the middle field is absent, both . and : are retained.

{given-monogram-allCaps}.:{surname-monogram-allCaps}

Giving a result like: "M.:D"

In such a case, you need to request that an additional pattern be added that doesn't have the given2 field, so that you can provide the right characters. Contact your coordinator about this.

Formality

Parameter	Description
`formal`	A more formal name for the individual. The composition depends upon the language. For example, a language might include the **title** and **credentials** and a full middle name (**given2**) in the long form.

informal	A less formal name for the individual. The composition depends upon the language. For example, a language might exclude the **title**, **credentials** and **given2** (middle) name. Depending on the length (and language), it may also exclude the surname.

Some languages will have more degrees of formality than just two, such as Japanese. In the first release, implementations for certain languages may handle these multiple degrees via the name data that they supply for formatting; for example, changing the **title** field to accord with the level of formality. If your language has more than two levels, pick the two that are most likely to be needed on computers, for instance: **informal** for friendly social greetings among friends, and **formal** for referring to or addressing business owners, professors, or formal communications.

Over time, the available styles may be augmented for languages that have more than two levels of formality such as Japanese, Korean, or Telugu .

Modifiers

Field names in name patterns may have one or more modifiers attached, as mentioned earlier. They can be used in combination with one another also, such as **{surname-initial-allCaps}**. A name record could contain any of these explicitly, but the software can produce default versions of them.

The modifiers we support right now are:

Modifier	Description
allCaps	Requests the element be completely capitalized. Note: this should not be used unless the field is *always* all-caps in this pattern in your language. There is a growing convention in certain languages to use all-caps for a **surname** placed before a **given** name, but that can also be done at runtime, following a "house style". *Examples: Joe ⇒ JOE, di Giorno ⇒ DI GIORNO*
initialCap	Requests that the first letter in the element be capitalized without affecting the rest. *Examples: mcAdam ⇒ McAdam* (Note, initialCap has no trailing 's')
initial	Requests an initial for a name field. The results vary by length and whether the name field has multiple words. The default uses **initial** patterns, and applies to multiple words. *Examples: Mary ⇒ M., Mary Sue ⇒ M. S.* Can be combined with **-retain** to keep existing punctuation that may be between words in a name part. See below.
monogram	Similar to 'initial', but just consists of a single letter, without punctuation. The default uses the first letter of the first word, without **initial** patterns. *Examples: Mary ⇒ M, Mary Sue ⇒ M*

`informal`	Requests an informal version of the name field, if available. The default is the normal form of the name field. *Example: {given} might be "Thomas", and {given-informal} might be "Tom".*
`retain`	This is used in languages that preserve punctuation when forming initials. For example, normally {given}=*Anne-Marie* is converted into initials with {given-initial} as "*A. M.*". However, if your language preserves the '-' between *Anne* and *Marie*, you will want to express this as {given-**initial-retain**}. In that case, the result is "*A.-M.*". (The periods in this example are added by the **pattern-initialSequence**.)
`genitive`	Suppose your language supports the genitive case, such as when the surname in Latvian needs to change to the *genitive* case with that pattern: Ozoliņš ➡ **Ozoliņa** {surname-genitive} {title} ➡ **Ozoliņa** kungs That can be accomplished by changing the pattern to be {surname-**genitive**} {title}. In this case the {surname} should only be genitive if followed by the {title}. So if the name record has no title it will appear incorrectly. In such a case, you'll need an alternative pattern with no title. You can ask your coordinator to add an alternative pattern like that, or submit a form.
`vocative`	If your language has different forms for names between addressing (vocative case) and referring (nominative case), and the data source is rich enough, it may be able to supply inflected name forms. In that case, typically the **referring** forms will be in the neutral (nominative) case, and the **addressing** forms will be in the *vocative* case. Some modifiers have been added to facilitate this, so that you can have a pattern like: *{given-**vocative**} {surname-**vocative**}*.
*The following are only used where parts of a **surname** to be separated sometimes in formatting.*	
`prefix`	Return the "prefix" ("tussenvoegsel") of a name field (eg, "van der Poel" ⇒ "van der"). Defaults to "" if the name has none, so "Smith" ⇒ "".
`core`	Return the "core" name, removing any tussenvoegsel (eg, "van der Poel" ⇒ "Poel"). Defaults to the same as the name field if the name field doesn't have a tussenvoegsel, "Smith" ⇒ "Smith".

Note: The survey tool shows the default behavior of the modifiers when it shows the samples. But the results can be customized by an implementation. For example, initials may be more than one letter in certain languages, such as Greek or Hindi. Moreover, an implementation can optionally request all *surnameFirst* formats to have all surname fields in allcaps. For example, that would cause given=Hayao

surname=Miyazaki to be formatted as “MIYAZAKI Hayao”, even if there is no -allCaps modifier in the patterns.

Foreign Name Spacing

There are two main challenges in dealing with foreign names formatted for your language that you need to think about. One was the ordering, which is dealt with under the Name Order For Locales topic above. The other is spacing.

Some writing systems require spaces (or some other non-letters) to separate words. For example, Hayao Miyazaki is written in English with a space between the two name fields, while in Japanese there is no space:

宮崎駿

1. If your locale requires spaces between words, the normal patterns for your locale are used. We see examples on Wikipedia. Note the space within the Japanese name on pages from English and Korean.

- “Hayao Miyazaki (宮崎 駿, Miyazaki Hayao...” or
- “미야자키 하야오(일본어: 宮﨑 駿 Miyazaki Hayao...”.

2. If your locale **doesn’t** require spaces between words, there are two cases, based on whether the foreign name is written in your script, or the foreign name is in its native script. In both cases, patterns from the **locale of the name** are used. For example, your locale might be Japanese, and the locale of the name might be German (Switzerland), such as Albert Einstein.

A. **The foreign name is written in your script.** In that case, the **foreignSpaceReplacement** is substituted for each space in the patterns from the **locale of the name**. Here are examples:
    - アルベルト・アインシュタイン
    - 阿尔伯特·爱因斯坦
B. **The foreign name is written in a different script.** In that case, the patterns from the **locale of the name** are used as is.
    - Albert Einstein

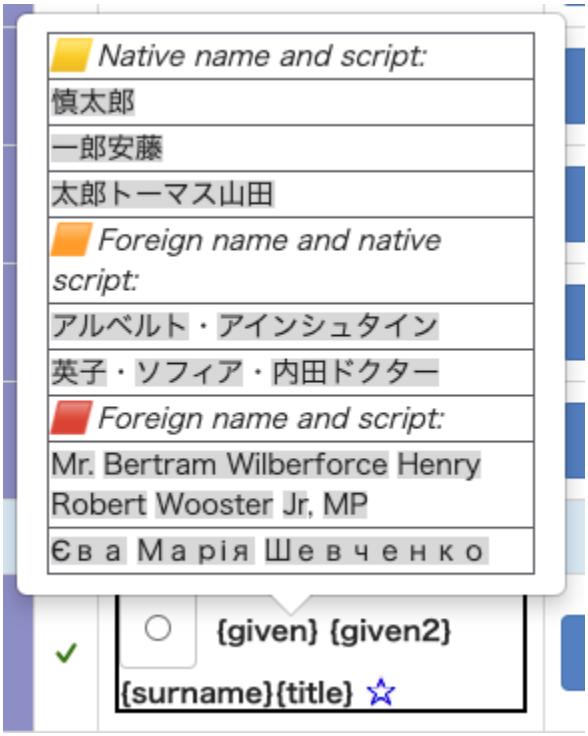
In both cases, the ordering may be changed according to the **Name Order for Locales** settings that you provide. If the name record does not supply a locale for a name, then a default locale will be derived based on other information (such as the script of the characters in the name fields).

foreignSpaceReplacement

Continuing from above ... In languages that don’t need spaces between words (such as Japanese), the ***foreignSpaceReplacement*** replaces any spaces in patterns when used for foreign names. The change is that any spaces in patterns will be replaced with the ***nativeSpaceReplacement*** in native names.

That will mean that you should insert spaces so that they work for foreign names. They will be removed for native names, and replaced by the ***foreignSpaceReplacement*** in foreign names (in your script).

You can see this below: for Japanese the spaces in the pattern are deleted for the native names, and replaced by "・" for foreign names in the native script.



NOTE: This is a screen shot before vetters have a chance to fix their sample names. So the Japanese name is odd: it is a non-Japanese name in Japanese script. We'll update the screen shot soon.

You have two distinct choices.

1. If foreign names in your script ***never*** need spaces or other characters (such as ・) between fields:
    - then you can dispense with spaces in all of your patterns.
    - Choose this option if it applies to your language.
2. *However*, if foreign names in your script more typically need spaces (or some other non-space punctuation such as ・) between fields:
    - then you should put spaces in *all* of the patterns, and set the ***foreignSpaceReplacement*** to be a single space or specific punctuation used in your language such as ・in Japanese.
    - That way the pattern will behave well both for the native names and the foreign names.

nativeSpaceReplacement

If your language *always* uses spaces between words, skip this section!

This and *foreignSpaceReplacement* are only for languages that don't use spaces between words.

AuxiliaryItems			
`foreignSpaceReplacement`		✓	· ☆
`nativeSpaceReplacement`		✓	❮EMPTY❯ ☆

Most languages use spaces between name parts, such as given and surname. For them both of the *SpaceReplacement* items above will be spaces. If your language doesn't use spaces between parts of name names, like {surname} and {given}, you'll have ❮EMPTY❯ as the *nativeSpaceReplacement*. When you compose your patterns ***you will still put spaces*** between name parts (with one exception, #4 below).

1. Remember that if foreign names are not in your script, they will be handled by a different locale. (The exact details for that step are in the spec, but you don't need to worry about them.)
2. If foreign names *(in your script!)* customarily have a particular separator between them, you'll put that separator in the foreignSpaceReplacement item, as above with •. This can also be a space, if that is used.
3. If foreign names *(in your script!)* also ***never*** have spaces in them, you'll put ❮EMPTY❯ in the foreignSpaceReplacement item.
4. If there are some fields, like {title} that always attach to a previous or following field — *whether or not the name is native or foreign* — then, and only then, should youyou can omit thea space between it and what it attaches to.
    a. For example, you would have "{given} {surname}{title}", with a space between the given and surname, but none between the surname and title.

Missing Fields

This section provides more information about how the characters between field placeholders are handled when a field is missing.

You normally don't have to look at this section unless your name patterns don't work right for sample names that are missing some fields.

For example, a name might only have {given}="Mary" and {surname}="Smith", or might only have a single field: {given}="Zendaya". When a **particular** field (such as {given2} or {title}) is in the pattern but not available in the name, then not only is that field omitted, but some of the literal characters (aka literals) around it may be as well. Here's how that works.

If a field has nothing between it and the neighboring fields, it's just omitted. If it only has whitespace between it and the neighboring fields, the field is omitted and any duplicate whitespace is collapsed

down to a single space. But when there are other literal characters, such as punctuation, things get more complicated.

- If you have one or more fields omitted at the very beginning or end of the pattern, everything–fields, whitespace, and any other literal text–before the first populated field or after the last populated field, is omitted.
- If two fields in a row in the middle of the name are omitted, so is all the literal text between them.
- And if you have one or more fields omitted in the middle of the name, so is any adjacent literal text, up to the nearest whitespace character or populated field on either side.

In the example below, the format has parentheses around each of the fields. (This is an artificial example, just for illustration). In formatting, the parentheses around {given2} would be removed, but not the parentheses around {given} or {surname}, because of the whitespace.

Pattern	**Name Record**	**Result**
({given}) ({given2}) (surname})	given=John, surname=Smith	(John) (Smith)

If this doesn't give you the effect you need, you can supply additional patterns to show what should happen if certain fields are left empty. For example:

Some languages like Spanish have both a surname *and* a surname2 field. Let's suppose that the formal sorting/index format has the following pattern (with a comma after the surnames). If there happens to be no surname2 field (for example, for a foreign name), then the comma would be removed when the surname2 field is empty, giving you the wrong result.

Pattern	given	surname	surname2	**Result**
{surname} {surname2}, {given}	Antonio	Lopez	Pascal	Lopez Pascal, Antonio ✅
	Antonio	Lopez		Lopez Antonio ❌

To fix this, you can ask for an additional pattern to be added, and specify it to be: {surname}, {given}. The second pattern will be used when there is no surname2 field, thus providing the correct result:

Patterns	given	surname	surname2	**Result**
{surname} {surname2}, {given} {surname}, {given}	Antonio	Lopez	Pascal	Lopez Pascal, Antonio ✅
	Antonio	Lopez		Lopez, Antonio ✅

For more information, see Q. The pattern doesn't work for a particular example. What do I do?

Special terms

❮EMPTY❯

If a value is to be empty (such as a ***NameOrder*** list of locales, or a ***foreignSpaceReplacement***), use the value ❮EMPTY❯ instead (including the two heavy angle-brackets). This is a special value that signals what you want, and is only accepted in those rows. In the Info Panel, you can copy this out of "... insert ❮EMPTY❯". Note that the heavy angle brackets are not the ASCII < or >; they are different Unicode characters.

> For example, suppose that a language doesn't normally use surname-first for any other language, writing Zedong Mao instead of Mao Zedung. In that case, ❮EMPTY❯ should be entered in the NameOrder surnameFirst field. Similarly, suppose that a language deletes all spaces in foreign names (that is, replaces them by "nothing"). In that case ❮EMPTY❯ should be entered for the foreignSpaceReplacement.

∅∅∅

When entering sample names for nativeFull, if a particular field is not used in your language, it can be replaced with the ∅∅∅ string (three U+2205 EMPTY SET characters) to indicate that that field is never used in names in your language.

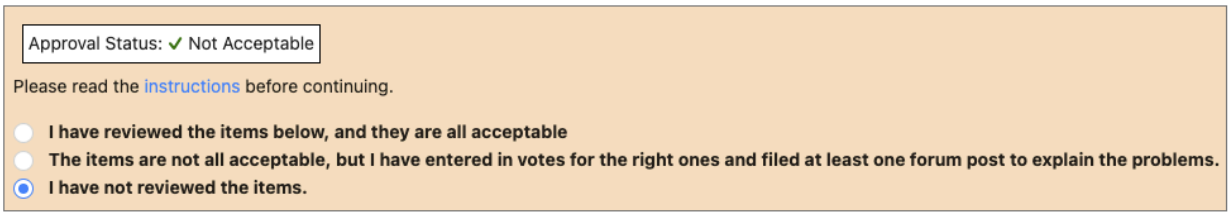
Review Report

After you've adjusted all the values as necessary, go to the **Reports** > **Person Names** to look over all of the examples there. *This is an important step, because the Examples don't show all the possible combinations of sample names with parameters!*

Make sure that the names look right for the gray settings in each of the 4 tables:

1. **NativeSamples:**
    a. **Main**
    b. **Sorting**
    c. **Monogram**
2. **ForeignSamples:**
    a. **Main**
    b. **Sorting**
    c. **Monogram**

Below is an example screenshot.



NativeSamples: Main

Order	Length	Usage	Formality	nativeG	nativeGS	nativeGGS	nativeFull	view
givenFirst	long	referring	formal	慎太郎	一郎安藤	太郎トーマス山田	恵子グレース佐藤ジュニアさん	view
		addressing						view
		referring	informal			太郎山田	けいこ佐藤さん	view
		addressing						view
	medium	referring	formal			太郎トーマス山田	恵子グレース佐藤ジュニアさん	view
		addressing				太郎山田	恵子佐藤さん	view
		referring	informal				けいこ佐藤さん	view
		addressing						view
	short	referring	formal				恵子佐藤さん	view
		addressing			安藤	山田	佐藤さん	view
		referring	informal	慎太郎	一郎安藤	太郎山田	けいこ佐藤さん	view
		addressing			一郎	太郎	けいこさん	view
surnameFirst	long	referring	formal		安藤一郎	山田トーマス太郎	佐藤グレース恵子さん	view
		addressing				山田太郎	佐藤恵子さん	view
		referring	informal				佐藤けいこさん	view
		addressing						view
	medium	referring	formal				佐藤恵子さん	view
		addressing						view
		referring	informal				佐藤けいこさん	view
		addressing						view
	short	referring	formal				佐藤恵子さん	view
		addressing			安藤	山田	佐藤さん	view
		referring	informal	慎太郎	安藤一郎	山田太郎	佐藤けいこさん	view
		addressing			一郎	太郎	けいこさん	view

- If the results are not all acceptable, click the "**The items are not all acceptable, …**" button.
    - Then use the 'view' link to go to the pattern in question to make fixes.
    - If you need other people to change their votes, *be sure to leave forum requests*. If there is a common problem (such as lack of spaces between fields) you only need one forum post on the first item exhibiting the problem.
    - Be aware that sometimes the problem will be in the **AuxiliaryItems** (foreignSpaceReplacement, pattern-initial, or pattern-initialSequence), so the 'view' link may not take you to the ultimate source of the problem.

    - Return to the reports page to see if your changes fixed the problem.
- Once all the results are acceptable, click "**I have reviewed the items below, and they are all acceptable**".
- In some locales you may find something that doesn't allow you and the other vetters to fix the problem. For example, in some circumstances Latvian requires the surname to change to the possessive grammatical case. In that case (and only that case), please tell your coordinator and file a ticket explaining exactly what the problem is.

FAQs

If you have questions, please forward them to your PM or coordinator. We can add them to the list here as well as respond to you.

Q. The pattern doesn't work for a particular example. What should I do?

A. So what do you do if you can't make a particular pattern work for your language — especially when particular name fields are missing (see Missing Fields)? Or if this guide is unclear?
    a. First, you can leave a comment in this document. (If you don't have a google account, you can register your email at https://accounts.google.com/SignUp.
        i. You can also select Tools > Notification Settings to see when comments are addressed.
    b. Second, if you are contributing on behalf of a TC organization, also contact your PM.

Q. In my language, no names are of the form "van den Wolf". Do I have to use the {surname-prefix} and {surname-core} fields in my patterns?

A. While you may not have names like that in your language, if someone called "van den Wolf" comes to your country to speak at a conference and needs to be included in an index, your formatter needs to handle the name (if transliterated) *as people speaking your language would expect*.

   If your language doesn't normally separate the parts, then you can just use {surname} in the pattern, and when it is formatted, the two parts will always be combined together.

   Some languages may sometimes suppress the surname-prefix, or separate the surname-prefix from the surname-core, such as in the sorting formats for an index. So for users of those languages, someone might see a list like:

   - ...
   - Humboldt, Alexander von
   - Mignone, Pablo
   - Ramée, Louise de la
   - ...

For those languages, you will want to use {surname-prefix} and {surname-core}.

See, for example, Guidelines for Preparing an Index of Names or Alexander von Humboldt — Sources

Q. In the field "SampleName Fields For Item: nativeFull" > "given2"; do I have to enter a value?

A. For native sample names, you should supply **given2** names if it is possible in your language; the same goes for the **surname2**, etc. The only hard requirement for the native sample names is to have a **given** name, and where called for, a **surname**. Note that surname doesn't mean just a family name; in many languages that don't have a tradition of family names, it can be a patronymic, or locale, or some other identifying feature.

Q. In the field "SampleName Fields For Item: nativeFull" > "given" and "given2": should we absolutely enter compound names for both?

A. If your language can have compound names (that is, multiple words like "Mary Sue") in a field, then you should enter them if possible.

Q. In the field "SampleName Fields For Item: foreignFull" > "surname2": should we absolutely enter a surname that is a compound name?

A. If your language separates words with spaces, then yes. Compound names can help to test out certain features of the formats that are otherwise not visible, such as how initials behave.
    a. Even if your language doesn't separate words with spaces, you should use spaces if they would be used in foreign names in your script.

Q. In the field "PersonName Patterns For Order-Length: GivenFirst-Long" > "addressing-formal": should we add "Mister" or "Madam" in front of the surname (when the title is not used)?

A. **Do not** add a literal title like that in the pattern yourself! Instead, use "{title}", which will be replaced by a title like Mister or Madam if that is available in the name record.

Q. Should I translate the credentials "MD PhD" (in the nativeFull sample) into my language?

A. No, you should pick **credentials** that would be used in names in your language for the native name sample, and **credentials** that you'd see in a foreign language for the foreign name sample.

The **title** and **credentials** should be created to reflect **titles** and **credentials** in common use in your language; they should *not* just be translations of the English. In the foreignFull sample,

they can be the English terms (translated or transliterated into your writing system) or other foreign terms of that sort.

Q. Should I include periods in credentials and titles *in the Sample Names?*

A. You need to supply them where necessary (used in your language). That is, there are three possible sources of periods in a formatted name:
   a. Those in the name record (eg, in a **credentials** like "M.D.")
   b. Those inserted by the mechanism that produces initials, ***IF*** you have one in the **pattern-initial** field (eg, "John" is transformed into "J.").
   c. Those explicitly in the pattern. (These are *almost always* a mistake.)

Q. I see "hu ja ko vi yue zh" in the surnameFirst field. What are those?

A. Those are language codes for languages like Hungarian or Japanese. These should be in the ***surnameFirst*** field for ***your language*** if and only if your language would put surnames first when formatting names from *those* languages.

   For example, if your language puts surnames first for names in your language, and it would also put surnames first when formatting Khmer names, then you would put "km" in the list. If it doesn't put surnames first when formatting Hungarian names, then you would remove "hu" from the list.

   If you want to know what the language codes in this field have as names in your language, click on them and look at the examples in the **info panel**.

Q. Should I just vote for those languages (hu ja ko vi yue zh)?

A. **Be careful. That is just a default list.**

   That means that names in each of those languages would be formatted *in your language* with the surname first. So for a Hungarian like Mónika Szeles (where Szeles is the surname), you would see her name formatted as "Szeles Mónika" because "hu" is in that list. Similarly for Béla Bartók, and other Hungarian names.

   If it should be Mónika Szeles for your language, then take "hu" out of the list, before voting for the list. The same goes for the other languages in that list.

Q. In French, credentials are not used. Should linguists in languages without credentials like French still include the {credentials} field in the patterns?

A. It really depends on what a French reader would expect. For example, would they expect to see "Dr. James McDonald, MD" in the middle of a news article in a French newspaper, etc.? Or just "Dr. James McDonald". So if they would expect to not normally see the **credentials** in French — even for foreign names — then you should omit {**credentials**} in the French patterns.

The formal vs informal may play into this as well: you may want the {**credentials**} in long formal patterns, but not in the others. Again, it should be based on the typical French reader's expectations.

Q. The guide says that each field in the name can have multiple words. Does this mean we have to supply multiple words for each field in our sample names for French?

A. The **...Full** samples are intended to 'stress' the patterns as much as possible, so that people can see whether the patterns can handle unusual cases.

The **nativeFull** sample name should contain a **title** that you would expect to see in French. (It would come out of a name record.) Now, if there are no multi-word **titles** used in French (eg, no equivalent of "Herr Prof. Dr. Schmidt"), then you can't have it be both French and multi-word.

Your **foreignFull** sample name should contain multiple words if possible, since that might come in from a name record: so something like Herr Prof. Dr. Schmidt.

In English, for example, you wouldn't normally expect to see "Herr Prof. Dr. Schmidt" in a nativeSample, but could in a foreign sample.

Q. As per the above guide, 'surnameFirst' field should be left empty for my locale (e.g. hi, fr) in the 'NameOrder For Locales' section. But that doesn't work; the Survey tool doesn't accept an empty value. What is the expected input here?

A. If a value is to be empty (such as a **NameOrder** list of locales, or a **foreignSpaceReplacement**), use the value ❮EMPTY❯ instead (including the two heavy angle-brackets).

This is a special value that signals what you want, and is only accepted in those rows. In the Info Panel, you can copy this out of the text underlined here: "... insert ❮EMPTY❯ ...". Note that the heavy angle brackets are not the ASCII < or >; they are different Unicode characters. They are deliberately chosen to be unusual characters.

Q. What does the error message "cannot have prefix without core" mean?

A. A name record in some languages will split a surname field like "von Neuman" into a **surname-core** and a **surname-prefix**. That is so that either the "von" can be suppressed, or the ordering can be changed (eg in a Sorting/Index pattern).

It would be malformed to only have the "von" in a name record, so that's why you have to add **surname-core** if you have a non-empty **surname-prefix**. To summarize:

surname-prefix	**surname-core**	**surname**	OK?

		von Berg	🆗
von	Berg		🆗
von	Berg	von Berg	🆗
von		von Berg	🤦
	Berg	von Berg	🤦
von			🤦
	Berg		🤦

Q. What should be entered in the Person Name Formats > NameOrder For Locales if our alphabet is non-Latin: The code in Latin characters (i.e., en-US) or in local characters (e.g. Hebrew or Urdu)?

A. The NameOrder fields must be a list of zero or more language codes from the Language column in the Locale Coverage table. Thus a valid value for English and German would be

en de

Note that if you want to specify ***no*** language codes, you need to use the special value ❮EMPTY❯, as described above.

Q. In my language, we wouldn't use an {given-informal} in a computer interface. What should I do?

A. The name formatting can be used in a wide variety of settings, not just business settings. The program will have control (and ideally give it to users) over the level of formality. For example, people might want to use the informal name in a mail merge letter to close friends using their contacts list. So you should honor the 'informal' setting — after all, if informal names are never used in your language, you probably wouldn't get them in a names database.

Remember also that if the {given-informal} is not in the source data, the {given} will be used. So you don't have to worry about {given-informal} being unavailable.

Q. For the sample name "Jean de Bouchart", the rendered monogram currently is "JD" while we would rather expect "JB". How do I fix this?

A. If you use {surname-core} instead of {surname}, you get the surname without the prefix. So, for example, "Bouchart" instead of "de Bouchart". For the monogram you would replace

a. {given-monogram-allCaps}{surname-monogram-allCaps}

by

b. {given-monogram-allCaps}{surname-**core**-monogram-allCaps}

Q. For the sample name "Marie-Agnès Gilot", the rendered monogram currently is "MG" while we would rather expect "MAG"; similarly the short name is "M. S. Gilot" where we'd expect "M.-A. S. Gilot". How do I fix this?

A. This is excellent feedback. For the monograms and initials, the programs using name formatting have the ability to either
   a. Allow users to have their preferred form explicitly, eg put "M.-A." explicitly in their contact list as the given-initial.

      OR

   b. Use their own algorithm for constructing initials and monograms, one that can be customized for given languages.

B. That being said, we want to make things work "out of the box" as well as possible, so we can see whether there is a good way to handle hyphenated names for languages that behave as yours does as well as ones that don't. So don't worry for now that M. A. appears instead of M.-A.