

Design Issues in Multilingual Applications

People have always perceived and organized the world in terms that are familiar to them. Why else would the constellation that was called “The Great Bear” in ancient Greece and Rome now be called “The Big Dipper” in North America, “The Casserole” in France or “The Plough” in England? Those “mental models” were strongly influenced by culture, environment and experience. Similarly, anyone who calls into an IVR (interactive voice response) system and experiences the “hear and feel” of the interface will necessarily form a mental model of the interaction. Drawing on past experiences and cultural norms, the caller’s expectations about how to interact with the system will be set.

Look Right!

While crossing the streets of London, tourists can be grateful for the signs painted on the roads’ crosswalks advising them to “Look Right” for oncoming traffic. If you’re a pedestrian from a country where cars navigate the right side of road, you’re conditioned to expect that when stepping off a curb, the most immediate danger will come from your left. It’s not intuitive to first look right when entering a crosswalk. The local government knows this and has judiciously stenciled arrows and words of warning onto their sidewalks, even though the majority of local pedestrians, being natives, don’t need them. The city has provided the extra instructions that we foreigners need to compensate for something that’s not intuitive to us, so we can navigate their streets without ending up on the hood of a London taxicab.

The lessons from the “Look Right” example are, as with countless aspects of day-to-day life, applicable to voice user interface (VUI) design. For multilingual applications to be effective and successful, they must recognize that callers from different cultures will have different notions of what is intuitive, and this can have a profound impact on the design.

Interactions that are intuitive, by definition, don’t require any additional instructions to the caller:

System: *“Do you want to hear the phone number again?”*

It’s a simple question in search of a simple “yes” or “no” answer. No further explanation is necessary.

Interactions that are not intuitive require additional guidance:

System: *“Okay, I’ll check availability for May 13. If I have the date wrong, say ‘back up.’”*

There’s nothing intuitive about saying “back up” in conversation to correct a date. But some form of instruction is called for because callers may be uncertain as to how to correct a misrecognized date. The above “say ‘back up’” strategy compensates for what may not be intuitive to callers.

One of Us, Or One Of Them?

The “hear” and “feel” of an application will be interpreted within the context of each caller’s culture. When it comes to a system’s persona, the same regional accent that sounds endearingly homespun to one caller may leave another caller thinking he or she is dealing with a small-time operation. It’s important to understand whether your Mexican Spanish callers will be put off in some way by a South American Spanish speaker, or vice versa. Does the other accent have positive or negative connotations for them? Some services are more sensitive than others when it comes to using accents. An accent may very well bring added value to a service with a regional identity. For example, in a travel application, the tourist guide with a local accent could bring the atmosphere of the tourist attraction to life. For more official public services, the designer may want to refrain from the use of accents because of the accessible-to-everyone character of the service.

During a ring tone download project, target group analysis showed that the service was primarily going to be adopted by preteens and teenagers, nine to 15 years old. Tests revealed that it was important to utilize the hip, slangy and informal language of this particular age group. Also, the prompts were recorded with young voices in order to connect with the callers’ mental model and expectations. The “being-one-of-us” tone brought added value to the voice application. A formal prompt style would very much have a counter-effect, being inappropriate for this particular service character and target group. In multilingual applications, the prompt designer needs to be extremely aware of what’s hip, slangy and informal in one language compared with another.

A banking application uses a banking language prompt style in a rather formal way. Even when the caller speaks more colloquially, he or she expects the system to adopt a more formal style. The mental model is set on banking, and to most of the callers, banking is serious business, with images of people wearing suits in formal settings.

In a certain bank, as a test, a young man was put behind one of the counters, dressed informally, with a baseball hat, sweatshirt and jeans. People entering the bank were avoiding him (especially younger clients!) and preferred to conduct their business with a formally dressed person. In this example, the “being-one-of-them” is a precondition in order to connect with the mental model and expectations of the caller. Formal language is a

concept that differs from culture to culture. The prompt designer needs to be highly informed about the professional jargon in order to connect with the target group in the most effective way.

“You” Can Make A Difference

One of the most important considerations for a multilingual application is to pay close attention to the level of formality. Many languages around the globe make a distinction between a more familiar and a more formal you (e.g., French tu/vous or German du/Sie). For convenience, let’s write the familiar and formal forms as “you” and “You,” respectively. The conventions associated with when to use which form is as culturally unique as the languages themselves. Prompting with “You” in an application will create different mental models for callers of different cultures. In a German language application, “You” would be expected in all cases, except with a target audience composed exclusively of young people (for whom “You” would come across as authoritarian and alienating). For the general population, though, “you” might prove distracting to the caller at best — impertinent at worst. Although the Swedish language makes a similar “you”/“You” distinction, prompting with formal “You” would be the unnatural case, for young people and adults alike. The effect would be a detached and old-fashioned feel.

Investigations conducted by the Fraunhofer Institute in Germany show that, instead of a formal prompt style, a friendlier and more personal style is more suitable and effective in making the caller accept recognition failures, and it initiates cooperative caller behavior. System prompts that address a problem concerning speech recognition in an open and sympathetic manner are mostly perceived as being friendly and helpful. (As in: “Speech apps make mistakes too.”)

VUI designers with a native-speaker feel for a given language might still exploit the effect that a shift in formality can provide. Consider this example, first in English.

System: *We have appointments available Thursday, Friday or the following Monday. Which would you like?*

Caller: *I’d like to come in on Saturday.*

System: *What was that?*

Caller: *Saturday.*

System: **Nope, can’t do it.** *Saturday’s completely booked. I can offer you only Thursday, Friday or Monday. If you want to check your calendar, you can say “pause” and I’ll wait until you say “continue again.”*

The lack of any “you”/“You” distinction in English does not prevent the language from making a shift in the formality level. The last prompt, spoken with a smile, serves to break down the dialog’s formality — for only a moment — to smooth over a rough situation and get the dialog back on track.

Now consider a German version of the same dialog, remembering that the German formal “You” is Sie, and that the familiar “you” is du.

System: *Es gibt mögliche Termine am Donnerstag, Freitag oder am nächsten (There are possible appointments on Thursday, Friday or on next*

Montag. *Am welchen Tag möchten Sie vorbeikommen?* (Monday. On which day want You to come?)

Caller: *Ich möchte gerne am Samstag vorbeikommen.* (I would like on Saturday to come.)

System: *Verzeihung, bitte?* (Pardon, please?)

Caller: *Samstag.* (Saturday)

System: *Neh du (haha), um Gottes Will. Das geht leider nicht. Am Samstag sind (No, you, for God’s sake, that works unfortunately not on Saturday are) wir zu. Es gibt nur noch Donnerstag, Freitag oder Montag. Bitte wählen Sie.* (we closed. There is only left Thursday, Friday or Monday. Please (you) make a choice.)

As with the English example, the system temporarily shifts away from the normal, more formal register in an attempt to bond with the caller, ease any frustration the caller may have, get the caller to refocus and bring him or her back on track. It’s the IVR equivalent of a friendly pat on the shoulder.

[Fe]Male

One question that generally would not fit the mental model of speaking with a human agent is whether you are male or female. Countless prior telephone calls will have taught you how well your voice alone suggests to people whether they should call you “Sir” or “Ma’am.” You know from past experience whether people in the local culture have any trouble guessing your gender from your traditionally male or female first name. You set your expectations accordingly.

Speech recognition technology does not yet reliably exploit the pitch of a caller’s voice to determine gender. And with a multicultural calling population, the system can’t depend on first names. For instance, the French pronunciation of “Daniel” could be interpreted as “Danielle.” “Pat,” to English speakers, may be used for both male and female persons and does not clearly provide any information about gender. With Asian names, most Western people can only guess; e.g., “Chao” (Chinese) meaning “great one” may be used both for males and for females. The cost of guessing wrong is great, as the system will come across as unintelligent and alienating. Additionally, instructing the recognition machine to sort out which names are male and which are female requires extensive

grammar design.

Consequently, the VUI designer needs to find out the intelligent way.

English is gender-neutral in certain cases:

English: *"Are you satisfied with the outcome of the survey?"* (gender-neutral)

The same example in Spanish demands more precision:

Spanish: *"¿Está contento con el resultado de la encuesta?"* (male)

"¿Está contenta con el resultado de la encuesta?" (female)

The way out in the example above could be rewording the prompt, avoiding the use of a gender-sensitive reference to the addressee, with a phrase that literally translates "Is the outcome of the survey pleasing to you?"

Spanish: *"Le gusta el resultado de la encuesta?"*

Another strategy is to create a gender-neutral prompt using the equivalent of "one" or "any person." The same can be used to address both singular and plural referents:

Swedish: *"Man kann åka taxi till stan."* (One could take a cab into town.)

But also: *"It's possible to take a cab into town."*

Where there's ambiguity, it's common for such languages to default to one gender. But, if the automated system is using the default gender in cases where a human generally wouldn't, the caller's mental model may be disturbed and the caller's evaluation of the system blemished. Therefore, even for a language like Spanish, alternative wordings that don't expose the limitations of the technology would be preferable.

... And For Icelandic, Press 94

What language will the dialog be in? Australian English, Japanese, touch-tone? Multilingual applications need to establish a mode of communication early on, as all subsequent interaction depends upon it. The "welcome prompt" needs to cover the initial information and, at the same time, to invite the caller to take part in the human-machine interaction, "inviting" the caller over the system threshold into a conversational dialog. Callers' expectations of the service they've called need to be taken into serious consideration. By leveraging the language instinct and creating a conversational dialog, the VUI design evolves in the direction of a natural, human-like interaction. VUI design is not about fooling callers into believing they are talking to a live call center agent, but rather making callers forget that they are talking to a machine.

When a multilingual service is accessed through a single telephone number, it's common to allow callers to express their preference.

System: *"To continue in English, say 'English.' Para servicio en Español, diga 'Español.'"*

The practicality of offering a language preference to callers diminishes as the number of languages grows. Callers could be forced to hear several prompts in languages they don't understand before being offered what they need.

Will the order of languages presented stir up political sensitivities? If there is a falsely accepted misrecognition, the only recourse to callers unable to navigate in a foreign dialog will be to hang up, and business could be lost. Depending on how well the caller base's language might map to location (think European Union), the caller's ANI (automatic number identification) might be used to presume which language should be used for the opening prompt.

Alternatively, the IVR could work with a companion Web site, on which callers could register their phone number (for ANI) and a language preference. The system, in turn, would know with which language to greet that person the next time he or she called. Imagine an example of a voice portal supporting hundreds of services. These single services cannot be "reached" through the use of short cuts; too many short cuts would lead to deteriorated recognition. In this case, the designer may want to give the caller the option to create a personal set of short cuts ("favorites") through a companion Web site. Take, as an example, the German toll-collect system, where truck drivers would need to select a language from a variety of welcome prompts in different languages. In this particular case, simply registering on a companion Web site once would solve the problem, connecting the registered phone number to the pre-selected language.

Some Assembly Required

Making concatenated prompts sound natural in any language requires careful planning during the prompt-writing stage, in addition to meticulous coaching during the prompt-recording stage. Each part must have the appropriate prosody (intonation, emphasis, pitch and rhythm), so the expressiveness of the concatenated whole will not be

compromised. Consider that addresses are expressed differently in different countries, with the house number either preceding or following the street name.

1492 King Street:

English: Fourteen | ninety-two | King Street.

French: King Street | Mille | quatrecent | quatre-vingt douze
(King Street | Thousand | four hundred | ninety-two)

German: King Street | Vierzehn | hundert | zweiundneunzig
(King Street | Fourteen | hundred | ninety-two)

Danish: King Street | Tusen | firehundrede | toochalfems
(King Street | Thousand | four hundred | ninety-two)

Dutch: King Street | Veertien | tweennegentig
(King Street | Fourteen | ninety-two)

Swedish: King Street | Ettusen | fyrahundra | nittiotvå
(King Street | One thousand | four hundred | ninety-two)

The address elements will need to be recorded with final or non-final intonation, based on the format used. Also, the variety of sentence structures that different languages present means that there may not be a one-to-one correspondence between the concatenated parts, even in languages as similar as English and German.

English: *"I found | three | Thai restaurants | in Toronto."*

German: *"Ich habe | drei | thailändische Restaurants | in Toronto | gefunden."*
(*"I have | three | Thai restaurants | in Toronto | found."*)

Speaking The Same Language

Multilingual designs must address both the linguistic differences and the cultural differences of callers. The issues go far beyond simply translating the prompts. "Speaking the same language" means more than just using words that the other person understands; it also means that the mental model, which a caller forms from the interaction, is the intended model and not a mental model that is tarnished by awkward prompting or cultural insensitivity. No one today will dispute the fact that the most effective and successful voice user interface designs are user-centered. For this to remain true for multilingual applications, it's vital that the needs of each linguistic community be addressed at the earliest stages of the design process. Otherwise, even the most brilliant design can get lost in translation. CIS