Just
It's Not ▾ Linguistics

Rebuttal to "Speech Tech University" article, which appeared in *Speech Technology Magazine*, June 2009

By Susan Hura, Ph.D. and Lizanne Kaiser, Ph.D.

Introduction

We were intrigued by an article in the June 2009 edition of *Speech Technology Magazine* called "Speech Tech University" on background and training for speech professionals, especially voice user interface (VUI) designers (http://www.speechtechmag.com/Articles/Editorial/Feature/Speech-Tech-University-54015.aspx). This is a topic that does not receive enough attention, and we were pleased to see that *Speech Technology Magazine* devoted an entire article to this subject and interviewed several of our esteemed colleagues for insights on this topic.   To our dismay, however, there were a number of disparaging, and in our opinion, misguided comments made about the value that linguistic knowledge— and even linguists personally—bring to the VUI profession. Specifically: "'Linguistics has nothing to do with voice user interface design.'" and "'The linguists don't know anything about design.'"[1] The overall picture painted in this article is that linguists have no place working in speech technology or VUI design, a sentiment with which we respectfully disagree. In the interests of full disclosure, both authors of this paper are card-carrying Ph.D. linguists pursuing careers in VUI, so we clearly have a vested interest in defending our training and the value it brings to our profession. We are not neutral parties in this discussion, but we hope the strength of our argument will convince the reader that an alternate viewpoint is necessary.

Linguistics Is Relevant to VUI Design

The first point we'd like to address is the relevance of linguistics to a career in speech and VUI. The *Speech Technology* article quotes our colleague as saying, "Linguistics has nothing to do with voice user interface design." As a stand-alone statement, this remark is both puzzling and shocking. We agree that studying linguistics does not give you formal training in design, but it absolutely informs the 'voice,' and 'user,' and 'interface' parts of the job.

To start, let's define our terms. Linguistics is the scientific study of language, including language in its spoken form, i.e. *speech*. Speech technologies are computer systems for recognizing and producing

---

[1] Direct quotes taken from VUI expert interviewees cited in the "Speech Tech University" article, *Speech Technology Magazine*, June 2009.

spoken language. Until reading this article, we never considered it controversial to suggest that knowledge of spoken language was relevant to people working with technologies that recognize and produce spoken language. The method for recognizing and producing speech via technology is not the same as the way people recognize and produce speech, but the object of the recognition and production is the same in both cases: speech. We find it nonsensical to suggest that a rigorous and thorough understanding of spoken language (as is offered through training in linguistics) is anything less than vital for speech professionals, including VUI designers.

Another puzzling suggestion from the article is that speech scientists in particular don't need to know much about speech. The article quotes another esteemed colleague at length about how speech science today is really all about algorithms, machine learning, and computer science. We agree that syntactic theory, for example, is unlikely to have a great deal of impact on the day-to-day work of a speech scientist. However, any knowledge speech scientists have about the object of their algorithms will only serve to ground their decisions and ensure that the technology fits the domain. There may be a sense in which speech science is just an exercise in data-manipulation, but even in that case, studying linguistics is valuable because it provides a deep understanding of the nature of the data set.

Studying linguistics provides indirect, but invaluable, training in taking a user-centered approach to design, as well as a framework within which to construct linguistically appropriate speech interfaces. This training is not about how to do design, per se, but more about having an appropriate mindset and knowledge base for doing good design work.

One of the primary lessons in linguistics is to observe what *is* happening, rather than what *should* be happening. As linguists, we're not concerned with the way people *should* speak according to some conventional rule, but rather on observing how people *actually* speak and discovering the underlying rules that govern real usage. Linguists describe this as a focus on *descriptive* language, rather than *prescriptive* language. (See case study in side bar.) A common misconception is that linguists are sticklers about proper grammar and pronunciation by virtue of knowing a lot about language. Linguists love hearing new uses of language, especially those that would make your eighth-grade English teacher's head spin! Linguists are taught to focus on people's real use of language and to carefully filter out our own preconceptions and biases about what should

Actual VUI Case Study: Prescriptive vs. Descriptive Language

A VUI designer was reviewing a dialog design for an automated telephone system with a business client. The business manager had added edits to the prompt wording, including changing all of the contractions into separate words (e.g., *you'll* to *you will*, *won't* to *will not*, etc.). The VUI designer explained that written language is different than spoken language, and in order for the voice talent to record the prompts in a natural way, the prompt wording needs to use conversational language.

The business manager commented, "That may be ok at some other companies, but we don't use contractions around here."

The VUI designer questioned, "You *don't*??"

To which the client replied, "No, we don't.", quite unaware that this response had just proven the VUI designer's point.

be happening, much in the same way that designers need to be able to understand what people really do when interacting with an interface without letting our own biases as experts (or those of our clients who know the domain) color our observations. What we're suggesting is that the linguist's focus on descriptive language is easily translated into a focus on, and even empathy for, end-users.

Linguistics Is No Hindrance to Good VUI Design

The article paints a dire assessment of the relevance of linguistics to the task of voice user interface design. The article devotes an entire section to comments by a VUI design colleague disparaging the field of linguistics and the prospects for any linguist in VUI design. According to the article, linguists have an "immediate disadvantage" in VUI design, as opposed to the biology student who went on to become a successful and talented VUI designer (an example presented in the same section). Unbeknownst to us linguists (recall the present authors are both linguists), we apparently entered the speech industry uniquely unprepared for the job ahead. Once again, we're perplexed. Is the article actually suggesting that linguistic training is a hindrance to being a good VUI designer? In fact, one of the present authors (SLH) chose to study linguistics specifically because of an abiding interest in speech technology.

A more charitable reading of this article leads us to believe that the overall point is not that studying linguistics is actively bad for prospective VUI designers, but rather that linguistics alone is insufficient training for VUI design. This is a well accepted fact that the article acknowledges elsewhere and that the speech community recognizes as well. In fact, a paper produced by the Association for Voice Interaction Design's (AVIxD) VUI Workshop in 2007 focused on exactly this topic (http://avixd.org/assets/WorkshopPapers/2007/VUIEducationTrainingMentoring.pdf). (Note that one of the colleagues quoted in the *Speech Technology Magazine* article is a co-author of this AVIxD workshop paper.) The AVIxD paper strongly makes the case that VUI design is a hybrid profession that requires training and skills in multiple domains, including linguistics. Because there is typically no such thing as a major in VUI design, people enter this field from a wide variety of backgrounds; the goal of junior VUI designers is to understand the gaps in their own training and educate themselves in these areas. Linguists are no better or worse off than those with other relevant professional backgrounds entering the field of VUI design. Linguists typically come into the profession without adequate design training, but by the same token, those trained as psychologists, human factors experts, and user interface designers lack grounding in spoken language.

Something really important to remember in this discussion is that speech technology is not a one-way street; there will always be a person on the other side of a speech system. The automatic speech recognizer is processing some person's speech; some person is listening to synthesized speech. If a speech technology solution is going to be effective and usable, those designing and developing the technology have to take these people into consideration. Among the many things we must consider about the users of speech technology is the way they perceive and produce language—insights which a solid grounding in linguistics is uniquely able to provide. There is much, much more to know about users, but linguistics undoubtedly gives you some important pieces of the puzzle.

What Linguists Bring to the Table

The article quotes one of our colleagues listing four criteria essential for success in VUI design: a good designer must be communicative, analytical, have "a business bent", and have the ability to break down a problem and challenge underlying assumptions. We may be biased, as linguists ourselves, but let us suggest that at least three of these are essential traits for successful linguists as well.

Linguists are intensely interested in communication (hence their chosen field of study), and most tend to be good communicators as well. Getting a degree in linguistics typically requires a great deal of writing, giving oral presentations, and conveying information coherently during teaching interactions. Not every linguist is an excellent writer, speaker or conversationalist, but they're no worse than others who might become VUI designers. Linguists tend to be careful in their communications because of their focus on language, which is a clear benefit for the task of crafting prompts.

An even stronger case can be made for linguists being analytical. Much of linguistics revolves around examining sets of data and looking for patterns. To write phonological rules or understand how languages change over time, linguists learn to pay close attention to minute details while searching for higher-order commonalities. Linguists are well-schooled in the scientific method of making a hypothesis based on observations of one set of data and then testing the hypothesis on new data. Many student who end up as linguists (the authors included) have stories of how they analyzed some aspect of language all on their own, just for fun, and only later learned that there was an entire academic discipline devoted to this endeavor. A nice side benefit of all this analysis is that the data being analyzed is spoken language – not bad experience for a VUI designer.

Successful linguists also excel at the last of the four criteria, breaking down problems and challenging underlying assumptions. From our perspective, breaking down problems and challenging assumptions is part and parcel with studying linguistics. Linguistics begins with a fundamental shift in the way you conceive of language; that what you thought you knew about language is flawed or incomplete. The most difficult task facing the Intro-to-Linguistics student is realizing that what they think of as the rules of language (never end a sentence with a preposition, don't say 'ain't') are not the rules that matter in linguistics. Rather, linguistics is concerned with revealing patterns of language that are so ingrained in our knowledge of our native language that they don't even seem like rules (such as in English you can't start a word with the sound /ng/ (but in Vietnamese you can); and in English adjectives come before nouns (whereas in Romance languages like French and Spanish adjectives follow nouns)). Linguistics students are taught from day one to question the surface level and focus on hidden commonalities within and between languages. Students who continue in linguistics tend to be those who like a good puzzle and take it as a challenge to discover the twist that allows for a unified explanation for a seemingly disparate data set.

As for having a 'business bent,' we have no argument with the idea that linguistics does not provide a rigorous foundation in business. (In fact, one could argue that majoring in linguistics shows a distinct lack of business savvy, given the job prospects for an average linguist!) But this admission puts linguists in a no worse position than many others who might choose to become VUI designers. UI designers,

human factors experts, and speech scientists do not necessarily have any better business acumen than linguists do.

Linguistics Gives you *Extra* Insights

Let's dive deeper into how linguistics gives *extra* insights into VUI design that other disciplines don't necessarily cover. Linguistics is uniquely positioned to uncover insights about human spoken language that may not be readily obvious to most speakers of that language who have not received specialized training and education in linguistics. Consider examples from six fields of study within linguistics:

Phonetics

Phonetics is the study of the sounds of human speech. It focuses on both sound *production* and *perception*. When real humans perceive speech, they combine multiple techniques: (a) *acoustics* (the science of sound) + (b) *articulation* (an awareness of how those sounds are produced by the human vocal tract) + (c) *context* (listeners guess the likely word(s) based on the meaning that fits best within the surrounding context). However, of these three, automated speech recognizers can only rely on pure acoustics (analyzed via statistical algorithms). In particular, speech engines key in on similar sounding vowels, since vowels – with their higher bursts of energy– are easier for the speech engine to distinguish than consonants. So, linguistically trained VUI designers can focus in on these acoustic properties and readily ascertain why having a grammar that recognizes both 'repeat' and 'delete' as commands could lead to problematic false matches, given the similarity in their vowel qualities.

Phonology

Phonology is the study of sound structures within human languages. One sub-area of phonology is *prosody* (the study of the rhythm, stress, and intonation of connected speech). Prosody is essentially the "music of speech." Prosody is often used to convey *pragmatic* inferences (more on that below) and/or paralinguistic meaning. For instance, consider the difference in meaning of this sentence, depending on whether or not the word 'was' is stressed:

(1) a. That was the last one.
b. That <u>was</u> the last one.

(1a) is the appropriate intonation just after the system has finished reading a list of items to the caller. And if the system has already told the caller this and the caller asks for the 'next one,' then (1b) is the appropriate intonation, because it emphasizes to the caller that there *isn't* a next one, since the one just played *was* the last one. However, if the system were to use the intonation in (1b) the first time it finishes reading through the list, it would sound awkward and annoyed. Having this awareness of the role prosody plays in conveying meaning and emotion is critical when coaching voice talent – especially since often times prompts are recorded separately out of context.

Syntax

Syntax is the study of word order and sentence structures. Word order patterns of course vary from language to language. For instance, in English the most common way of saying dates is Month Date Year

(April 10th, 2010). Correspondingly, a common touchtone input format is MMDD(YY)YY. Whereas in Romance languages, like French and Spanish, the word order is Date Month Year (e.g., *Le dix Avril, 2010*). And the most common touchtone format is DDMM(YY)YY. So, in creating a French or Spanish version of your automated phone system, the VUI and grammar design and implementation teams need to be aware of these cross-linguistic syntactic differences.

Semantics

Semantics is the study of the literal meaning of words and phrases. (We'll examine non-literal interpretations under Pragmatics.) In designing a speech system, ambiguous menus and commands should be avoided to minimize caller confusion and increase successful interactions. For example, after playing information to a caller, some automated systems follow this with the instructions shown in (2a):

(2) a. Please say 'Repeat' to hear that again.
　　 b. Please say 'Repeat to hear that again.'
　　 c. To hear that again, please say 'Repeat.'

However, callers have been known to misinterpret (2a) instead as the expression as in (2b). As a result, some callers literally respond, 'Repeat to hear that again.' If the VUI designer and grammar developer did not anticipate this sort of semantic interpretation, then the speech grammar may only recognize 'Repeat.'; the alternative 'Repeat to hear that again.' would be out-of-grammar and treated as a misrecognition. A better VUI design approach is to use the sentence structure shown in (2c), which is semantically unambiguous.

Pragmatics

Pragmatics is the study of the ways in which context contributes to meaning. It examines inferred meaning, beyond the literal semantic interpretation of the words. It's the difference between 'what is said' versus 'what is meant.' A VUI designer needs to consider the nuances and implied meanings based on how a prompt is written and delivered by the voice talent. (Again, prosody greatly impacts pragmatic interpretation.) For instance, one of the problems with an error prompt such as:

(3)　　　The PIN you entered is invalid. Please enter a valid number.

is that it emphasizes the caller and their action ("you entered") and is a negative statement. So, it has overtones of the caller being bad, wrong, and at fault. Furthermore, it goes on to overtly state that the caller should enter a *valid* number, as if to imply that the caller couldn't figure out on their own that a valid number is required.

Socio-linguistics

Socio-linguistics is the study of how society affects the way language is used. It examines the interaction among factors like cultural norms, expectations, social classes, gender roles, hierarchical relationships, age, etc. in terms of language. These factors also need to be considered as part of VUI design. For instance, if the recorded voice of the automated system is female, and the system is used in regions of the U.S. with a high percentage of speakers of southern dialects, then the yes/no grammar needs to contain "Yes, ma'am." as one of its recognized phrases. Also, there are well studied and documented

differences in the way men versus women generally speak.[2] So, if the voice of your automated system is a female, and your VUI designer (or design team) is male, your designer needs to be able to "style-switch" in terms of their prompt writing style, so that the wording of the prompts matches the persona and gender of the automated system.

Linguistics Adds Value

All of the fields of linguistic study provide added insights that can enhance VUI design. Could some of these insights also be discovered through trial-and-error user-centered design and evaluation techniques? Yes, some. But having a VUI designer with linguistic training helps the designer anticipate linguistic factors which could impact the design. Linguistically-aware designs are more efficient and ultimately more usable, because they leverage an understanding of real human language, which is the implicit *mental model* when interacting with speech-enabled automated systems. (Not to imply that the VUI design can or should mirror real-human conversations 100%. VUI designers of course also need to be aware of the capabilities and limitations of automated speech technology, and factor these into the optimal design as well.)

Why is Linguistics so Threatening?

Given that linguistics can be a very informative knowledge-base as part of one's multi-disciplinary approach to VUI design, why does there appear to be such a backlash against this field of study? Why have the interviewees quoted in the *Speech Technology Magazine* article used such stark language as "Linguistics has nothing to do with user interface design." Or "The linguists don't know anything about design."? What's so threatening about linguistics that these VUI experts see absolutely no value in it??

Frankly, we two linguists are a bit at a loss to explain those statements. But, we suspect that it may be related to the fact that linguistics is a rather arcane knowledge. Linguistics isn't readily obvious to a lay person, and most of it can't be picked up off the street. Even though we're all "experts" at speaking our native language, that does not entail that we're "experts" at understanding the underlying rules, patterns, and properties of human language. This is something which must be dissected and formally studied. So, just like computer science may seem rather cryptic to someone who has never studied programming before, linguistics may seem rather opaque – in fact potentially even irrelevant – without the benefit of having had linguistic training.

Conclusion: It's Not *Just* Linguistics

To be clear, we are _not_ advocating a linguistics-only approach to VUI design. What we and our colleagues can all agree upon is that VUI design requires a multi-disciplinary approach. Our advocacy of linguistics as part of VUI design should in no way be interpreted as denouncing the value of many other relevant fields – e.g., design, human factors, computer science, psychology, business, marketing, drama, music, etc. But linguistics should not be cast aside as irrelevant.

---

[2] Deborah Tannen, a Georgetown University linguistics professor, has written a series of books on this topic.

Spoken language is the medium of VUI interaction. It behooves VUI designers to have some understanding of how humans produce, perceive, and interpret human spoken language. A VUI design team needs some appreciation for the underlying discipline of linguistics to ensure that human speakers will be able to use the system – as linguistics can impact usability and business results. You _do_ need a multi-disciplinary approach. Just don't forget to include linguistics.