## Background and career journey of Katarzyna

8:06  
**Alexey**This week, we'll talk about human-centered AI for disordered speech recognition. We have a special guest today—Katarzyna Foremniak is a computational linguist with over ten years of experience in NLP and speech recognition. She has developed language models for automotive brands like Audi and Porsche and specializes in phonetics, morpho-syntax, and sentiment analysis. Katarzyna also teaches at the University of Warsaw and is passionate about human-centered AI and multilingual NLP. Welcome to the show!

8:41  
**Katarzyna**Thank you. I'm very happy and honored to be here with you today.

8:47  
**Alexey**How accurate was the bio? I asked GPT to summarize my longer bio, so I hope it was accurate.

8:56  
**Katarzyna**Yes, it was accurate. It was quite rich for a summary.

## Transition from linguistics to computational linguistics

9:06  
**Alexey**Before we dive into our main topic of human-centered AI and speech recognition, let’s start with your background. I think GPT already provided a lot of good insights, but could you tell us more about your career journey?

9:21  
**Katarzyna**Sure. I’m a computational linguist, which is a short answer. On one hand, I’m a researcher and a teacher at the Department of Italian Studies in the Faculty of Modern Languages at the University of Warsaw. On the other hand, I work on NLP projects in the automotive industry and collaborate with companies in data handling. My background is mainly in linguistics, which was my starting point. I studied Italian and Polish in parallel and added some technical skills along the way, leading me to my journey as a computational linguist. My main field of interest is phonetics, which is why this topic is relevant for today's meeting. Sorry for being a bit lengthy!

10:42  
**Alexey**It wasn’t long at all. I’m curious, to be a linguist in Italian, do you have to speak Italian perfectly?

10:55  
**Katarzyna**It helps!

10:58  
**Alexey**Do you speak it well?

10:58  
**Katarzyna**Yes, I do.

11:00  
**Alexey**I usually go to Italy as a tourist. The funny thing is that I live in Germany, and when I go to places like Garda, which is a popular vacation destination for Germans, people just look at me and start speaking German.

11:21  
**Katarzyna**Maybe that’s why! Italians are usually very open and appreciate it when you speak their language.

## Merging linguistics and computer science

11:38  
**Alexey**I’m curious about how difficult it was for you to transition from linguist to computational linguist. Linguistics, as I understand it, is less mathematical than other disciplines.

11:47  
**Katarzyna**There was an audio problem; I only caught the first part of your question.

11:53  
**Alexey**How difficult was it for you to become a computational linguist? It seems like linguistics isn’t as math-heavy, so how was the transition for you?

12:25  
**Katarzyna**You raised two interesting points. First, linguistics can indeed seem distant from mathematics. However, it depends on the approach; literature might seem far removed, but if you focus on data and relationships between datasets, it’s a different story.

13:22  
**Alexey**You’ve touched on how important the use of data is. Is it safe to say that computational linguistics merges linguistics and computer science?

13:34  
**Katarzyna**Yes, that’s correct. When I began, there were not many studies in computational linguistics in Poland. At that time, it was necessary to acquire programming skills along with linguistic knowledge. I had to catch up and was fortunate enough to meet several people who were willing to teach me.

13:47  
**Alexey**When you say that linguistics might not be too far from mathematics, I remember reading about syntax trees. They represent language in a more concrete way, almost like algebra, allowing us to work with mathematical abstractions instead of just letters and characters.

14:32  
**Katarzyna**Indeed, that’s one of the main perspectives in linguistics: viewing language as a structured system. This aligns closely with your description.

14:47  
**Alexey**When you were studying linguistics in Italian, did you learn Italian first and then focus on linguistics, or did you approach it from a linguistic perspective before using the language more practically?

15:04  
**Katarzyna**I started learning Italian first. I managed to communicate in the language before diving deeper into its linguistic aspects. I also specialize in Polish linguistics, which felt much more natural since it’s my first language.

## Understanding phonetics and morpho-syntax

15:25  
**Alexey**In your biography, summarized by GPT, it mentions that you specialize in phonetics, morpho-syntax, and sentiment analysis. I’m familiar with sentiment analysis, but could you explain what phonetics and morpho-syntax are?

15:47  
**Katarzyna**Sure! Phonetics is the study of sounds in the language system, focusing on how we produce speech. It often intersects with phonology, which explores sounds on a mental level. Morphology, on the other hand, deals with how words are formed and how they interact with one another. In morphologically rich languages like Polish, this includes inflection and the use of prefixes and suffixes. It’s essential to consider both phonetics and morphology, as many aspects of our speech are interconnected.

## Exploring morpho-syntax and its relation to grammar

17:28  
**Alexey**So we have linguistics, and within linguistics, we have morphology and phonetics, right?

17:34  
**Katarzyna**Yes, we can delve even deeper into other areas like semantics (the meaning of words) and pragmatics (the use of language). However, for today, I believe phonetics and speech are the most relevant.

17:58  
**Alexey**I’m not sure if you answered this already, but what exactly is morpho-syntax?

18:03  
**Katarzyna**Morpho-syntax combines morphology and syntax. It studies how words are constructed and how they fit into sentences, which are larger segments of text. There's a strong connection between how we use words and their arrangement in the overall structure.

18:34  
**Alexey**How does morpho-syntax relate to grammar? Are they basically the same thing?

18:39  
**Katarzyna**Grammar is linked to both morphology and syntax. When we discuss grammar, we often start with the word itself, looking at its structure and inflection, such as verb conjugation or tense.

19:00  
**Alexey**So syntax is a higher-level concept, then? Each word is correctly formed and used, and syntax governs how we arrange them in sentences.

19:18  
**Katarzyna**Exactly. Syntax is all about constructing sentences and how words relate to each other within those sentences.

19:24  
**Alexey**Is syntax also related to word order? For example, in German, the verb must always be in the second position, right? That’s part of syntax, correct?

19:34  
**Katarzyna**Yes, that’s a syntactic pattern.

19:41  
**Alexey**For me, syntax was always associated with programming languages. For instance, in Java, you have to use curly braces. I didn’t realize that the concept originates from linguistics.

20:03  
**Katarzyna**That’s a great example! It illustrates the same patterns we see in both natural languages and programming languages, such as what comes first and what should follow.

## Connection between phonetics and speech disorders

20:33  
**Alexey**When it comes to phonetics, it’s about how we pronounce words, right?

20:40  
**Katarzyna**Exactly, it’s about the production of sounds.

20:42  
**Alexey**And we want to discuss speech disorders, so how are phonetics and speech disorders connected?

20:51  
**Katarzyna**If we speak, it can be in a standard or non-standard manner. Speech disorders are examples of communication disorders. They can affect articulation, meaning the pronunciation of sounds and sound clusters, as well as fluency and voice quality. We might articulate a sound incorrectly, or have issues with fluency, such as interruptions or stuttering. These disorders can lead to difficulties in understanding speech.

23:19  
**Alexey**Did I understand correctly that an accent can be seen as a speech disorder, or is it simply a normal variation?

23:34  
**Katarzyna**When we talk about foreign accents, we don’t typically classify them as speech disorders. They may lead to comprehension difficulties, but they’re generally accepted as part of normal variation. Speech disorders, on the other hand, are often linked to biological or neurological causes that hinder normal speech production.

## Improvement of voice recognition systems

24:41  
**Alexey**That’s interesting! I remember that even five years ago, many voice recognition systems struggled to transcribe what I said accurately. For instance, YouTube's transcription was quite poor in the beginning. But now, with systems like Whisper, they work much better with my accent. Perhaps my accent has improved too, but I think the systems have advanced more significantly.

25:23  
**Katarzyna**That’s true! The models have improved considerably. Any non-standard speech can present challenges, but we generally don’t view foreign accents as a problem. We can choose to work on them if we want, or embrace our accents as part of our identity. However, discussing speech recognition for disordered speech is very similar to recognizing atypical speech, including child speech, foreign speech, and idiosyncratic pronunciations. Dialects and variations also fall under this category.

26:22  
**Alexey**For example, the Scottish accent.

26:26  
**Katarzyna**Some languages have rich regional varieties, and others do not. In the case of Scottish, a local variant can differ significantly from the so-called standard.

26:46  
**Alexey**I remember watching a video from the British Parliament where a representative from Scotland spoke, and nobody could understand him. They asked him to repeat himself several times. It was quite amusing!

27:06  
**Katarzyna**In such cases, it’s often best to use the standard version for effective communication, but that’s not always possible. Sometimes, not conveying everything can have its advantages.

## Overview of speech recognition technology

27:31  
**Alexey**Can we talk about speech recognition? How does it work in general? Before diving into the differences between typical and atypical speech recognition, what’s the proper term to use for standard speech?

27:54  
**Katarzyna**That’s a very timely topic, as there’s so much change happening in this area. We have large language models with generative AI. Traditionally, models were trained on precise datasets, typically collected according to a prepared scenario. A good example of standard speech is data collected from speakers like those from the BBC.

28:35  
**Alexey**So if we consider British English, that would be a standard reference, right?

28:35  
**Katarzyna**Yes, usually these datasets are designed to reflect standard speech. The systems are trained to map the output of spoken language to the model phrases they were trained on. However, with the advent of large language models, we can incorporate contextual training, which improves recognition. There’s hope for better recognition of atypical speech as well.

28:54  
**Alexey**Right. It’s important to distinguish between standard and atypical speech.

29:02  
**Katarzyna**Yes, and it’s essential to recognize that we all speak atypically at times. Factors like speaking fast or slow can affect clarity. Automatic speech recognition systems are typically trained on data from speakers without speech disorders. Therefore, they can struggle to recognize atypical patterns, which leads to poor accuracy.

30:12  
**Alexey**So if I understand correctly, ASR (automatic speech recognition) systems struggle with speech that deviates from what they were trained on.

## Challenges of ASR systems with atypical speech

30:24  
**Katarzyna**Absolutely. This misalignment between training and real-world usage can lead to significant recognition challenges.

30:33  
**Alexey**So what can be done about this?

## Strategies for improving recognition of disordered speech

30:53  
**Katarzyna**There are several strategies we can implement. One approach is to collect and curate specialized datasets that include speech from individuals with various disorders, using this as a subset in training. We can also employ transfer learning techniques to adapt models that were trained on standard data for individuals with speech disorders.

37:02  
**Alexey**What if we don’t have such data available?

## Data augmentation for training models

37:07  
**Katarzyna**If data collection is challenging, we can utilize data augmentation to expand the training dataset by artificially simulating disordered speech. For example, if we know specific sounds or consonant clusters are problematic, we can create artificial variations.

37:31  
**Alexey**That’s interesting.

37:33  
**Katarzyna**Another strategy is using multimodal outputs. While we learn from audio, adding visual data—such as lip reading or gesture recognition—

## Transfer learning in speech recognition

40:17  
**Alexey**Yeah, not yet, of course. But I've worked with images, and in a typical situation, you have an ImageNet neural network trained on ImageNet. Then you have your own data, which could be tractors or anything else not included in ImageNet. You might find tractors there, but also something else. You take, say, 1,000 examples and fine-tune your network that was trained previously on ImageNet. The same process applies to speech. There is a model trained on standard data, and then you collect disordered speech. It doesn’t have to be a very large sample, like with images, and you apply transfer learning to fine-tune your model, right?

41:10  
**Katarzyna**You mentioned data collection, and I said it’s not always easy because of the variety of speech disorders. For people with motor speech disorders, it can be challenging to organize the entire collection process. Also, considering we’re dealing with health issues, GDPR regulations come into play. It would be ideal for research purposes to have large corpora of disordered speech.

41:55  
**Alexey**In terms of transfer learning, we don’t need a lot of data. A few hundred examples are usually sufficient to get started, at least with images.

42:09  
**Katarzyna**Indeed. However, when we consider different disorders, languages, and non-English languages, the volume of data needed becomes substantial.

## Challenges of collecting data for various speech disorders

42:18  
**Alexey**You need to account for each subset. For instance, stammering is a relatively common disorder. It might not be too difficult to gather data from American English speakers who stammer. But if we look at a particular dialect or a less common language—

42:47  
**Katarzyna**For example, take a language that isn't widely spoken.

42:48  
**Alexey**That's a good example. An interesting case is bilingual individuals, as studies show stammering can occur in one language and not in another. This is linked to specific semantic clusters or syllables that may be more common in one language than in another.

43:31  
**Katarzyna**There are significant differences in these cases.

43:53  
**Alexey**Exactly. I sometimes experience this myself. Maybe it’s because I'm thinking about what to say next, and then I start stammering. It doesn’t happen often, but I’ve noticed it happening to others, too.

## Stammering and its connection to fluency issues

44:31  
**Katarzyna**Fluency issues are normal human behaviors, but stammering should be diagnosed as it differs from what most of us experience. Fluency is more common when using a foreign language; we often need time to find the right words.

44:43  
**Alexey**With stammering, individuals know what they want to say, but they struggle with certain sequences of sounds.

44:48  
**Katarzyna**Yes, it can block speech at the beginning of a sentence, while the rest flows fine. Sometimes, it’s linked to specific consonant clusters that are tough to pronounce. If a language has many of these, like English or Polish, it can be challenging.

## Polish consonant combinations and pronunciation challenges

45:16  
**Alexey**Polish does have many challenging consonant combinations. I recall a funny story from Krakow about a central street called "Czarnowiejska." It's packed with consonants, and it’s always amusing to see a British person trying to pronounce it.

45:51  
**Katarzyna**Exactly! There’s a trick with certain sounds, where two consonants together create one sound. This can make pronunciation tricky.

## Use of Amazon Transcribe for generating podcast transcripts

46:17  
**Alexey**By the way, I use automatic speech recognition for podcast episodes after recording. I utilize Amazon Transcribe, which is supposed to recognize English.

46:32  
**Katarzyna**So, you’re using that for generating transcripts?

46:34  
**Alexey**Yes, but it expects English. Then, all of a sudden, I might throw in a Polish word, like "sensu." I think it’s interesting how LMs are being used more frequently in this area. I do this as well. After getting the output from Amazon Transcribe, I feed that into an LM or an LLM. Even if the transcriber misses something, the LLM, using the surrounding context, can often correct it.

## Role of language models in speech recognition

47:28  
**Katarzyna**Absolutely. LMs can assist at various levels. Adding context is crucial for recognition and transcription. We should focus more on meaning preservation rather than traditional metrics like word error rate or accuracy.

## Contextual understanding in speech recognition

49:19  
**Alexey**That makes sense. In my case, I had an interview about Scikit-learn. Many in the machine learning community know it, but the speech recognition system struggled to identify it correctly, especially since I was speaking with a French accent. However, because we provided context in our discussion, the LLM was able to infer the correct term.

50:29  
**Katarzyna**That's a fantastic example! This two-step process—first speech recognition and then using an LM—works well. There are also models that can integrate both steps into one.

## How voice recognition systems analyze utterances

51:27  
**Alexey**I suppose it’s about understanding how voice recognition systems operate. They analyze utterances, right?

51:44  
**Katarzyna**It can be anything; it can be a word, it can be a phrase. A phoneme is what...

51:50  
**Alexey**One sound, but like a word that I say, an utterance, right? For example, I have a disorder; I say "whiskey" instead of "whiskey." If I have a language model in my voice recognition system, it can understand from the context that I'm not talking about the alcoholic beverage, but rather about something that carries a lot of risk. Because it has access to the other words, the goal of a language model is to predict the next word based on the context. We can utilize language models, and we probably do, in voice recognition systems, right?

52:31  
**Alexey**It can be trained to predict it or, in post-processing, it can say, "Okay, it’s strange that it’s in this context; it should be something else." All predictions we make are what happens next, as LLMs do. It can indicate that "risky" should come instead of the beverage from your example. So, it can be done before and also at the second step. One thing that came to my mind now, if we still have time, is that ASR models can also be personalized for specific individuals with specific disorders. In this case, this first stage, the training, and preparing the model to expect atypical productions or articulations can be done. It’s a bit different if we want to have a model that recognizes both standard speech and atypical productions.

## Personalization of ASR models for individuals

54:05  
**Katarzyna**I guess with personalization, the way it works is I first need to train it as a user. It asks me, "Hey, can you pronounce this sentence?" I record myself saying the sentence, and then it asks me to pronounce something else. I do this for about ten sentences, and then it builds a model tailored specifically to the way I speak.

54:35  
**Alexey**Individual speaking features, yes, tailored to your style of speaking.

54:42  
**Katarzyna**There was a system like that before Whisper existed. I think I even tried it. But now with Whisper, it’s doing such a good job. I use it in ChatGPT; when I dictate whatever I want, it recognizes it, and even if it misunderstands something, ChatGPT figures out what I want. So it's quite convenient.

55:13  
**Alexey**Amazing! But, I know, it’s a bit terrifying. It knows me so well—maybe better than I know myself!

55:23  
**Katarzyna**I assume that when we talk about speech disorders—not accents or similar issues—having a personalized model is quite useful, right?

55:39  
**Alexey**Of course. Especially when it’s difficult for you to use standard models. The personalized one can be a tool for communication. Today, we also discussed speech disorders, but there are also language disorders, especially as a result of neurological diseases, where individuals struggle to find the right words or use words in context. These issues are more connected to the content itself than to articulation.

## Language disorders and their impact on communication

56:25  
**Katarzyna**How does it work? For example, if I have a disorder, I cannot pronounce a specific word?

56:33  
**Alexey**You might not remember the right word in context, for example.

56:39  
**Katarzyna**That happens to me all the time.

56:42  
**Katarzyna**Yes, it occurs to all of us from time to time.

56:46  
**Alexey**Especially in foreign languages.

56:47  
**Katarzyna**If it happens too frequently, it can become problematic.

56:52  
**Alexey**When it happens in your native language and interferes with your daily life, then it’s a problem.

56:59  
**Katarzyna**Absolutely. Then it’s a real issue because stuttering and other speech problems can occur to anyone, and that’s completely natural. However, if it becomes excessive, it leads to communication problems.

57:24  
**Alexey**Maybe this is something I should have asked you at the beginning: how do we actually use this? What are the applications? I might guess, given your work with the automotive industry, that what we discussed could be used in cars. Can you provide some examples?

57:50  
**Katarzyna**When you ask me about where speech recognition models can be used...

57:58  
**Alexey**Especially regarding disorders.

## Applications of speech recognition technology

58:00  
**Katarzyna**Of course, in cars. But that wouldn’t be my first answer for people with speech disorders. It can be used as a communication tool because sometimes it’s hard for humans to understand atypical speech. A pre-trained, personalized model can do this more easily. So that would be my primary answer, as I think that’s the most important application today.

58:47  
**Alexey**How does it look in practice? Like, let's say I have an app on my phone, and I speak to the app, and then I show the result to someone I want to communicate with.

59:00  
**Katarzyna**For example, it's not common yet—common enough, but we're talking today about...

59:04  
**Alexey**I do this all the time with German.

59:09  
**Katarzyna**But we are not— I think you see any serious pronunciation problems in your case. We are discussing human-centered AI today, and yeah, AI shouldn't be the obvious one. The number of centers that dedicate their work to human-centered AI proves that this is an important topic and not obvious for everyone. Human needs should come first. It's not yet standard to give each person with serious disorders this kind of tool. But I think there is hope.

59:55  
**Alexey**What you're working on right now to make it possible.

59:58  
**Katarzyna**Yes, that's one small step. But...

60:02  
**Alexey**Are these models heavy? Because especially when it comes to personalized models and fine-tuning—if we talk about mobile devices, then these apps need to run on devices, and not everyone has the latest iPhone Pro. They have to be conservative regarding resource consumption. I guess that's quite challenging.

## Challenges of personalized and universal models

60:34  
**Katarzyna**It's a challenge. One is creating a personalized model for a person—that's quite doable. But creating one that can be universal for many speech disorders makes things much more complicated. We’ll probably include some readings in the episode description. There are some studies from Asia for Korean and Chinese that managed to create such tools. We often talk about European languages, but let’s not forget about what’s happening a bit farther away from us.

## Voice recognition in automotive applications

61:23  
**Alexey**And when it comes to cars, how is it used? How is voice recognition used in cars? Like, play Spotify with Alexa or...?

61:36  
**Katarzyna**Yeah, you can ask it to behave in a certain manner. For example, if it...

61:48  
**Alexey**If I cannot pronounce "R," then please park, and the car has no idea what I'm talking about.

61:53  
**Katarzyna**And it's parking, and it's parking! Everything you need and what is planned by the producers and car designers includes opening the windows, air conditioning, seat heating, steering wheel heating, radio, calling, etc. That’s also an interesting example—not for disordered speech but for recognition in general. For years, it was part of IR systems that developed a bit slower because the purpose was to make it work all by itself, even when not connected to the internet. So these built-in systems were more traditional, based on n-grams. Now things are changing, and LLMs are being introduced, so recognition is improving. Still, there are many videos on YouTube with Porsche drivers and others trying to convince their cars in Polish, Italian, Spanish, etc., to do something, and the cars are doing completely different things. So there is so much work to be done.

## Humorous voice recognition failures in cars

63:27  
**Alexey**There’s this hilarious video with two Scottish guys trying to go to the 11th floor in an elevator using voice recognition. Have you seen that?

63:37  
**Katarzyna**No, you have to send it to me!

63:40  
**Alexey**I’ll send it if you just Google...

63:42  
**Katarzyna**Alex, we’ll put it in the post like the icing on the cake.

63:52  
**Alexey**If I just Google "11 elevator," then there’s a video. I’ll include this in the description; it's amazing. So, I think we’ll run out of time.

64:10  
**Katarzyna**Sorry, but it was fantastic!

## Closing remarks and reflections on the discussion

64:13  
**Alexey**I think we covered only three questions out of—I don’t know how many we prepared, but it was...

64:18  
**Katarzyna**I think that's good! There’s still something to think about, something to read, to dive deeper into.

64:27  
**Alexey**I’m going to send the video to you right now. Then for the rest, I'll send it to Zoom, and I will also put it on YouTube. That should be a good way to end this interview today. There’s still so much to discuss. I think this video is like ten years old, if not more. You probably know the systems are better now.

64:58  
**Katarzyna**Ten years old and it’s still valid. We still have the same problems. I think the situation is a bit better, but probably such things still happen.

65:13  
**Alexey**Thanks, Kasha, for joining us today, for answering questions, and for sharing your experience and what you work on. That was amazing! I really enjoyed this, and I'm sure everyone else did too.

65:25  
**Katarzyna**Thank you. Thank you for the invitation, and really congratulations on the great series of podcasts, but also for the fantastic platform that you created. I feel really impressed, and as I said at the beginning, I feel honored to be here.

65:42  
**Alexey**Thank you! If you happen to be in Berlin, please let me know. If I also go to Warsaw, I’ll say...

65:48  
**Katarzyna**Hi! Please do. Please come to San.