**Globally Speaking**

**Podcast 88 Transcript**

**Talking Transcription Part Two**

**Featuring Jason Chicola of Rev.com and Temi.com**

Michael I’m Michael Stevens.

Renato I’m Renato Beninatto.

Michael And today, on Globally Speaking, we’re taking a look at the future of transcription. How does automation play into our space? Are people going to be outdated and not used? Or is there a way that we can work smarter and better? And we’re going to talk to a technologist. Renato is like a kid in a candy store during this conversation about a specific company that’s doing some interesting things related to transcribing our content.

Renato So, let our guest introduce himself.

Jason Hi, I’m Jason Chicola. I’m the founder of rev.com. Rev’s mission is to create great work-from-home jobs powered by AI. Today, we’re entirely focused on the voice-to-text market, which means we offer services to convert voice to text. Those include audio transcription, closed captioning and foreign subtitling. We also have some AI services that I think you’re familiar with, where we can transcribe audio automatically—that is, without a human—and currently those services are offered under the brand name temi.com. T E M I dot com.

Renato Very good, and how did you get into this space? What attracted you to this voice-to-text environment?

Jason You know, it’s a bit of an unusual path. You know, the classic origin story is like Uber, where you had Travis and I think it was a guy named Garrett Camp in Paris, and they couldn’t get a taxi. So, they said, “Wouldn’t it be great if we had an app that would send SMSs to the taxi companies,” and they formed Uber. So, they tried to solve their own problem. That’s not how we built Rev.

The origin of the company runs through the history of a company called Upwork, which is the largest marketplace for remote labor, where I was an early employee and cofounder going back about 15 years ago. And the company at the time was called oDesk, and oDesk was a platform, and Upwork is a platform, where you can hire programmers, designers around the world. And it’s a website that connects supply and demand. They create market mechanisms whereby you have feedback scores on the workers. It’s designed very much like eBay, and that company’s been pretty successful; it went public late last year.

And when I was setting up to start a new company in the end of 2010, I knew that remote work was going to be this powerful trend in my lifetime, because there’s so much great talent that lives far from a big city, and the internet will ultimately make it possible for work to come to the people, as opposed to forcing people to drive always to an office. You know, I look at the world and see so many people spending an hour or more a day in traffic and limited to jobs within a 20-mile radius of their house. And I think it’s temporary.

I mean, in the same way that you no longer have to shop at the stores by your house, you can order from Amazon, I think it’s likely that before I’m dead, there’ll be a billion people who work for customers that aren’t 20 miles from their house. And, you know, I’m kind of devoting my career to accelerating the migration from, you know, local work to remote work.

We kind of started with that concept. As we thought about how would we make remote work better than other platforms, what we felt was that for remote work to really scale, the people paying for the services were going to want quality to be guaranteed. That the existing, you know, what I call the first generation of remote-work platforms like Upwork or, say, even Fiverr, they don’t really guarantee quality. Quality’s kind of hit or miss. And so, we felt that we should pick a service where we can guarantee quality.

And we did a couple of experiments; we tried a few different things. And we’ve had the most success with voice-to-text. It just turns out that it’s a big market, that technology can be really helpful and that it’s possible to develop standards, because what you might want in a transcript and what somebody in Oklahoma wants in a transcript are probably the same thing: you both want it to be accurate to the words that were spoken. So, it really just checked all the boxes of being a really great market.

And we sort of view it almost as a test case for our business model, which is to provide remote work, but to do it in a standardized way with SKUs where we can guarantee quality. And we have aspirations to do other services over time, but we found so much opportunity in the voice-to-text market that we’re going to stay here for a little while before we add other things.

Renato Well, one of the things that I notice, and we here at Globally Speaking use the services of Rev occasionally to transcribe our podcasts and share them with our listeners. But one of the things that is also interesting in your website is that you provide translation services as another similar service that you provide on your platform. How big is that part of the business and why did you choose that as the other service to provide on your platform?

Jason So, it’s a small part of our business today, but what is a bigger part, and a rapidly growing part, is what we call foreign subtitles. Where if you have a movie, let’s say, that’s in English, and you want the words in, let’s say, Spanish, we can have somebody watch the movie, type the words out in Spanish, so that you can, for example, distribute it in Latin America, whether it be on YouTube or Netflix or iTunes.

So, our document translation business was our first business that we launched, and, you know, candidly, it didn’t meet our financial targets for a lot of reasons. I think we weren’t probably in the right market segment and we largely put it on the backburner as we saw bigger growth opportunities elsewhere. We view foreign subtitling as a phenomenal market simply because there’s so much content and video content that is being distributed globally, there’s just a big demand for it. It’s kind of at the intersection of transcription, which is mostly what we do, and translation, which I think is what your podcast typically covers, because it involves a little bit of both.

Renato Well, transcription is one of the foundations. I mean, there is a lot more content in video and audio format that that needs to be transcribed before it can be translated and subtitled again. And actually, one of the questions that I wanted to ask you is—because I had that problem literally earlier this week—I received three videos in Portuguese from a friend and she was asking me how do I get this transcribed? And the first thing that I did: I went to your platforms to check whether you did it in Portuguese and you don’t; you’re only doing transcription in English. Isn’t the technology available already to do it in other languages?

Jason Well, technology is, but our transcription, the bulk of the work, is done by humans, and so we, as you rightly point out, today, we only transcribe and caption in English, English audio going to English text. That’s not because we don’t believe your friend’s use case is quite important, but simply because we have to focus on what we think are the best opportunities for us at the moment. And we see such great opportunities in the English market that we’re going to stay put there for a bit.

I mean, as you might imagine, having to provide that service that you just described, to transcribe this Portuguese video, we need a different workforce than we currently have. I mean, clearly, we could develop that workforce, but we haven’t yet, and we have such great customer interest in our English offering that we’re focused on optimizing it before we add other languages. Although we’re clearly going to do it.

Renato You mentioned in your introduction that you have this other website, temi.com, which I must confess I have also used. Tell us a little bit about it and how do you see the implementation of AI in this type of service? So, what are the challenges, because they’re very, very close and similar to what we have in the language services business as a whole?

Jason Yeah, you’ve asked probably one of my favorite questions, because every day I open up the news and I see some article about how robots are going to take away all the jobs, and robots are going to come for us. Everybody watched the movie “Terminator,” and they associate AI with this sort of dystopian nightmare scenario, and that’s not how we see it, right?

We see technology making everything cheaper, better and faster, and so, I think we’re seeing AI have a couple of pretty significant effects on our market. So, first, let’s talk about what you just said. Our most popular offering is human transcription which we sell on rev.com for a dollar per minute, and we guarantee accuracy of about 99%. So, it’s pretty accurate.

Some customers, let’s say podcasters, if they have a larger budget, they will use Rev to transcribe it. In fact, if you made a list of the top podcasts in the country, things like “This American Life,” you’d find that many have used us. In fact, “This American Life” is a Rev customer.

Renato We have something in common with “This American Life” now. [Laughs]

Jason Yes, that’s right, and there’s many other top podcasts and podcast networks, like, I think, Gimlet Media and others, that use Rev to transcribe their podcasts. It’s great, but what about the podcaster where it’s one guy in his garage with, you know, no sponsors and no advertising, but he still wants to do a good job of this podcast? A dollar a minute is too expensive for him, because a 40-minute episode costs 40 bucks, and he doesn’t have 40 bucks to transcribe it.

So, he can use Temi, temi.com, which is our cheaper, automated service, and he can get a transcript that is probably going to be 90-plus percent accurate if he’s recording good podcast audio. Because with AI, the thing you have to know is garbage in/garbage out. If you produce a really clear recording, you get a pretty good transcript, and vice versa: if you produce kind of a crappy recording with bad audio, you get a crappy transcript.

He can transcribe that 40 minutes for one tenth the price that he would pay Rev for the human service, and so that would cost him $4 instead of $40, and that price difference is a big deal. You know we also offer software at no additional charge that helps the customer to correct the transcript and make it, you know, accurate for their needs and share it with coworkers and do all the kind of collaboration stuff that they need to do.

I would point out for your listeners that if they use Rev, I’m happy to report that by the end of next month, that Temi service will be available on Rev, from your Rev account. So, you’ll be able to use all of our services from one location. We launched it under this separate name, Temi.com, because we weren’t sure if it would be a success, and we wanted to make sure people liked it before we brought it over to our main site.

Michael Into the main brand.

Jason Yeah, the mothership.

Renato But one of the things that I love about Temi is that you submit one-hour worth of recording and you get the transcription back in ten minutes because you don’t have that human limitation of somebody needs to listen and type it. And it’s really, really fast because you’re just looking at data, right? I imagine that that’s how the artificial intelligence algorithm works. It doesn’t listen to it, it just looks at the patterns and it generates a transcription really fast.

Jason You know, you’re absolutely right. In fact, hearing you say that, it makes me think that we should be charging you more than Rev because it is better in that aspect.

Renato [Laughing] It saves time, absolutely.

Jason But let me come back to your question. So, how does AI play out in the market? There’s really a couple of effects, okay. So, I wanted to first set the groundwork by explaining to your audience that we have a more expensive human service and a less expensive AI service.

We believe that one of the major impacts of this is that it’s market-expanding, meaning that many of our customers who have been using Rev for years had a bunch of audio or video that they did not transcribe because they didn’t have the budget. It wasn’t important enough. They would transcribe their most important content, but not some of the other stuff that’s kind of just okay.

Like, for example, I was talking a few weeks ago to a video production firm that uses Rev for most of their videos, but they have some nonprofit clients that have less budget and they can’t use it. They can’t afford it. So, for them, Temi is a great, great fit. So I think one of the major impacts of AI is that you can bring prices down for services, and there can be some price elasticity, which means that demand can go up in response to lower prices.

And, you know, for example of a similar trend playing out, I would point out look at what Uber has done for the taxi market as they’ve introduced lower-cost services like ride-sharing which they call UberPool. It used to be that you’d spend 15 bucks/20 bucks to get across town. Now in a lot of cities, you can get across town on five dollars using UberPool, and the result is that tons of people that before would have walked or used the bus or stayed home are traveling across town in these services. And, you know, we, I think, see the same dynamic: that by making something really good, but really cheap, people use it in ways they wouldn’t have used it before when it cost more.

So, I think that’s one of the major effects, just simply market expansion, because transcription becomes a lot easier and better and cheaper. I would, of course, point out that while the speech engine behind Temi, we believe, is best in class, it’s not static. You know, it’s getting better every few months, and so, I think if you look forward a couple years from now, you’ll see even more of this sort of expansion. A lot of people that would have used human will use that instead, and I think that that’s great for the market and the world.

We’ve talked about the customer perspective—that is, you guys are in a business where you deal with audio and video, so therefore you need our services. The other side of our business is the people that are doing the work. These are freelancers that work from home and they work flexibly, they work when they want. I’ll keep using kind of Uber metaphors.

It’s kind of like being an Uber driver, except they can do it in their pajamas from their couch from wherever they live, and they don’t need to get in their car and drive into Manhattan. They can do it from their home, which might be in rural Arkansas, and that’s a-ok, that’s perfect, as long as they have [an] internet connection, they’re good to go.

And, what we’re seeing from their point of view is that the AI is making a big positive difference, because we have, for example, right now in beta, a version of our work platform where the people on Rev, the freelancers that are typing out your podcast, have two options. They can either start from scratch, or they can start from a transcript that was done by AI and then edit it.

What we find is that while the AI is not always great, it’s usually pretty good, and it usually saves them time and allows them to work faster, earn more money per hour. And, you know, most interestingly to me, this is almost a surprising benefit: it actually makes them enjoy the job more and work longer because it alleviates wrist strain.

Think back to the times when you wrote a lot; I don’t know if you can remember being in college writing a ten-page paper. If you write a lot, your wrists get really tired, and, you know, unless you’re superhuman, you can’t do it forever. And by working from the AI, the transcriptionists are editing rather than typing, so instead of typing every word, you might edit one word in five or ten. And so, that ends up becoming a much more pleasant job for the freelancers as well.

So, I mean, there’s probably half a dozen other ways that I think AI can play out that are maybe a little more subtle and technical, but the big ones, I think, are those that from the customer side, I expect massive growth in the market, and from the provider side, I expect rising productivity and frankly, higher job satisfaction because it’s faster, it’s less painful and also, AI tends to be really good at work that is really boring and mundane. AI really excels at the work that is kind of the easiest, and that tends to be the work that people don’t enjoy quite as much.

Renato Yeah, it’s the classic clean audio in a controlled environment with the good recording, but I imagine that if you’re having to transcribe something that is recorded in the middle of the street with sirens going in the background, the situation might be a little different, and require more effort from the transcriptionist, and also from the AI in this case.

Jason Yeah, you know, you’re absolutely right. Our transcriptionists are quite good, our software is pretty good, but they can’t work magic. One of my favorite things that I see once in a while is that we’ll have a customer that will submit a file for transcription, and someone will listen to it and say the content here is, you know, indecipherable because of, say, background sirens. And the customer will say, ‘yeah, you know, I couldn’t make it out, but I thought maybe you guys could.’ And we sort of laugh. It’s like, you know, our view is like, if you can’t hear it, what powers of hearing do you think our people have? I mean, it needs to be intelligible for us to be able to transcribe it.

Renato It’s very similar to what is going on in the translation space. You can use machine translation and post-edit it, but in some situations, it’s much better to just sit down and do it from scratch without using the technology. The quality element is added by the human intervention.

Jason Absolutely, I mean, you’re totally right, there’s a lot of parallels. I’m not nearly as close to the market for translation as you are, but my sense is that in the translation market, the AI is a bit better. In some ways, it’s simpler that you start with a text file and not an audio file, because audio files can have this infinite complexity of, like, the background siren, which you don’t have when you’re going text to text. But, you know, my sense is that Google Translate is pretty good, but of course human can be infinitely better. It’s a very similar trend broadly, I think.

Renato And I imagine, I’m just speculating here, that accents also add a certain level of complexity. How do you handle that?

Jason Humans! I wish I could tell you we had a silver bullet, but you’re right, there’s four or five things that make it harder. Audio quality is kind of the big one. If the recording is done poorly, that’s going to be a problem. If you have a microphone, and somebody is eating a bag of Doritos next to the microphone, you know, you’re not going to hear a lot. But there are other challenges as well. As you say, accents can be quite tricky, for sure, as well as certain kinds of terminology. Nearly every industry has words that they use that not everybody knows.

I would say that on average, our freelancers are pretty good with accents and they’re pretty good with terminology. On terminology, there’s kind of two ways they tackle it. One is they are expected to do research. So, for example, if we have a human transcribing a recording, and someone says, “the CEO of Apple, Tim…,” and then they cough, we would expect that person to know that you can Google “CEO of Apple” and find out that the name is Tim Cook. So, they do some, I would say, basic research, probably more than customers would expect.

And then with accents, it depends on the individual. I mean, some people are great with accents; some people aren’t. We try to steer the work in the direction of folks that can handle the accents and the complexity. We’re not perfect, but we certainly make an effort, and I think by and large we can handle accents pretty well—as well as anybody can. I mean, certainly, you know, there are accents which are tricky to hear, but you can hear, and there’s accents that are indecipherable, so we can some but not others.

Renato Yesterday I was interviewing a Frenchmen and I had looked at a description from a colleague of his and the colleague said that he had learned what in Frenglish, what “Pokemon department” meant. And, during the interview, during the conversation, the Frenchmen mentioned that they had a hard time selling to “Pokemon departments,” and I understood that he meant *procurement* department.

Jason Ah!

Renato [Laughs] So, that’s something that would be hard for a transcriptioner to figure out what a “Pokemon department” was, but it was quite cute. But these are the intricacies of human communication. It happens even with native speakers in the same language; sometimes you have a hard time understanding a word here and a word there, and that’s the beauty of the complexity.

But one question that I have for you, Jason, is how does one become a Rev supplier? Do you have an onboarding process? Some of our listeners might be interested in becoming suppliers because if you can do translation, you probably can do transcription.

Jason Yeah.

Renato If you’re an interpreter, you can do transcription for sure.

Jason Well that’s right. So, there is a link to apply on our website. If you go to rev.com and you go to the footer, there will be a link that says, “Apply for work on Rev,” and there’s a couple of steps. We first have applicants apply, and to apply we have them not only provide a little bit of info about themselves, but we have them take a very short test of English grammar skills, because it turns out that having a good grasp of the English language is a really good predictor of how good you’re going to be at the job.

It’s not academic stuff, it’s just understanding punctuation, language and sentence phrasing. Because the one other thing I didn’t mention is that a lot of why humans are so important compared to AI is that humans can see patterns and read between the lines in a way that machines can’t. Humans can hear where the story is going and kind of know where the story is probably about to go next in a way that machine aren’t able to do that. And so, the better they understand language, the better they’re able to fill in the gaps and kind of interpolate the data that’s there.

So, we first do a short test of grammar, and then we ask them to transcribe a very short audio snippet. We have a one-minute, roughly, two-minute audio file that we have them listen to and type out using a simplified version of our tool. And we think that’s a pretty good way to test them because it allows them to do the actual job, see if they like it and see if we think they’re going to be good at it.

We admit a lot of people, but less than half of those that apply, to offer full disclosure. Of those who then are admitted, we have a three-level system internally for our freelancers. You can almost think of it as gamification, like three levels in a video game. So, if you apply and make it in, the first level that we have we call it Rookie. So, we say welcome to Rev, you’re now a Rookie, and we use the term Rookie intentionally because we consider it kind of a provisional status where we need them to learn a lot about how to use our tool and to prove themselves by doing, you know, good work on a handful of small, simple jobs.

While they’re a Rookie, we try to give them some of our easier, shorter jobs, and for each of those jobs, we carefully scrutinize their work and we give them a lot of feedback on what they did well and what they could do better. We try to make that process constructive in pointing out errors and mistakes so that they can improve. And if they do well on those jobs, we promote them to the second level in our hierarchy, which we call Revver.

So, the levels are: first, Rookie, second, Revver, third, Revver Plus. As a Revver, they can do nearly any job on the site, and I believe the pay is also higher than at the Rookie level. As a Revver, they, like I say, can do any job, they can work as much or as little as they want, and again, we are evaluating their work across a number of dimensions for quality, reliability, things like that. And we have certain performance standards that we hope that they meet.

If they meet those standards, all good. If they fall below those standards for some period of time, they won’t be able to continue to work on Rev. We also have a higher standard that, if they meet that, which means their work is highly accurate, they’re very responsible, they’re very reliable, then we promote them to third level, which we called Revver Plus.

Now, if you’re Revver Plus, that has some major advantages. The biggest advantage is that we give them first dibs on new jobs. So, when a new job comes in, for the first hour or so, it’s only available to the top freelancers, and that has a couple of important implications. The biggest one is that the Revver Plus group gets much more choice over what they work on.

Because we have so many audio files and video files coming through every day, and the topics range from academic to media to legal to healthcare, people really care about what they work on, both the content and also the audio length. We have a lot of stay-at-home moms on the platform that are often doing work with family obligations around them, and so they really value the opportunity to claim short jobs. Some people care more about the content, and they may say, “I really want to work on, for example, religious content because I’m religious and I love doing church work,” right. Some people will go that way.

Some people say, “I have a small baby next to me, I have a little bit of time while the baby’s taking a nap, and so I want a very short job, I don’t care what it is.” And so, having that first dibs is really, really powerful. There are some other benefits as well, but that’s kind of the overview of the system for the freelancers.

Renato So, you mentioned church, you mentioned podcasts before, what other segments are heavy users of your services?

Jason Universities are increasingly capturing lectures, recording lectures with video, and getting them transcribed or captioned. There’s certainly media, and media has many components: everything from, you know, YouTube to Netflix and everywhere in-between. So, just video on the internet is a mega trend.

There’s market research; you have market research firms out there that are doing focus groups and then they want them transcribed. There’s the legal industry, there’s depositions. For example, you have some companies that will transcribe meetings. That’s becoming a big trend, especially as people move from using phone calls to things like Zoom. It’s easier to record meetings than it used to be. Reporters will often transcribe interviews. Those are, I think, some, you know, good examples. There’s many others.

I mean, AI is becoming a big driver as well. There’s plenty of companies that are building various kinds of AI systems, and some of them will need to use human services either because the AI can’t do the job or because they want to train their AI engine and they need accurate inputs. So, those are some of the kinds of reasons that people would choose to use Rev.

Renato We talk about the power of voice and how voice is increasing as a user interface, but at the end of the day, if you want to be productive, if you want to do something searchable, practical, usable, you need text. So, it all comes back to the old text format that it’s easy for you scan and it’s much easier to be more productive with text than you would with voice. It brings us to a very human level, again.

Jason If you want to derive insight from content, you want text. I mean, I’ll give you an example of a typical use case for transcripts. If you’re making media, if you’re making video—let’s say you’re at CNN and you’re going to produce a five-minute piece that’s going to go on the air—how much video do you think you record to produce that five minutes of final product?

You might record 50 hours of content and whittle it down to five minutes that gets shown on national TV at eight o’clock at night. And so, how does somebody go from 50 hours of raw footage to five minutes of tightly edited final product? Well, you can have the editors watch all 50 hours, right, but then, you know, they’ll never get anything done. So, they want a transcript, so at CNN, there’s a lot of people that use Rev, for example, and there’s many other media companies that are doing similar kinds of analysis, and saying that if we’re going to produce great content, we need to be able to efficiently deal with all the audio/video that’s coming to us. And it’s never been easier or cheaper to capture audio and video, so the amount of audio/video is soaring, and that creates arguably even more demands, as you point out, to make sense of it all, and the text is crucial for that.

Renato Very good. Michael, do you have any questions?

Michael Haha! No, you’ve done a good job. I just let you run with it on this one.

Renato I love this topic, as my ex-wife, when I met her, she was a transcriptionist using cassette tapes. I’m originally from Brazil, and in the United States, you had pedals that you could buy, but in Brazil, it wasn’t available commercially and it was very expensive to import. So, her brother had created a pedal for her with a doorbell and he attached a couple of cables inside the cassette player, and she used that as a pedal to stop the thing. It was a cassette tape and a typewriter, so we’ve come a long, long way in this process.

Jason You know, it’s funny, so let’s talk about that. I mean, you know a lot about this market and believe it or not, a good percentage of Rev freelancers today use foot pedals, and we, early on, we wrote software that controls them. So, they can use our software to do their work and using the software that runs, like, say, in the Systray of a Windows PC, they can control a foot pedal that plugs into the USB port.

And it’s a big deal because, now, it sounds a little weird to someone that has never tried this, but if you tried to type a lot and you have to control the audio playback, you quickly realize that you’re constantly moving your hands from the typing to use the mouse to hit the play and pause. You want a third hand, but you don’t have a third hand, so you use your foot for the play/pause. That, in our opinion, is the best way to transcribe and until people…

Renato Absolutely.

Jason Until people can grow a third hand, you know, we think that’s the best way. But the other part to that, the cassette and the typewriter, those things we have better versions of today, for sure.

Renato [Laughs]

Jason I can remember when I was a kid, I had a teacher whose wife did medical transcription, and she would get cassettes mailed to her house, and she would type it up. I don’t know if she submitted on paper or electronically, but, I mean, literally, cassettes by mail. That’s how it used to work.

Renato Yes.

Jason As long as there have been governments, there have been transcripts. I mean, there have been transcriptions for thousands of years. The difference was thousands of years ago, the only things transcribed would have been a government proceeding because nothing else was of high value enough, right. And what’s changed is the cost to transcribe has come down a lot, and so more and more stuff is transcribed. So, we think that the market will keep growing as the technology makes it cheaper and better quality.

Renato Fantastic. Jason, thank you so much, this was a fascinating conversation.

Jason Hey, it’s been a pleasure.

**End of conversation**