**Globally Speaking**

**Podcast 75 Transcript**

**What’s the Latest with Neural MT?**

**Featuring John Tinsley of Iconic Translation**

Renato I’m Renato Beninatto.

Michael And I’m Michael Stevens.

Renato Today on Globally Speaking Radio, we’re coming back to a topic that was one of the most popular topics we had. We talked about neural machine translation in a three-part series back in episodes 20, 21 and 22, and it’s currently one of the topics that is most pressing when it comes to technology in the language space.

 Our guest today is working in this environment at a very high level. And my favorite part of this interview is when we talk about neural machine translation as it applies to real life.

 Michael met with our guest in Dublin and here’s his interview.

John I’m John Tinsley, the CEO and co-founder of Iconic Translation Machines. Iconic is an enterprise machine translation provider headquartered in Dublin, Ireland, with offices in Spain and the UK, but obviously, as is the nature of the beast, we operate on a global level.

Michael John, you’re one of our reoccurring guests. You joined us in the last three-part episode where we introduced the concept of neural machine translation. It seemed it was much more on the hype curve at that point. It’s been over a year now since that happened. Tell us first a little bit for those who haven’t listened to those episodes both about Iconic and neural machine translation in general.

John Back then, neural was the new kid on the block and it had made that initial splash of hype. We were in high-level expectation-management mode. Yes, it was really exciting from everyone’s perspective, but it wasn’t production-ready. It worked at the core, but there were a lot of things that needed to be done before you could incorporate it into kind of production workflow. So, we’re really trying to manage everybody’s expectations, saying yes, this is coming, but let’s work on a few things.

John So, trying to kind of be upbeat about it because it was so transformative, but also not trying to come across like being a damp squib and saying, you know, it’s not as great as…

Michael Yeah, you’re not the Luddite saying no, no, we won’t move forward.

John Somebody told me not too long ago that they were really interested in neural machine translation, but they heard me speak at a conference and decided not to go for it. And they were then surprised to hear that we were working on it because I was so negative about it before. So, I thought, ‘oh maybe I should really change my tone a little bit’ because it is actually really, really exciting and it’s just a case of trying to find the balance there.

Michael What have you guys found out? What have you experienced that has made you more positive about the direction it’s going?

John I think it would’ve been impossible to predict—and if you go back and listen to old episodes, you’ll see that we didn’t predict—how fast a pace things would develop at. Back with statistical MT, we kind of hit a plateau in terms of, you know, how far we could go with it. Improvements were very incremental, so it was really a case of ‘okay, this is what we have, what can we do with it?’

 And then all of a sudden, you know, neural comes along and in some cases, which is where the hype came from, it, like, blew things out of the water. So, you would do so much effort for such a small increment with statistical MT, and all of a sudden, we’re getting, like, order of magnitude levels of improvement and we’re like, ‘this is bizarre.’

 And that hasn’t really stopped to some extent. BLEU scores are a thing that people look at; and, you know, if you’re in research, if you’re publishing a paper and you had half a BLEU point improvement, that was enough to, like, write an article about at a top-tier conference and kind of draw attention.

 And now, we’re still, you know, we try something new and you’re getting like a 7, 8, 9-point improvement and it’s like, wow! The avenues of things to explore, things to try, they’re still opening up. That’s really exciting from a technology perspective, from an R&D perspective, and then, obviously, from the commercial potential, what more can we do with that? As opposed to saying, “this is what we have, where can we fit it in?,” now we’re looking at new doors that are being opened, new opportunities where they weren’t there before.

Michael Some of that has come from your work around the source; that there has been a focus on the quality and cleanup of the data.

John One kind of characteristic of neural machine translation that has become clear is that it’s a lot less robust when it comes to noise. Statistical machine translation, for all of its flaws, it was kind of robust. It would produce a reasonable translation for anything, you know; it would have a go. Whereas with neural MT, it knows what it knows, and if it sees something it doesn’t know, it can be a bit unpredictable. And so, very much from a data perspective, there was a ‘less is more’ when it comes to kind of quality data—so, focusing on having, you know, less cleaner data than just reams and reams of data where there might be a lot of noise in there.

Michael It might be helpful for some of our listeners when we talk about ‘that data,’ what makes up that data, what is it?

John Translation memories are a perfect example of, like, the ideal training data for machine translation. So, what we need for machine translation is examples of previous translations: here is a sentence, here is how it was translated in another language, and we learn from that. And so, TMs are the embodiment of that, and you supplement that with glossaries, terminologies for specific domains, for specific industries, and that’s training data.

 So, if you’re trying to do an online machine translation tool that is, you know, consumer-grade, can translate anything for anyone at any time, you’re going to need serious amounts of data. We’re talking, you know, hundreds of millions, perhaps billions, of words of training data, whereas if you have a more narrow focus in a specific use case, a specific industry or a specific client, then we can get away with less better data.

Michael What are some of the other things that folks in this space are talking about now? The advancements or different areas where they’re excited or seeing opportunity?

John I mentioned that it’s kind of opening up new avenues. So, one thing we used to talk about in the company was predicting the likelihood of success for a machine translation project depending on a number of factors, and the key factor being the language that you’re looking at. So, really, there was a small enough set of languages that MT could do well at and could do production-ready, and then outside of that, you were struggling.

 I mean, just to give a few examples, you know, if you’re dealing with your French, Spanish, Portuguese, German, Italian, they were kind of the much more doable languages, but I guess if you talked to anyone who tried machine translation a few years ago for Japanese or Korean or Arabic, or something like that…

Michael They would start to go cross-eyed.

John Yeah. And we saw a lot of big enterprise who had machine translation successfully deployed. It would be for seven or eight languages, and then they would use kind of traditional workflows for the rest. Now, everything is back on the table. So, for example, in the last year and a half, we’ve put engines into production for English into Hungarian, Finnish, Russian, Czech, Polish, Ukrainian, Serbian—like, languages where if you asked me two and a half years ago, we would have said, “No, probably that won’t work very well; we recommend you do these alternatives.” So, that’s probably the most exciting thing.

Michael Are you also exploring the non-English language pairs? Is there research being done in that yet?

John It doesn’t really matter what the language pairs are once you have data to feed into it. The market will always drive what language combinations you need. So now the doors are open for, let’s say, English to, you know, Korean, which was one of the really, really hard languages before, or perhaps combinations that don’t include English where there wasn’t data before. The biggest question probably being asked on the research side of the industry at the moment is: okay, we know that the technology is probably good enough to work with these languages if we had the data, but we don’t have enough data—how do we get around that problem? So, the traditional, the low-resource-languages issue.

 And so that’s where a lot of the research is being focused on: okay, what can we do with very little data; what can we do with no data? So, it’s like, we call it unsupervised machine learning. So, supervised machine learning is where you learn from previous examples of something. Unsupervised is where you kind of iterate the learning from a starting point of not really having many examples of data to build on.

Michael That’s interesting. And then you see LSPs getting involved, some of the larger ones, with the more public engines working to create some of that data. It’s just, how much do we need to have before we can get a viable engine running in this particular language pair, or from English into whatever it may be?

John So to give you a kind of a cool example of how this is working: let’s say you have a small amount of Georgian. There’s only three million people in Georgia and they only speak it there. So, it’s not a widely spoken language so there’s not a lot of data.

 So, maybe you have a small amount of English/Georgian data but not really a lot. So, you can build a very small prototypical system, let’s say. But, we have a lot of English data. We use our small, probably-not-very-good English-to-Georgian system to translate a lot of English data, machine translated, into Georgian. And then you use that, even though it’s noisy, to train and see what more you can learn to reinforce what you know is good. Once you’ve reinforced a little bit, translate all of the English again with a slightly better system, and go again and keep iterating like that.

 And so, that’s kind of unsupervised learning or supervised learning from a very low starting point, and this is something that a lot of people are working on and has the potential. We’re doing it in the company now as well, and it’s promising, and it’s another one of those things where my eyes are opening here. If a client said to us before, “We don’t really have any data to give you,” again, that was another factor where we might say, “Okay, this is a hard language *and* you can’t give us any data? Probably can’t do anything.” Now, it’s a case of, “Okay, we can probably still try something here.” And it’s probably going to be more effective than anything you’ve seen before for this language pair.

Michael Previously, we may have said, “We can’t do it at all,” or “You need to go and create this data.”

John To give you an example that reflects what I said as well at the beginning about the kind of the pace of change: towards the middle of 2018, we decided to do a blog series on our website, because we have a lot of scientists—PhDs in machine translation in our team—so we said let’s do like a six-part series on what’s in-vogue in machine translation.

Michael Yeah, and in the six parts we’ll cover it pretty well, sufficiently.

John Data cleaning, low-resource languages, domain adaptation and, you know, that’ll be a good primer. But, like, by the time we got to the end of the six weeks, there was something new.

 So, we just published issue 26 this week and we’re already at the point in a seven-/eight-month period where we’re revising topics that we covered a few months ago because there’s already an update on them. And that applies to this kind of notion of training when you have no data or low resources for a particular language. I think we’ve done three issues on that now because it’s…the updates keep happening. Somebody tries something new and it’s really promising, and you compare it to the previous approach and it’s approved upon it. It’s really cutting edge.

Michael Well, it sounds like a great spot for our listeners to be able to stay aware and read up on the latest that’s happening. Where do they find it?

John So, it’s on our blog. It’s called “The Neural MT Weekly,” so I think if you go to iconictranslation.com/blog, it’s quite prominent there. If you check any of our social media channels, it’s there. So, we have our team, like I said; they kind of write issues on a weekly basis. We have some guest posts as well; we have some cool guest posts coming up.

 So, basically what they do every week is they’ll take a topic, so let’s say unsupervised training. They’ll look at a range of different people and different universities or companies who’ve kind of been working on this topic, and they’ll write a summary post of ‘here’s what they did; here’s the impact of it; here’s what this might lead to,’ and then move on to the next one. Sometimes we have somebody who wrote those papers write a guest article.

Michael That’s cool. And from what I’ve read on it, it is accessible, so you start reading an article and you’re like, ‘yeah, yeah, I get this,’ and then suddenly you are in a pretty challenging academic piece…

John Yeah.

Michael So, it has both the levels.

John It’s unavoidable. And so, that’s kind of the guidelines we have for the guys when they’re writing it. It’s like, try and make the intro and outro at least kind of generally accessible, but obviously it’s hard to talk about detailed approaches to unsupervised learning for neural networks without giving some…

Michael You gotta get into the weeds.

John …level of detail, exactly. So, yeah, that’s the balance we’re trying to strike.

Michael Probably one of the key questions you’re getting asked from buyers of this service is: Is it still too early for me to get into it? Or is now a good time to start testing and playing with some of these engines and seeing what happens? What’s your advice there?

John It’s 100% the right time. We were cautious at the beginning; we weren’t kind of going straight gung-ho saying, “Right, we’re going to do neural everything from now on.” But that’s where we are now. So, we don’t build statistical MT engines. There are legacy ones that are still in production, and I think that’s the case for any sort of technology. If something’s not broken, don’t fix it, with some users, and so, everything we do is neural now, and even, like, the evolution of neural—recurrent neural networks, convolutional neural networks, now it’s attention-based—it’s still evolving. So, it is definitely the right time. We’re seeing new use cases. We’re working in kind of a lot of industries where, maybe (a) they were kind of slower to adopt this type of automation, this type of technology in the past due to concerns over quality, and (b) kind of use cases that maybe we couldn’t—weren’t practical before—and languages and things like that, that are now kind of a lot more open.

Michael Yeah. I’ve been hearing different things in the industry: life science companies being heavily committed to seeing their content being translated through neural machine translation.

John We’re experiencing exactly the same, because the pressures of multilingual content are on more industries. Tech, IT or the kind of automotive were the traditional ones, but as you know, pharmacovigilance, clinic trials, are getting more complex, and you’re having to reach further around the world to kind of get people on trials. The language issue is becoming a lot more challenging, and life sciences companies used to kind of pass that issue off to wherever the trial was running onsite, but that had cost issues, had speed issues, time-to-market issues, it had quality issues. And so now, they’re seeing neural MT at the center of a workflow as an opportunity to transform to some extent how those operations are working.

 And there was always, like, the concern in life sciences that there were so many quality checks—for some of the certifications, there’s even back-translation. If you have something, you know, professionally certified translated from, let’s say, Chinese into English, you’ll have somebody else professionally translate that from English back into Chinese as a, you know, an extra sanity check.

Michael Yeah.

John So, what the technology will allow you to do is just get to those points faster. You know those quality checks have to happen. You know how long they take, but can we get to starting them faster through automation of some of the translation process?

Michael And with the companies that are doing this, there’s still some level of editing involved with it. The human element is still there? Or are they going fully automated for a back-translation?

John It depends. It depends. So, that’s what we’re seeing. So yes, there is the traditional workflow of a human in the loop, and there’s revision, there’s editing, there’s QA happening at the end of a machine translation process. But, what we’re seeing—as I’ve mentioned a couple of times about new avenues opening up—is what can we do with raw machine translation or, you know, MT as-is? What that requires is getting to a point where you trust the quality enough that it’s not doing anything nefarious.

 So, to use the life sciences example, one area there is pharmacovigilance. So, that’s where, you know, pharmaceutical companies have to allow all consumers, users of their drugs—so be that an individual, be that a hospital, a doctor—to report any side effect they have with the drug. And they’re coming in in all sorts of formats and all languages, and companies have to react to them within a certain short period of time to meet FDA requirements, regulatory requirements. That just creates a big backlog.

Michael Depending on the volume that’s coming in, there’s no way a human person could assess—

John There’s no way, no, and so basically, you say, ‘okay, these go into the backlog; we’ll take the fines that we get because we can’t resolve them on time,’ but that’s not sustainable because it’s not going to go away. So, that’s something where we see a machine-only part—at least for the first part of getting it translated, getting things categorized before they’re acted on.

Michael They can sort of flag those translations that may be most likely something serious, and then go back and check them with a human resource.

John Yeah, and to give you another example, of another industry that has a legal bent on it, because it is the legal industry: litigation e-discovery is a big case where the opposition council will dump terabytes of digital data and say, “right, go and find the smoking gun in there.”

Michael Yeah, like, two days before they’re going to trial or something, yeah.

John Exactly, and so, you know, a lot of the law firms are used to doing this in English, so they’ll have processes and workflows for doing searching and complex analytics, but only on English. And so, they need to get something into English quickly so they can act on it. So, again, there machine translation is the only option. The reason it doesn’t have to be perfect is because you’re establishing a legal position on the basis of something.

 And so, the fact that you’re using MT is fully transparent. So, you can go into the court and say to the judge, “We are taking this position on the basis of a machine translation of this document into English,” and that’s legally defensible. Increasingly so, courts are accepting the fact that okay, that’s a fair enough reason for you to take that position. And so, that gives you time, that buys you time. That allows you then to have, you know—so, you filtered out a lot of the documents because they weren’t relevant and you’ve established your position in time using the MT, and then you can go back to the original workflow of having people put eyes on it and maybe having certified translations where you need them.

 The other big use case that we’re seeing for the machine translation only is in customer support: so, time-to-resolution for an issue, or you want to get as many first-level support issues resolved without people if you can. So, that’s where knowledge databases on the website come from. You have them translated into multiple languages, but then, even once the human is involved in the loop, if you’re selling into, you know, 70 different countries, and you have to maintain a multilingual support team, that can be onerous.

 So, we’ve a couple of large IT clients who use MT in the middle of their ticketing system. They have an English-speaking customer support team who receives all tickets from customers in English because they’ve gone through MT. And they respond in English and the customer sees it in their language. They’re solving an additional 30-40% of their issues with that before you then have to engage a support who speaks the language of the customer as well. And there’s a massive cost implication there.

Michael Yeah, I was thinking the same thing. Every phone call where you’re talking to a real human being has a dollar figure, and when you’re able to reduce that, that’s significant savings.

John Yeah, it’s massive, and you’ll see a lot of the technology providers in the localization space are looking at that as a use case, because everyone has customers, everyone needs to support them, everyone wants to have more customers in more places, so it’s a problem that almost every company has.

Michael Yeah, that’s fabulous. You gave us some really good use cases and a good update. Anything else our listeners should be aware of that’s happening right now in this space?

John I don’t want to say no because, you know, because there’s so much happening! If you think from a commercial provider, like our perspective, our interest is less on the pure research and more applied. Our guys are keeping a, you know, close eye on what’s happening and where and considering how we can apply that to our applications.

 So, one of the big areas is around what we call domain adaptation. So, if you think of the likes of the online translators like Google and Microsoft—they’re really good. So, you know, sometimes when you hear people say, “Oh man, I used Google Translate, I know it was rubbish.” And we say it’s not rubbish; it’s, like, really, really high-quality technology going on there. What their comments are maybe reflecting is that in trying to do a little bit of everything, you have to give something up, and that’s, you know, being very, like, extremely good at one thing.

 And so, that’s what we try to focus on, and a lot of research is trending towards how can we maybe have multiple different data types within our engine? So, how can we have, like, a single engine that has training data from the legal area, from the pharmaceutical area, from IT, and do this kind of on-the-fly adaptation depending on the input that comes in?

 So, rather than having to have a single system for each different customer, for each different domain that you work in, how can we do this kind of on-the-fly adaptation? And that’s the intersection of research and kind of commercial deployment of MT.

 How can we build upon what’s already there? Anyone who’s providing machine translation, whether it be a provider like ourselves or an LSP, they’re up against what people can access freely online, and so it’s looking at how can we get an advantage over that? How can we adapt technology that we have more to an industry, more to a particular client?

 We’re tweaking things a bit more, making things, having kind of technology that at its foundation is more general, but then ultimately having that be something that’s bespoke for a particular client and their use case.

Michael It’s a way you describe your business: the bespoke.

John That’s intentional on our side. The word that that we used to use that a lot of other providers use was ‘custom.’ But now everything is custom. So, you can customize Google Translate, you can customize Bing Translator, but I guess there are levels of customization. So, dropping in a few dictionary terms is customization. So, how do you distinguish between dropping in a few terms versus taking your core technology and really, really adapting it for Client A’s very specific use case in the pharmacovigilance division of their life sciences company?

 And so, that’s what we describe as bespoke. So, the constants are our technology that we have and the methodology that we have for adopting it and the team that we have, and the variable that comes in there is the client and their use case and their data. And so, combining them is what we say produces a more bespoke solution for them. Though, with a, like I said, a consistent foundation to allow that to kind of scale better as well.

Michael Yeah. Well, that’s great. You’ve given our listeners a lot to think about and a really good perspective. I can’t wait until the next time we catch up.

John Yeah, definitely, let’s do it in 12 months’ time and see where we are then.

Michael Sounds great!

Renato Well, as you guys just heard, John Tinsley is very knowledgeable about this space, and shows us the real status of neural machine translation. How does it apply to you? How does this change your world? Think about it and let us know what you would like us to talk about the next time we approach this subject.