LTJ 27 4

[Start of recorded material]

Interviewer: From the University of Leicester in the United Kingdom. This is Glenn Fulcher with another issue of Language Testing Bytes.

In Issue 27(4) of *Language Testing* contains an article by Carol Chapelle and colleagues, entitled, *Towards a Computer Delivered Test of Productive Grammatical Ability*. This is one of many paper's that's being published at the moment on the assessment of grammar, and we thought that this would be an excellent opportunity to look at what's going on in the testing of grammar, and to try and get an overview of the field. We turned to Jim Purpura, who is Associate Professor of Applied Linguistics at Teachers' College, Columbia University in New York, and also the author of *Assessing Grammar*, published by Cambridge University Press in 2004.

I telephoned him in his office in New York and we apologise for any of the noise in the background, New York can, of course, be a very noisy, if exciting, place.

Jim, welcome to Language Testing Bytes and thanks for agreeing to talk to us about assessing grammar.

Respondent: Oh, I'm very happy to do it, this is a great thing you guys are doing, Glen.

Interviewer: Thanks Jim. Teaching and testing grammar have been out of favour for quite some time but over the past few years I've noticed that it's starting to come back into fashion, and this reminded me of something Alan Davis wrote in an article in *Language Teaching*, it was, in the early 1980s, and I just want to read you one of the quotations from that, and what Alan said was, "What remains a convincing argument in favour of linguistic competence tests, both discrete and integrative, is that grammar is at the core of language learning. Grammar is far more powerful in terms of generalisability than any other language feature. Therefore, grammar may still be the most salient feature to teach and to test. Jim, do you think Alan Davis was, and still is, right?

Respondent: Well, yes, I do think Alan is, was and still is, right actually. I mean, as I've said in many of my talks, you know, grammar is really the critical linguistic resource for communication, and so, without that piece, you know, we don't have language really except for sign language. So, you know, I do believe that that's true. Also, more recently, Glen, there has been a number of studies that really provide convincing evidence that the relationship between grammatical knowledge and performance is a very strong predictor, grammatical

knowledge is a very strong predictor. I mean, in the old days, McNamara 1990 and his colleagues did some research on looking at the relationship between grammatical resources in speaking and that was a strong relationship. More recently there's been some very sophisticated articles on grammatical knowledge and reading, and [UI 03:13] dissertation, or Eolic Leo's [SL 03:16] dissertation on grammatical knowledge and reading and listening. Kirby Rebowski's [SL 03:21] on pragmatic knowledge, and Chewa Dakin's [SL 03:24] dissertation on the relationship between grammatical knowledge and social studies knowledge, and, you know, it goes on and on.

So, for those who reject the notion that grammar's important or shouldn't be assessed in some way on tests, I really don't understand where they're coming from because, for me, this is the heart and soul of language ability, so, yes, I agree with Alan.

Interviewer: And, your own approach to assessing grammar as its evolved is, kind of, very different from what we think of as the traditional multiple choice discrete type structure items, and in Issue 27(4) of the Journal, the current issue of *Language Testing*, Chapelle, Carol Chapelle and her colleagues begin their paper with a quotation from your widely acclaimed 2004 book entitled, *Assessing Grammar*, and that's taken from page 37, in which you particularly criticise, and I quote here, "Scoring rubrics with descriptions of grammatical development that have little support from SLA findings."

Well, it seems to me that second language acquisition research generally offers language testers very little that we can actually use in designing language tests, with the one exception of the acquisition of grammar, you know, from ground breaking studies like those of Pienam [SL 04:52] back in the 1980s, and even before that with the Moore theme [SL 04:56] acquisition research. How do you think this evidence can be used in the assessment of grammar?

Respondent: I do think it can be used and I think there's information on developmental sequences as, and also in the morphing studies that can be used for general understandings of how people acquire certain grammatical features. And people have used these in tests, I mean, Chang's study looked at relative clauses and how we could use these features, you know, the information from SLA and how we can design tests that relate to that. And then, you know, so we can use some morphing studies, we can use developmental sequences in terms of sequencing certain items.

But, the problem with SLA studies is they're… first of all, there aren't very many features that they've looked at in that, in the grand scheme, they haven't looked at very many features. Now, in Carol's articles with her colleagues she put in a number of extra features that I haven't seen in studies before, but the problem with all of these

studies is that, first of all, they're not drawing on a comprehensive model of grammatical knowledge. So, as a result, there are issues of representativeness and coverage, so my question, in looking at that study and some of the other studies that use SLA research is, how do we know that these features are the, we have the best features that we should be testing.

And so, on the one hand, and I'd like to say we need to look at SLA research and try to understand, you know, how they've looked at these sequences in the orders and what they've found, but we have to understand that this is not the be all and end all there's a whole lot more research that needs to be done, basic research that needs to be done, in order for us to be able to use this research to assess grammatical knowledge in a much more comprehensive way.

And so, what I was trying to accomplish in my book was to provide at least some kind of a beginning of a comprehensive framework, where that would allow us to know that we're not just focusing, for example, on morphus and tactic [SL 07:35] forms. But, there are a number of other types of forms that need to be accounted for, and, just as importantly, we can't overlook the research that, now there's been almost 20 years of research on forming mappings, and for us to go back, think that it's only about syntax and is just, you know, wrong minded as far as I'm concerned, and a lot of the SLA research is really focused on morphus syntax.

So, even though we should consider what has been done in SLA I think we need to go way beyond that and think more about how some of these forms relate to their meanings, and meanings that even might go beyond literal meaning or semantic meaning, but meanings that are used pragmatically. So, yes, there's more and more research in SLA and I think, I just don't think we can completely ignore that, but we need to use it judiciously.

Interviewer: When it comes to actual items that we use to assess grammar, however we define that, it's clear that you prefer not to rely entirely upon multiple choice items anymore, and Chapelle and colleagues in this month's issue recommend items like, gap filling, jumbled word order, and form changing items. When it comes to the variety of item types that we can use in our tests to assess grammar, are we currently going anywhere innovative? I mean, do you see anything new on the horizon?

Respondent: Well, you know, first of all, I didn't say in my book that I preferred any kind of items, I don't think, I'm sure I didn't. What I was trying to say in the book was that I don't think we should have [UI 09:39] item types. In other words, depending on the context of assessment and what exactly we're trying to get at we should freely be able to use any kind of, you know, selective response, [UI 09:54]

production, or [UI 09:55] production tasks, in order to elicit grammatical knowledge.

So, I don't think that I would say that we should not do multiple choice, that in certain circumstances, that is, you know, the way… that's how we might want to elicit that knowledge. But, again, you know, Keogh used a bunch of limited production task in this, and I don't think in the end, Glen, to answer your question, that there's all that much that's innovative in terms of items, item types, that I've seen so far.

I think that we could be more innovative in that respect if we were going to use the full capacity of computers, but, in what I've seen so far and including in the Oxford tests that I've worked on, as you know, we have been using just these very traditional item types and we've just been putting them on the computer, but what I think we could do is look at issues of visual media, and context, and try to elicit grammatical knowledge from students in a much more interesting way by means of providing them with rich contexts delivered over the computer.

So, I think that's a way to go in terms of item types, but I don't think what's innovative so far is that. I think what is innovative, though, is how we're thinking about scoring the actual performance that we're seeing, and Carol has used a rating scale model in terms of doing that, where she has a zero to two scale, but there are some other ways of looking at that also. Carol and her colleagues basically looked at morphology and syntax, and they, kind of, played around with meaning a little bit, but meaning really is an explicit so, basically, the way they scored this was really on morphus syntax, but I think there's a different dimension here that, especially with the limited production tasks, that could be scored, and that is the meaning that's being conveyed and the extent to which meaning is conveyed in those items.

So, in terms of scoring we can look at scoring in some very different ways. We could look at the way Carol did in terms of rating scales or in terms of what Pineman [SL 12:48] did or suggest, and Rod Ellis suggested in terms of developmental scores, where you take information from SLA and then give scores based on those stages, or, I think another way of doing this is what Boffin and Palmer [SL 13:03] suggested was to look at right/wrong scoring but using multiple criteria for correctness when the construct that's been elicited suggest there are different dimensions.

And then, in Carol's data, for me, the construct being elicited is multiple dimensional, there's a form dimension and there's a meaning dimension, and unfortunately in the study, I mean I don't want to criticise the study too much because I really liked it, but I think there was an opportunity lost in the study in that they only got

very ambiguous information from the scores, where they could have got, you know, information on the students' knowledge of forms, and also their ability to convey those meanings associated with those forms.

So, I think the scoring is the more innovative aspect of what's happening in grammatical assessment.

Interviewer: Well, from what you've just said, and indeed form your recent work, what you're doing is bringing the focus to the meanings that are associated with grammatical choice, and you brought to testing, I think, in your work, some of the really critical insights from applied linguistics, going back to the work of grammarians, Halliday and Sinclair spring to mind. But, of course, applying it, as you indicate, is where the devil creeps in.

If we focus on meaning, someone might say that we're assessing pragmatic competence rather than grammatical competence, or an ability, say, to detect illocutionary intent or something like that, but we know that grammatical choice is bound up with being able to communicate with appropriate illocutionary intent in pragmatically suitable ways.

I guess, as we're coming to the end of the podcast now, I suppose what I'd really like to ask you, what I'd like to know, is how you see language testers coping with this kind of complexity and still being able to say that they're assessing grammar?

Respondent: That's a really good question but I just want to begin by saying that we're always measuring grammatical semantics and pragmatic competence to some extent, and whether we do this explicitly or implicitly, from the moment that we begin using language we are assessing forms, meanings, and extensions of those meanings, as long as we have some context that allows us to do that.

And, that's one piece of the answer, so, another piece of the answer is that when, you know, there's a lot of evidence now that shows that second language learners learn form and meaning differentially. In most cases they acquired meaning before they acquire form, and so, since we acquire them differently, this tells me that there seems to be two things happening here, and that we should take account of this.

So, and this is normal I think because, you know, the goal of communication in my view is to convey meaning, so with meanings we want to convey are pretty simple then we can use simplified forms as resources of communication. For example, if you say, you know, I no wanting potatoes, and I say that in Greek, and I'm able to say that in Greek, where I have almost no knowledge of Greek. I would be really, really happy by saying, I no wanting potatoes, in a restaurant, and they would understand me and the way feel is that I

should get some credit for being able to say, communicate my meaning even though the forms I am using are obviously not accurate. So, why not give credit for the ability to convey semantic meaning rather than to only convey semantic meaning accurately. Of course, this depends on the purpose of test and the context of assessment and all that, but I'm arguing here that we should consider this.

If we want to convey those subtle meanings, or meanings that convey basically how academic we are, then obviously we need to draw on our grammatical resources that are probably more precise and possibly more complex and exhibit, maybe, a wider range of structures that we're able to use. In this case we need to recognise, you know, that there's complexity to it and we need to assess that in a different way, but from the moment that we use grammar, whether it's forms on a [UI 17:53] morphus syntactic level, we almost always map these forms onto meanings, and I say almost always because there are only a few exceptions in the language where I think the form doesn't have a meaning associated with it. For example, when we say, what time is it? The, it, that expletive, it, I really can't figure out what kind of meaning that has except for a syntactic need in our particular language group.

But, 99 times out of a 100 we're going to have forms that are associated with these meanings, so even though we think we're not measuring meaning we actually are measuring meaning, and we learn these meanings differentially. So, I would argue we need to begin looking at this and consider it from day one. I mean, you can… if someone asks you, for example in a restaurant, what can I get you, and you say, I'd like a chicken salad, for example, well, you know, that the meaning associated with, I'd like a chicken salad, is very clear, very literal, there's no inference that needs to be made and so you're getting here both form and meaning on these two different dimensions. And you can have all different types of varieties of, I'd like chicken salad, I want chicken salad, I'm liking chicken salad.

By the same token if you say, would you like tea or coffee, and you respond, well, coffee would be great, thanks, then in that case basically the relationship between that question and response isn't very direct. It's direct enough that form this context we understand what we're saying but the response to, would you like tea or coffee, is, yes, I'd like coffee, and not, coffee would be great. Coffee would be great, adds a different dimension into the meaning here.

By the same token if you say, for example at a servicing counter, you say to someone, can I talk to you, and they say to you, I have a meeting now. What you're really getting at here is you're getting at forms, all these forms here the asking of a question, but the response, I have a meeting now, is really not a direct answer to that question

can I talk to you, the answer would be, yes. Well, here you have context that the meaning is, no, you can't talk to me right now, I have a meeting, I have something else to do, so there's an extension of the meaning of I have a context where we are getting at pragmatic knowledge here.

And this is on a very, very low level, and you'll see these in all different types of tests whereas, what people think they're measuring… they think they're measuring grammatical knowledge but they're… they are to some extent, they are of course measuring grammatical knowledge, but they're also measuring pragmatic meanings here. Or, better yet, if you say, can I talk to you, and someone says, whatever, then, you know, think of all the types of meanings that are being elicited by just that one word, and, again, you wouldn't necessarily see… you might see this on a test but we do this at the very basic levels of language learning to the very advanced levels of language learning.

So, I would argue here that, you know, we have to include the meaning dimension with all kinds of assessment of grammatical ability. It's just, it's part of that construct and that was what I was trying to do in that model, you know, I would rethink that model a little bit now but still basically I still believe in that idea, and a lot of research now in SLA is showing that forming dimensions is really important.

I know that was a long-winded answer to your question, I'm sorry.

Interviewer: No. Jim, thanks for giving us this insight into what is a fascinating area of language testing assessment, and I think we're certainly going to see a lot more published in this area in the coming years, so it's really great to have your take on it, and we thank you for putting it all into context for us, and thank you for joining us on Language Testing Bytes.

Respondent: Oh, I'm very happy, thank you for having me Glen.

Interviewer: Thank you for listening to this issue of Language Testing Bytes.

Language Testing Bytes is a production of the *Journal of Language Testing* from Sage Publications.

You can subscribe to Language Testing Bytes through iTunes, or you can download future issues from ltj.sagepub.com, or from languagetesting.info.

So, until next time we hope you enjoy the current issue of *Language Testing*.Transcript Divas

[End of recorded material]
NOTES: [SL 00:00] Sounds Like
[UI 00:00] Unintelligible