Since the seminal publication of Norris and Ortega (2000), meta-analysis has emerged as an important tool for synthesizing the research in individual domains of applied linguistics including second language acquisition (SLA). In recent years, meta-analyses have mushroomed (see e.g. Goldschneider and DeKeyser 2001; Mackey and Goo 2007; Li 2010; Spada and Tomita 2010; Jackson and Suethanapornkul 2013; Shintani and Ellis 2013; Kang and Han 2015), fueling the conception that this type of research synthesis is more accurate and reliable than the traditional narrative review. At the same time, though, issues and concerns have arisen about the validity and reliability of meta-analyses.

This special issue is the first attempt to examine the complementarity of the meta-analytic and the traditional narrative review approaches and as such is very timely. Guided by a coherent framework provided by the editors, authors reported syntheses of the research on three topics—processing instruction (PI), the role of aptitude in (instructed) second language learning, and the teaching of pronunciation—as well as explicitly addressing how they went about their syntheses. Consequently, this special issue makes both a substantive and a methodological contribution to the field of applied linguistics in general and to SLA in particular.

In his introduction to this special issue, Rod Ellis (2015) provides an account of the scope and procedure of both approaches as well as of how each approach deals with problem formation, literature search, data evaluation, data analysis, interpretation of results, and public presentation—the core aspects of any type of synthesis. The introduction invited readers to come to their own understandings of how the two approaches fare relative to each other. In the limited space granted to this Epilogue, I have decided to focus on the pair of syntheses on PI, DeKeyser and Prieto Botana’s narrative review and Shintani’s meta-analysis. I will conclude with some general thoughts about the two approaches and on how applied linguistics can benefit from their complementarity.

CASE IN POINT

DeKeyser and Prieto Botana’s narrative review focused on three questions relating to (i) the role of explicit instruction (EI) in PI, (ii) the difference between PI and production-based instruction (PB), and (iii) factors that mitigate
the impact of PI. The specificity of these questions is advantageous given the range of primary studies sampled, which differed in both design and choice of outcome measures. Using the same sample of studies as those in the meta-analysis, the researchers coded the primary studies for the learners’ L1, the L2 target structures, the number of participants, the age of participants, the type of treatment, the outcome measures, and the main findings. They also included some additional studies to provide contextual information.

DeKeyser and Prieto Botana began with a summary of two critical studies, VanPatten and Cadierno (1993) and VanPatten and Oikkeno (1996), to create a ‘problem space’. The central question—what is the role of EI in PI—is thus thrust to the fore, and led me as a reader to formulate my own questions such as: Are the studies’ outcome measures sensitive enough to capture the contributions of EI in PI?

DeKeyser and Prieto Botana concluded that EI plays an important role in PI and that its effect is mediated by a number of factors such as its intensity, the complexity of the target structure, task essentialness, the learners’ aptitude, and whether strategy training is included in the instruction. Their review confirmed several points already known about the role of EI. For example, it showed that EI can provoke rapid changes in learners’ behavior, and the more intensive EI is, the more powerful it is in bringing about change in both receptive and productive use of the target feature. The conclusions they reached align with DeKeyser’s long-held view of L2 acquisition as skill acquisition (see e.g. Anderson 1983).

I found this narrative review highly informative not least because it led me to question some of its conclusions. For example, I began to question DeKeyser and Prieto Botana’s conception of what makes a structure complex. By my reckoning, the Spanish subjunctive—in the absence of any empirical evidence to establish its acquisitional complexity—may not necessarily be more complex than, say, the Spanish OVS word order or Italian gender agreement. Indeed, what counts as a complex structure is a vexed issue in SLA research (see e.g. Han and Lew 2012). Whether or not it is easy to determine the form-meaning relation of a given structure—a criterion DeKeyser and Prieto Botana seemed to have used to judge complexity—may not stem from the structure itself but from whether there are sufficient contextual cues that would allow meaning to be inferred. I found myself questioning whether there was a basis for claiming that complex structures are more sensitive to the influence of EI in PI.

A good narrative review also allows the reader to ask if there are any issues of significance not addressed. One such issue came to my attention. The majority of the PI studies have failed to identify the specific processing strategy to be addressed by PI. This is the bedrock of PI, as I understand it (see also Wong 2004). The aim of PI is to alter learners’ processing strategies in order to engender better intake (i.e. form-meaning mapping, not just the noticing of the form). If a PI study has not taken on board the processing strategy that needs to be addressed through instruction, the whole question of the importance of EI becomes redundant.
The point that I want to make about this and other narrative reviews (e.g. Skehan 2015; Thomson and Derwing 2015) is that this kind of review is inevitably idiosyncratic (cf. Norris and Ortega 2006). As such, for all its merits—depth of analysis, latitude in sampling primary research, specificity of argumentation, promise of insights, and the like—we must recognize that each review comes from an individual who brings to it his or her own conceptual baggage or particular epistemological stance, which mitigates against objectivity. One way of attenuating this would be through multiple independent narrative reviews of the same topic.

Turning now to the counterpart of the narrative review on PI, Shintani’s meta-analysis, the sample of primary studies here is the same as that of the narrative review, 33 studies, yet the attention given to individual studies received is markedly different. In the meta-analysis, the 33 studies, broken down into 42 experiments, which then received more or less equal treatment, unlike in the narrative review where 12 studies underwent detailed analysis with the rest mentioned only in passing.

The equal treatment of the primary studies—in keeping with the general purport of the meta-analysis—led to a number of general conclusions:

Although PI was more effective than PB [production-based instruction] for developing receptive knowledge, PB was just as effective as PI for productive knowledge. Furthermore, the PB proved superior to the PI for productive knowledge when both groups received the same explicit information. (Shintani 2015: 306)

The analysis of the moderator variables yielded additional findings (e.g. the effectiveness of PI was not influenced by the provision of explicit explanation and strategy training, whereas the effectiveness of PB for receptive knowledge was).

It is clear that the general conclusions are, in the main, orthogonal, both in spirit and in essence, to those of the narrative review. Similarly, the findings from the analysis of the moderator variables are largely at odds with the insights provided by the narrative review. By way of illustration, here is what the narrative review had to say about PI vis-à-vis receptive knowledge and PI and PB vis-à-vis productive knowledge:

While it has often been argued that PI is better than PB for comprehension ability and that there is no difference for production ability, our review shows that the PI advantage for comprehension is found in only half of the relevant studies, and that PB can be more effective than PI for productive outcome measures, depending mainly on how the output practice is implemented. The lack of advantage of PB documented in several studies may be the result of the non-communicative, drill-like nature of the practice they provided. (DeKeyser and Prieto Botana 2015: 301)

In order to understand these discrepancies, one needs to look into how the meta-analysis generated its results. First of all, the analyst adopted a mixed set of content-based and technical criteria in sampling the studies. Then, the
studies were coded for independent and moderator variables, mostly in dichotomous terms (e.g. PI vs. PB; receptive vs. productive), thereby reifying what is otherwise a very heterogeneous set of studies. Next, effect sizes were calculated and aggregated for the independent variables and moderator variables and their significance determined via Q tests.

These are standard procedures in performing a meta-analysis (see also Lee et al. 2015; Li 2015). They enabled the primary studies to become comparable, thereby avoiding the ‘apples and oranges’ problem. However, as a reader I still have a number of questions:

1. Given that conceptions of PI have evolved over the years, especially since the debate in Language Learning on the role of EI and production (DeKeyser et al. 2002; VanPatten 2002a, b), does it not matter if the primary studies were driven by different conceptions of PI?
2. If the sample of primary studies is non-homogeneous, as acknowledged by both reviews, what merit is there in a coding based on largely bottom-up, data-driven criteria along with arbitrary decisions (e.g. a post test is ‘immediate’ if taken within six days of the treatment but ‘delayed’ if taken seven days after the treatment)?
3. Do the uneven sample sizes for the moderator variables (e.g. 19 vs. 4 for participants’ age) not matter?
4. To what extent are the findings of the meta-analysis artifacts of how the data were analyzed [e.g. separating out studies directed at the First Noun Principle from those directed at the Meaning Primacy Principle, or lumping TI (traditional instruction), MOI (meaning-oriented instruction) and PB (production-based instruction)]?

Perhaps my gravest concern about the meta-analysis is that there is no critical examination of input processing theory. Questions have been asked elsewhere (e.g. Carroll 2013; Han and Rast 2014) as to whether the principles apply to all target languages and whether or not they apply equally to learners at all developmental stages. While such questions need not deter the meta-analyst, they are surely worthy of addressing if the findings of the meta-analysis are to make a useful contribution to our understanding of input processing principles and their pedagogical correlate, PI.

Summing up, the two syntheses of PI studies manifest systematic differences in their purposes, procedures, interpretations, and conclusions. The findings and conclusions are mostly stated in probabilistic terms in both reviews but, in my view, more coherently and less categorically in the narrative review. The complementarity of the two reviews is evident not just in their overlapping findings but also in their differences. For example, both identified strategy training in EI as an important moderator variable; both recognized the implication of the skill-acquisition nature of PI and PB; and both pointed out the scarcity of research on the role of individual difference variables in tempering the effects of PI and PB. Overall, the meta-analysis paints a gross picture and the narrative review a more nuanced one. For example, both reviews identified
structured input (SI) as the key to the efficacy of PI, but only the narrative review pointed out that it was the ‘task-essentialness’ of SI that was fundamental to the efficacy. The meta-analysis allows us to see the scope of research to date and its main findings while the narrative review illustrates the importance of a more probing analysis. It is in these ways that the two approaches demonstrate their complementarity and thereby enrich our understanding.

CASE WRIT LARGE

The two ways of doing research synthesis correspond to the qualitative (the narrative review) and quantitative (the meta-analysis) approaches in research in general. The former arguably engages with the primary research directly and at a deeper level and the latter less directly (i.e. via statistical analyses) but potentially more broadly. Moreover, the narrative approach relies on the researcher’s ability to not just sample key articles and factually report what they show but, more importantly, on the ability to interpret what is not explicitly stated in the primary studies. In reading comprehension terms (e.g. Grabe and Stoller 2013), the researcher attempting a narrative review seeks not only to build a text-based model of comprehension but also a situation model of the primary research, drawing on his or her ability to delve beyond surface details and to provide insights that are thought-provoking and heuristically rich and that can guide future research.

The meta-analytic approach, on the other hand, follows a set of ‘rules’ and hence is arguably more rigorous and objective than the narrative approach. In addition to aggregating the results of the primary studies into a form that allows for reliable generalizations, this approach also leads to emergent findings (i.e. findings not found in the primary studies themselves).

Still, I cannot resist addressing what I consider to be some myths about meta-analysis. First is its presumed objective nature. What often has evaded general attention is its subjective nature. For example, data coding relies largely on the criteria that the meta-analyst has established. A second issue is the purported validity of data sampling. The process of sampling the primary research can result in the exclusion of a substantial number of studies (e.g. 33 out of 62 studies were included in Shintani’s meta-analysis). As such, there is an issue of representativeness to begin with. In particular, the sample size for the analysis of moderator variables in many meta-analyses in SLA is typically small. Third, the reliability of the findings hinges on the primary studies being conceptually and methodologically homogeneous, which is not usually the case in the primary studies assembled for meta-analysis in SLA.

Applied linguistics has much to benefit from both approaches to synthesizing research as Rod Ellis has argued in his introduction to this special issue. The findings that emerge from a meta-analysis provide the big picture while the detailed and nuanced readings of the individual studies uncover critical issues—‘evidence that quantitative research cannot access’ (Thomson and Derwing 2015: 326).
I can see three ways of achieving complementarity in research synthesis. One is the approach adopted in this special issue, namely, to provide both a narrative and meta-analytic review on the same topic and invite readers to consider both. The second is to combine a narrative and meta-analytic review in one article as attempted by Jackson and Suethanapornkul (2013).

The third way is to complement an existing narrative review with a meta-analytic review. For example, in SLA research there are abundant narrative reviews of critical period research but not one meta-analysis. Conducting a meta-analysis would help paint the big picture, which is often lost in the narrative reviews, especially in those that adopt a particular position.

Finally, we need to recognize that meta-analyses, by their nature, can only synthesize the results of experimental studies and that there are other kinds of research—for example, case studies of individual learners that shed light on developmental trajectories—that are of central importance to our understanding of L2 acquisition and that can only be synthesized in narrative form. I would not like to see the current fashion for meta-analysis lead to the exclusion of such reviews.

ACKNOWLEDGEMENT

I would like to thank the editors for the opportunity to contribute to the special issue and for their constructive feedback on an earlier draft.

REFERENCES


Lee, J., J. Jang, and L. Ponsky. 2015. The effectiveness of second language pronunciation


