CAF: defining, refining and differentiating constructs

Gabriele Pallotti
University of Modena and Reggio Emilia

Abstract

This article critically scrutinizes a number of issues involved in the definition and operationalization of CAF constructs. It argues for maintaining clearer distinctions between CAF on the one hand and notions such as linguistic development and communicative adequacy on the other. Adequacy, in particular, should be considered both as a separate performance dimension and as a way of interpreting CAF measures.

The notions of complexity, accuracy and fluency (CAF) have been employed in a number of studies on the acquisition and use of a second language, although they do not constitute a theory or a research program in themselves. They are dimensions for describing language performance, most frequently used as dependent variables to assess variation with respect to independent variables such as acquisitional level or task features. CAF measures can also be used to describe performance by native speakers or first language learners.

Each of these constructs has been operationalized in various ways and the articles in this issue bear witness to this lively methodological debate. Researchers agree on the usefulness and validity of the constructs, but they also agree that they should be further specified and that they don't exhaust performance description.

Defining constructs

Before introducing the discussion on constructs and their properties, I will briefly address two broader issues. The first might be called the necessary variation fallacy. Researchers frequently seek to identify the 'best' measures of language performance, and the answer is typically that the best measures are those that most clearly show variance among subjects, both over time and across tasks, correlating with other equally varying proficiency measures (e.g. Wolfe-Quintero, Inagaki and Kim 1998). However, a measure can be scientifically valid and informative even if it doesn't show any difference among groups of subjects. Research should be concerned with variations and differences, but also with constants and similarities. If after an experimental treatment two groups of subjects don't show any difference, then this is not a non-result, but a result just as interesting as their being different. Likewise, if a measure does not change over time, this does not make it a poor measure, but perhaps a measure pointing to a trait that does not actually vary. For a measure to be valid, it does not necessarily have to vary across subjects, but it must adequately represent its underlying construct. If the trait related to that construct does not change, then so won't the measure. This search for variation is manifest in the present issue as well. Norris and Ortega in their contribution seek to identify variables with greater 'predictive value' or that are 'best indicators of differences' among students at various proficiency levels; Skehan writes that 'the value of D was not particularly revealing regarding task differences ... Accordingly, we will focus here on the findings related to Lambda'. There is nothing wrong in this approach, but it seems to me an orientation more appropriate for language testing - with its

practical need to discriminate between different groups of learners - than for basic research in SLA. In the case discussed by Skehan, if different tasks produce variations for Lambda but not for D, this may mean that they do not produce variations in text-internal lexical variety, while they require different proportions of difficult words - *two* interesting results, rather than one interesting and one 'not particularly revealing' result.

A second cautionary note related to the search for the 'significant result' concerns the use of multiple measures in comparisons of groups of subjects. The fact that two groups 'significantly' differ on just one measure out of many might indicate that we are actually dealing with two samples coming from two populations that are completely equal except for one characteristic. However, it is also possible, and indeed quite probable if p is close to .05, that the significance of that difference is due to chance - a likelihood increasing with the number of measures (with 25 measures, the chance of one of them reaching $p < .05$ is 72%; Woods, Fletcher and Hughes 1996: 128). One should thus be cautious in interpreting results on CAF variations when only one or two measures out of many reach significance.

Turning now to issues of construct definition, Norris and Ortega (this issue) note that 'researchers have not done sufficient thinking about what we are measuring or why'. In their paper they provide an exemplary treatment of syntactic complexity, carefully pinpointing theoretical constructs and their operationalizations. In the remainder of this section a similar attempt will be made at clarifying some constructs underlying CAF measures, in order to arrive at more internally consistent formulations.

Fluency

Fluency can be defined as 'the capacity to produce speech at normal rate and without interruption" (Skehan, this issue) or as 'the production of language in real time without undue pausing or hesitation' (Ellis and Barkhuizen 2005: 139). Both definitions imply some normative reference, usually assumed to correspond to native speakers' behavior. Fluency is a multidimensional construct, in which sub-dimensions can be recognized, such as breakdown fluency, repair fluency, speed fluency (Tavakoli and Skehan 2005). Once it is established which of these sub-dimensions is at issue, it is in principle relatively transparent what is being measured.

Accuracy

Accuracy is perhaps the simplest and most internally coherent construct, referring to the degree of conformity to certain norms. At least since Bley-Vroman's (1983) article on the 'comparative fallacy', it is clear that accuracy *per se* is not a direct indicator of interlanguage development. Wolfe-Quintero, Inagaki and Kim (1998: 33) acknowledge this when they write that 'the purpose of accuracy measures is precisely the comparison with target-like use. Whether that comparison reveals or obscures something about language development is another question'.

Despite the dubious validity of accuracy as a measure of language development - Norris and Ortega (2003: 737) suggest it may be appropriate mainly in the description of advanced levels - it is at least a well-defined and coherent notion. There is a risk, however, of making it a spurious construct covering distinct dimensions, and this happens when errors are classified according to their gravity. For instance, some authors (e.g. Homburg 1984) propose to calculate accuracy

scores by giving errors different weights according to the extent they compromise communication. However, a 100-word production with ten errors not compromising communication is not more 'accurate' than a text of the same length with ten errors hindering comprehension, but just more 'understandable' or 'communicatively effective'. One can have perfectly accurate but communicatively inadequate messages (*colorless green ideas...*) or perfectly intelligible messages violating various L2 norms (*me no likes go dance*), which proves that we are dealing with different constructs (accuracy vs. comprehensibility), not relative degrees of the same construct.

The same argument holds for the proposal to grade errors according to their level in a developmental sequence (e.g. Cumming and Mellow 1996). A 100-word text with ten errors on subjunctives and conditionals is not 'more accurate' than one with ten errors on articles and pronouns, but simply 'more developed' or 'advanced'. In this case, too, we may find texts that are very accurate but scarcely developed (i.e. advanced on a language evolution scale) and texts containing many errors but exhibiting several traits of evolution. 'Accuracy' and 'development' are thus separate constructs and should be assessed with different measures, not with measures merging the two dimensions.

Complexity

Complexity is certainly the most problematic construct of the CAF triad because of its polysemous nature. To begin with, the same term complexity is used to refer to properties of both tasks and language performance. It would be preferable, following Skehan and differently from Robinson, to call the former 'difficulty', specifying that it means *objective difficulty,* i.e. inherent to the task, and thus

different from *subjective difficulty,* which arises from the encounter of a subject's (in)competences with a task. Climbing Mount Everest is *objectively more difficult* (i.e. more difficult for everyone) than climbing the Ben Nevis, even though climbing both mountains is for me *subjectively more difficult* than for an experienced mountaineer.

Even restricting the use of complexity to performance description, it still retains multiple meanings, because it can be applied to different aspects of language and communication. Ellis and Barkhuizen (2005: 153-4) list eight of these aspects, including lexical, interactional, propositional and various types of grammatical complexity. One might wonder whether it is appropriate to consider all these aspects as dimensions of the same construct or as different constructs altogether. In this issue, for example, Skehan doubts whether lexical and syntactic complexity are 'different aspects of the same performance area' or two separate areas, while for Norris and Ortega even syntactic complexity is a multidimensional construct with several sub-constructs.

Secondly, for each of these constructs the word *complex* may have different meanings. At least three such meanings can be distinguished. One is purely structural, 'composed of two or more parts' (Merriam-Webster), which corresponds to the etymology - from Latin *com* (together) + *plectere* (to braid). Linguists and teachers have long sought structural criteria for distinguishing between simple and complex grammatical structures, for example 'the number of linguistic transformation rules' (Spada and Tomita 2007: 229) or 'the number of criteria to be applied in order to arrive at the correct form' (Hulstijn and De Graaff 1994: 103). Another widely used structural definition of complexity refers to variety, to the existence of multiple alternatives - this is the main sense of many definitions of

lexical complexity. The Merriam-Webster Dictionary indicates a second meaning of *complex*, 'hard to separate, analyze, or solve'. This definition is frequent in CAF studies which equate *complex* with *difficult*, *cognitively demanding*. For example, Skehan in this issue defines *complex* as 'challenging language'. Thirdly, some authors identify *complex* with 'acquired late'. In this issue, for example, Ellis (table 1) defines complexity as 'the capacity to use more advanced language', while Skehan speaks of 'more advanced language, leading to complexity'.

These three senses of complexity are certainly related in many circumstances, and in several cases one can say that a structurally complex sentence requires many cognitive resources for its production and that it is therefore acquired late. However, this correspondence has to be demonstrated, not assumed, and there is no guarantee that it holds universally. For instance, some linguistic forms may be acquired late because they are infrequent or not communicatively relevant and not because they are structurally complex or require special cognitive efforts (Hulstijn 1995). It is thus necessary to separate the notion of complexity from that of 'progress' or 'development'. *Advanced* and *developed* are time-related notions, meaning 'appearing late' and they are not included in dictionary definitions of complexity. Furthermore, if complex equals advanced, the growth of complexity over time is no longer an empirical finding, but a tautology. In order to meaningfully apply the notion of complexity to development, the second construct should not be contained in the first. The increase over time of structural complexity (e.g. subordination ratio or lexical variety) or cognitive complexity (e.g. the distance across which grammatical features are unified in a sentence) should be a falsifiable empirical proposition, not an essential part of the definition of complexity itself. For this reason, attempts to validate complexity indicators by showing that they increase over time appear to be

questionable.

The consequence of this discussion is that development should be considered as a separate dimension from CAF. CAF refer to the properties of language performance as a product, while development is a process, with its sub-dimensions such as route and rate. CAF measures can empirically be related to developmental dimensions, but the latter should not be analytically considered part of the former.

Tasks and CAF: general and specific measures

Numerous studies have investigated the relationship between task properties and linguistic performance measured along the CAF dimensions. In this issue, Ellis reviews the effects of planning time, while Skehan and Robinson et al. focus on the interplay of various aspects of task demandingness and language complexity. This interplay has been the object of a rich research program whose results are still rather mixed. Several studies found positive correlations between task demandingness and linguistic complexity, but others did not. Theoretical models have been proposed to account for these results and Skehan's article is a good example of such efforts. Drawing on psycholinguistic research on attention and on Levelt's (1989) speech production model, Skehan seeks to identify the conditions that promote linguistic complexity and those that may lower it. In some cases, the task may require learners to pay a high amount of attention to the content of their utterances (mainly involving the Conceptualizer in Levelt's model), which depletes the attention available for the Formulator that is responsible for constructing more stucturally complex utterances. However, there may be cases in which these higher conceptualization efforts do not interfere with the Formulator, thus allowing for more complex *and* more accurate

performance on difficult tasks. This account is akin to Robinson's idea of resource-dispersing and resource-directing task dimensions, with the former lowering performance levels and the latter increasing it. The main difference seems to be that Robinson's Cognition Hypothesis makes the prediction that complex tasks will lead to more complex and accurate performance in general, while Skehan maintains that 'each influence is a task design feature' rather than a generalized effect.

This leads both authors to look more closely at the relationship between specific task conditions and specific effects on performance. The claims thus shift from asserting that task difficulty (in general) affects linguistic complexity (in general) to the observation of how certain task features produce effects on the production of specific linguistic features. This approach makes good sense. Why should *all* sorts of task complexification lead to higher complexity of *any* linguistic feature? Why should telling a story in the past stimulate the production of more rare or varied lexicon than the production of the same story in the present tense? And why should making a decision with more elements produce a higher subordination ratio than making one with fewer elements?

It is more logical to expect that a task involving conterfactual reasoning stimulates the productions of subjunctives and conditionals more than a simple narration, or that a task requiring to talk about astronauts and their space adventures elicits more uncommon words than one describing children playing with a ball in the park. In this way, the relationship between task difficulty and linguistic complexity becomes more reliable, but less theoretically interesting. Discovering such relationships looks more like validating the tasks as elicitation procedures for specific linguistic features than like confirmations of general theories about speech production.

In this issue, Skehan in fact explains the finding that narratives elicit more uncommon words than decision making tasks by the different lexical demands of the two communicative situations, concluding that 'narratives seem to push second language speakers ... into using less frequent lexis, presumably as they are responding to the events within the narrative'. Robinson and colleagues further pursue this investigation on specific measures. Robinson (2007) found a strong association between performing tasks requiring reference to complex psychological states and a higher use of psychological terms, taken as an index of more complex language. The result was clear-cut, but rather predictable - if you need to describe complex intentional chains, you will need to use more psychological terms. Further studies discussed in this issue look for more subtle correlations, but their results are less conclusive. In their first study, Robinson et al. find a higher use of atypical aspect marking in the condition where subjects had to tell a story in the past, assumed to be more complex than telling the same story in the present.[1] The trend is in the expected direction, but the difference is significant only if figures for past and progressive are summed up. However, one is left to wonder why learners in the Here-and-Now condition used so many past tenses, nearly as many as those in the There-and-Then condition. Perhaps they were using them as unanalyzed chunks. If this is the case, then it is not surprising that these stereotyped forms appear with the most prototypical aspect. Furthermore, one might question the choice of treating Accomplishments as 'atypical' contexts for perfective and progressive markings, as they share features of both Achievements and Activities. In the second study no or weak correlations were found between the There-and-Then (complex) condition and higher use of complex and more appropriate lexicalization patterns. This too is predictable, as there is no clear communicative or cognitive reason for referring to

past events with more satellite-framed verbs or greater event conflation.

These results support the idea that ingenuous manipulation of specific task characteristics stimulates the production of certain linguistic features, but not that task complexity in general promotes linguistic complexity in general. Such a finding is pedagogically relevant in that it demonstrates that carefully designed tasks can make learners engage with structures deemed to be particularly complex, in a beneficial combination of meaning-oriented communication and focus on form. Theoretically, they cast doubts on any broad framework relating general task complexity with general linguistic complexity, further demonstrating that performance complexity is multi-faceted, like its relationships to communicative circumstances. The main conclusion seems to be that linguistic complexity grows when this is specifically required by the task and its goals, and not for the sake of it, as if learners aimed at complexification by default, as if they were eager to produce a subordinate clause or a rare lexical item whenever they have a pinch of free attentional resources.

Adequacy

Most studies have assessed CAF within communicative tasks, but very few of these discuss how the communication unfolded and whether it was successful in achieving its goals. If in an information gap task a learner were to utter unhesitatingly *colorless green ideas sleep furiously on the justification where phonemes like to plead vessels for diminishing our temperature,* her production would score extremely high on complexity, accuracy and fluency, in spite of being

completely irrelevant, and probably counterproductive, for task success. In contrast, an utterance such as *No put green thing near bottle. Put under table* is neither complex nor accurate, and may not be fluent either, but can turn out to be perfectly functional for achieving the speaker's (and the task's) intended communicative goal.

It is surprising how few CAF studies report data about the communicative success and adequacy of the tasks and the learners investigated (for exceptions see De Jong et al. 2007; Kuiken and Vedder in press). This is in striking contrast with how the two dimensions are treated in language teaching and testing, where an effort is made to independently assess notions such as communicative adequacy and success on the one hand and linguistic accuracy and complexity on the other.

Adequacy can be seen both as a separate dimension, theoretically independent from CAF (although it may be empirically found to be related to it), and as a way of interpreting CAF measures themselves. As a separate performance descriptor, adequacy represent the degree to which a learners' performance is more or less successful in achieving the task's goals efficiently. As the examples above show, a text scoring high on CAF measures may prove to be quite inadequate for reaching the task's goals, while another can be quite effective in this regard despite its limited C, A and F. As a descriptor of communicative success and efficiency, adequacy can be measured in several ways. In closed tasks with correct/incorrect outcomes it can be rated straightforwardly as the ratio of correct items achieved. In open tasks with no predefined correct answer, adequacy can be evaluated by means of qualitative ratings, using pre-defined descriptor scales like the ones of the Common European Framework of Reference (Vedder et al in press).

A second way in which adequacy should be considered in CAF research is in the interpretation of CAF measures themselves. Many studies seem to implicitly

assume that the more CAF the merrier and that less CAF depends on limitations in language processing capacities. However, both assumptions are questionable.

Even for accuracy it is not always the case that 'more is better'. Sanell (2007) for instance shows that Swedish learners of French as a foreign language initially produce 'wrong' negative sentences of the form *je vais pas*. Later on, the standard *je ne vais pas* becomes more frequent and almost categorical. However, advanced learners tend to use it less frequently, preferring *je vais pas*, which is not correct for standard French norms but extremely common in colloquial speech, and thus an index of sociolinguistic ease.

As regards fluency, it is well known that there is an optimal amount of syllables per minute, beyond which a faster speech rate negatively affects comprehension. Pauses and dysfluencies too can be communicatively functional, from both a psycholinguistic (Chafe 1994) and an interactional (Goodwin 1981) point of view.

But the dimension where assuming a linear increase is most problematic is that of complexity. First, because this is the dimension with the largest individual stylistic differences. It is indeed possible to report group means, but, as Skehan reminds us in this issue, such means may result from very different individual profiles and much variance is to be expected even in native speakers' baseline data. In short, Beckett is not Joyce, and this has nothing to do with (in)competence, but with stylistic choices. Secondly, even taking group averages as a baseline, optimal complexity levels may vary according to communicative tasks. Pallotti and Ferrari (2008) have compared native and advanced nonnative speakers in two different oral tasks, a film retelling and the opening of a service phone call. While the natives' syntactic complexity (measured as AS unit length and subordination ratio) is higher

in the narrative task, the exact opposite occurs in phone calls' opening. Here, nonnative speakers tend to produce longer and more complex syntactic units, which are grammatically correct but pragmatically unnatural and ineffective, leading to frequent misunderstandings and repair sequences. In a subsequent study, Pallotti and Ferrari (in preparation) have observed the longitudinal development of the same learners over four years, finding that they tend to approach native speakers' behavior with a gradual increase in syntactic complexity in the narrative task and a decrease in phone calls' opening, producing in the latter case an omega pattern. These data empirically confirm Ortega's (2003: 494) intuition: '"more complex" does not necessarily mean "better". Progress in a learner's language ability for use may include syntactic complexification, but it also entails the development of discourse and sociolinguistic repertoires that the language user can adapt appropriately to particular communication demands'.

Especially for fluency and complexity, native speakers' baseline data are crucial, not because learners' aim is necessarily to behave like native speakers, but because looking at what native speakers do may overcome the researchers' bias towards seeing learners as defective language users, who always need to 'do more'. For example, Skehan and Foster (2007: 223) describe cases in which complexity and accuracy decrease as having 'a damaging effect on performance' so that 'performance suffers'. If this may appear to be evident for accuracy, it is not clear why lower syntactic complexity should necessarily indicate suffering and trouble. Norris and Ortega in this issue correctly warn that syntactic complexification indexed by subordination ratio may not always increase linearly, but they claim that this may happen because syntactic complexity grows in other ways, e.g. by phrasal and clausal complexification, seeming to imply that language production *must* necessarily grow

more complex over time, albeit in different ways.

Likewise, the study by Robinson et al. on the Aspect Hypothesis in this issue assumes that an increased ratio of atypical aspectual markings is 'more developmentally advanced'. However, no mention is made of the semantic or socio-pragmatic appropriateness of such atypical choices. A learner utterance such as *she was coming in while he washed the dishes* would certainly be an atypical application of progressive and past tense markings, but also inappropriate for most discourse contexts. Briefly, it is not clear to what extent atypical aspectual marking denotes accrued flexibility - the 'virtuosity' (Andersen 2002) of native speakers and advanced learners - or just maladroitness. Similarly, authors do not discuss the fact that so many learners in the Here-and-Now condition used past tenses - what were they doing? Was their performance adequate for the task's demands?

In cases like aspect or plural marking one might say that (grammatical) adequacy corresponds to accuracy, so that studies like the one by Robinson et al. should distinguish between atypical-accurate (acceptable) uses of aspectual marking and atypical-inaccurate (unacceptable) ones. However, in cases like syntactic complexity a sentence may be pragmatically inadequate but still grammatically accurate. For example, it is well-known that long and complex sentences are quite appropriate in the academic writing of a language like Spanish, in contrast with a more paratactic English prose. At certain levels, Spanish speakers may transfer their L1 habits to their English academic writing, producing over-complex but not ungrammatical sentences (Neff et al 2004). If socio-pragmatic adequacy were taken into account, for example by looking at native speakers' baseline data, it would appear that after a certain point a decrease in syntactic complexity might be interpreted as a sign of higher proficiency.

Wolfe-Quintero, Inagaki and Kim (1998: 2) write that 'we are not interested in measuring the ability to "write well" in a second language ... but in measuring language development as it manifests in a written modality'. It is certainly possible to measure anything in and for itself, but one wonders how performance can be meaningfully interpreted without reference to its quality and effectiveness. More specifically, one should be aware that fluctuations in CAF do not depend exclusively on psycholinguistic factors such as memory, automaticity of cognitive efficiency, but they may responsive to the task's semantic and pragmatic demands. A lower level of lexical or syntactic complexity may be due to some limitations in the mental lexicon or in the Formulator, but it may also depend on the speaker's pragmatic choices.

Conclusions

CAF measures are a good starting point for describing linguistic performance and its multidimensionality. In order to apply them it is first necessary to clearly define underlying constructs, so that each measure or group of measures refer to a well-identifiable construct. In this article I have proposed to set CAF constructs more clearly apart from other constructs. The first is interlanguage development and acquisition of new forms, which I propose to treat as an independent dimension instead of as a specification of complexity or accuracy. The second is adequacy, the appropriateness to communicative goals and situations, which should be seen as both an independent construct based on task success and as a way of interpreteing CAF measures.

Applying all these measures every time performance is described would be very onerous, and most studies will focus only on some of these, possible looking for

correlations with other studies that have analyzed other dimensions. In order to do so, it is necessary to reach some consensus about constructs and their operationalization, so that results can be compared across studies. A significant progress in the field would thus be the identification of a a limited set of standardized measures to be used across studies. It is also important to remember that one has to make choices and that measures are necessarily partial. It is therefore not advisable to group too many notions under the same term which, instead of providing a wider picture, just makes it more blurred.

References

Andersen, R. 2002. 'The dimensions of pastness'. In R. Salaberry and Y. Shirai (eds): *The L2 acquisition of tense-aspect morphology.* Amsterdam: Benjamins.

Bley-Vroman, R. 1983. 'The comparative fallacy in interlanguage studies: The case of systematicity,' *Language Learning* 33: 1-17.

Chafe, W. 1994. *Discourse, consciousness, and Time.* Chicago: University of Chicago Press.

Cumming, A. and D. Mellow (1996). 'An investigation into the validity of written indicators of second language proficiency'. In A. Cumming and R. Berwick (eds.): *Validation in language testing*, Clevedon: Multilingual Matters.

De Jong, N.H., M. P. Steinel, A. Florijn, R. Schoonen and J. Hulstijn 2007. 'The effects of task complexity on fluency and functional adequacy of speaking performance'. In van Daele et al (eds.).

Ellis, R. and G. Barkhuizen. 2005. *Analysing Learner Language.* Oxford: Oxford University

Goodwin, C. 1981. *Conversational Organization: Interaction Between Speakers and Hearers.* New York: Academic Press.

Homburg, T. J. 1984. 'Holistic evaluation of ESL compositions: can it be validated objectively?' *TESOL Quarterly:* 18, 87-107.

Hulstijn, J. 1995. 'Not all grammar rules are equal: Giving grammar instruction its proper place in foreign language tecahing'. In R. Schmidt (ed.): *Attention and awareness in foreign language learning.* Honolulu: University of Hawai'i Press.

Hulstijn, J.H. and De Graaff, R. 1994. 'Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal'. *AILA Review, 11:* 97-112.

Kuiken, F. and Vedder, I. (in press). 'Communicative adequacy and linguistic complexity in L2 writing: a case study'. In R.M. Manchon (ed:, *Learning to write and writing to learn in an additional language.* Amsterdam: Benjamins.

Levelt, W. 1989. *Speaking: From intention to articulation.* Cambridge MA: The MIT Press.

Neff, J., E. Dafouz , F. Díez., R. Prieto and C. Chaudron. 2004. 'Contrastive Discourse Analysis: Argumentative texts in English and Spanish'. In C. Moder (ed): *Contrastive Discourse Analysis .* Amsterdam : John Benjamins.

Norris, J. and L. Ortega. 2003. 'Defining and measuring SLA'.Iin .Doughty, C. J. and M. H.

Long.(eds): *The Handbook of Second Language Acquisition*. Oxford: Blackwell.

Ortega, L. 2003. 'Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing'. *Applied Linguistics, 24:* 492-518.

Pallotti, G. and S. Ferrari (2008). 'La variabilità situazionale dell'interlingua: implicazioni per la ricerca acquisizionale e il testing linguistico'. In G. Bernini, L. Spreafico, A. Valentini (eds): *Competenze lessicali e discorsive nell'acquisizione di lingue seconde*. Perugia: Guerra.

Robinson, P.2007. 'Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty.' *IRAL* 45: 193-213.

Sanell, A. 2007. *Parcours acquisitionnel de la négation et de quelques particules de portée en français L2*. Stockholm: Stockholm University

Skehan, P. and P. Foster 2007. 'Complexity, accuracy, fluency and lexis in task-based performance: A meta-analylsis of the Ealing Research'. In van Daele et al (eds.).

Spada, N. and Tomita, Y. 2007. 'The complexities of selecting complex (and simple) forms in instructed SLA research'. In van Daele et al. (eds.).

Tavakoli, P. and Skehan, P. 2005. 'Strategic planning, task structure, and performance testing.' in R. Ellis (ed): *Planning and task performance in a second language*, Benjamins, Amsterdam.

van Daele, S., A. Housen, F. Kuiken, M Pierrard, M and I. Vedder. (eds) 2007. *Complexity, accuracy and fluency in second language use, learning and teaching*. Brussels: Contacforum.

Wolfe-Quintero K., S. Inagaki and H. Kim 1998. S*econd Language Development in writing: measures of fluency, accuracy, and complexity*, Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.

Woods, A., P. Fletcher, P. and A. Hughes. 1986. *Statistics in language studies*. Cambridge: Cambridge University Press.

1   In the There-and-Then condition subjects had to tell the story in the past *without* looking at picture strips, while in the Here-and-Now condition they told the story in the present tense *with* visual stimuli under their eyes. The There-and-Then condition could thus be said to be more difficult for two reasons - memory burden and use of past tense marking - and it is difficult to tease them apart in interpreting results.