A Quantum-Cognition Approach to the Study of Second Language Acquisition

Stefano Rastelli

Centre for Applied Research and Outreach in Language Education (CAROLE) - University of Greenwich (UK)
stefano.rastelli@unipv.it

In this paper, I want to draw attention to the fact that – following other linguistic disciplines such as Semantics, Natural Language Processing, and Information Retrieval – Second Language Acquisition (SLA) research could also benefit from a quantum-cognition approach. In order to illustrate the potential of a future collaboration between SLA and quantum-cognition, I describe a case-study demonstrating a quantum effect in SLA, namely, the superposition of a grammatical and a statistical rule of noun-article agreement. I argue that a second language (L2) competence – like other phenomena in both nature and the human mind – is measurement-dependent, meaning that we can access it only contextually and probabilistically.

Keywords: *second language acquisition, quantum cognition, superposition, comparative fallacy, interlanguage*

Journal of Cognitive Science **17-2: 229-262, 2016**
Date submitted: 12/24/15 Date reviewed: 06/19/16
Date confirmed for publication: 06/28/16

1. Introduction

In this paper, I propose to apply some insights from quantum cognition to the study of second language acquisition (SLA). The combination of quantum cognition and linguistics in itself is not a novelty. Methods from quantum cognition have already been spreading across cognitive sciences and linguistics for at least twenty-five years (Bruza et al., 2015; Busemeyer & Yang, 2015). For instance, quantum probability[1] bears on non-compositional, vectorial semantics which considers some concepts to be inseparable and sentences to be combined in a non-compositional way, similarly to quantum-entangled, non-separable atomic particles (Heunen et al., 2013; Bruza et al., 2011). There are also quantum approaches to Natural Language Processing (NLP) and Information Retrieval (IR). For instance, a quantum model has been developed by Coecke and colleagues in order to overcome some methodological weaknesses of competing symbolic and connectionist theories of meaning[2] (Coecke et al., 2010). The quantum-cognition approach utilises the general framework of quantum physics to address decisive methodological problems related to the measurement of cognitive entities. For this reason, I propose that quantum cognition can offer a fresh perspective on the measurement of a second language learner's competence. In this paper, I will not provide the mathematical elements necessary to model L2 developmental data into a formal quantum framework. To support my view, I will instead discuss some plausibility arguments in order to describe the potential of quantum cognition theory for SLA research. This potential can be explained and understood by readers with an elementary knowledge of linear algebra, classical probability theory, and statistical methods. In the references section, readers will find more sophisticated information on the quantum theory framework along with a number of technical details that I cannot deal with in this paper.

In the last fifty years of SLA research, the competence of an L2 learner

---

[1] Quantum probability is the abstract foundation of quantum physics formalised by von Neumann and Dirac in the 1930s.

[2] This model computes the meaning of sentences based on both the meaning of words (in a vector space) and their grammatical roles by Pregroup analysis.

has always been acknowledged as being difficult to measure and assess. SLA researchers have often struggled to agree on the markers of acquisition and to answer the apparently simple question as to when an L2 learner eventually has acquired the target language. This question actually has two dimensions: One addresses the dynamics of the learning process, how it unfolds (in what stages and at what rate), and the various reasons why it is more or less successful. The other one aims at what exactly we can know of this process and at the best methods of eliciting such knowledge. For instance, it is traditionally disputed whether it is more fruitful to analyse learners' performance or learners' mental representations and processing routes. In most SLA research, regardless of the scientific paradigm that is adopted, the issue of describing the object and the issue of its knowability are kept apart. In fact, SLA is treated as a biological, cognitive, or socio-cultural phenomenon that unfolds independently of the various means adopted by researchers to describe it. In this paper, I argue that an important methodological fallacy of SLA research stems from the fact that the separation between the object and the procedures by which the object is measured is not questioned enough. My claim is that (a) the operation of observing and measuring an L2 learner's language from the standpoint of the target language (TL, the language that is acquired) contributes to determining both the nature and the dynamics of the object that is measured (an L2 learner's competence); and that (b) the foundational principles of SLA theories (e.g., the concepts of 'rule', 'stage of acquisition', and 'interlanguage') are measurement-dependent; that is, they are by-products of the procedures that are traditionally adopted to measure L2 development. Points (a) and (b) mean that decisive aspects of a learner's competence do not exist prior to measuring, but are elicited by measurement itself. The act of measuring L2 competence in fact literally brings into being those features of the phenomenon that fit the measurement. When it comes to SLA research, L2 competence is universally measured in terms of its distance from the TL. The output of such a measurement procedure based on a distance-metrics is that L2 items[3] can only be 'similar' or 'dissimilar'

[3] It is very important to note that the term 'item' can refer to either single words, sentences, or even cognitive functions such as expressing tense relations, expressing agentivity, etc.

to the TL. In the former case, L2 items are dubbed as ‘target-like’, while in the latter case, they are dubbed as ‘non-target-like’. Much of SLA research is about investigating two things: (a) *the cause* of similarity/dissimilarity; (b) *the dynamics* of similarity/dissimilarity over time. As to (a), the cause of similarity/dissimilarity has almost always been equated to a learner’s knowledge and internalisation of language rules. Learners’ knowledge of rules is considered to be the factor that ultimately determines whether a learner’s output is similar or dissimilar to a native speaker’s. As to point (b), the idea that learners progress through ‘stages of acquisition’ – which is still popular in SLA research – stems from the distance-based metrics and is measurement-dependent as well. In fact, ‘stadiation’ (the operation of establishing a learner’s level of proficiency according to a stage-based model) is a direct function of counting similarities and dissimilarities[4] in a learner’s comprehension and production. The higher the number of similarities that are found in a learner’s output, the more advanced their L2 stage is assumed to be. There are of course streams of L2 research that consider also non-target-like rules (differently principled, tentative language behaviors) which do not immediately result in similarities (see section 2.1). Still, these non-target-like rules (and their outcomes) are believed to be gradually dismissed by learners as their proficiency increases.

The adoption of a distance-based measurement by researchers has consequences. When bringing into being the objects ‘language rule’ and ‘stage of acquisition’, such measurement eclipses other aspects of L2 development that cannot be analysed in terms of similarity/dissimilarity to the TL. There are well-known cases in the SLA literature where similar, apparently target-like outcomes stem from the application of non-target-like rules (Rastelli, 2014). In section 4, I will instead discuss the opposite, i.e. less well-known cases in which L2 learners apply target-like rules, but produce output that can be dissimilar and non-target-like. This happens when L2 forms originate from the application of two different rules (statistical and grammatical rules) that coexist in a learner’s competence and have exactly the same scope. These rules can constrain the same portion of language (sequences of words, phrases, sentences) simultaneously, but

[4] This similarity/dissimilarity can be evaluated based on the correctness of form-meaning mappings or based on patterns displayed at the level of cognitive functions.

independently of one another. When these two target-like rules superpose, the resulting outcome might not resemble anything one could expect from the application of a single target-like rule. For the remainder of this paper, I will propose that SLA researchers could profitably integrate a theory of L2 competence with a theory of measurement. The quantum-cognition approach could provide the optimal theoretical underpinning for such an integrated 'Quantum SLA approach'. To describe this course of action, I will first provide a description of how certain measurement conditions affect first and second language studies (section 2). Then, I will describe the main quantum effects and how they could be extended to language studies (section 3). Sections 4 and 5 are dedicated to describing 'quantum L2 competences' and the functioning of quantum probabilistic rules in SLA, while section 6 describes what a learner's "intralanguage" is and why this concept is important for the quantum SLA approach.

2. Native language competence as a measurement-dependent object

The fact that any measurement has consequences on the thing that is measured can also be observed in language, namely in speakers' intuitions about their native language. Acceptability judgements[5] about well-formedness of sentences are manifestations of a native speaker's intuitions about her own language competence. They still constitute the empirical basis of generative syntax and a substantial proportion of generative-oriented SLA research.[6] In the past, criticism mainly questioned the extent

[5] Acceptability judgements can take the form of binary or n-point (as in Likert scales) choices, or magnitude estimations, and can be made on- or offline (temporised and associated with covariate analysis).

[6] A recent survey of articles that appeared in the journal Linguistic Inquiry from 2001 through 2010 (Sprouse, Schütze, & Almeida, 2013) estimated that 77% of articles were based on acceptability judgements (the remaining 23% were based on meaning/ambiguity judgements). Contrarily, a survey of articles published in Second Language Research, Studies in Second Language Acquisition, and Language Learning between 2007 and 2012 showed that only 85 out of 418 articles used judgement data (Gass & Polio, 2014: 157). The latter data seem more in line with the trends of generative-oriented SLA research also described in Slabakova et al. (2014).

to which acceptability judgements represent a natural and informative task as well as the extent to which parsing problems and intuitions about sentence grammaticality can be kept apart. More recently, the criticism also addressed the issue of relational influences. Even in the case of binary (yes/no) judgements of sentences presented to participants, relational influences from preceding decisions may affect speakers' judgements on subsequent ones. There are at least two instances of this relational effect, one is sentence-internal and the other one holds across sentences. The first one arises from multiple possible sources of unacceptability within a single sentence. These sources may combine additively or super-additively in speakers' judgements (Casasanto et al., 2010). It might be that participants are uncertain as to whether an item is correct or not. At a certain point in the sentence they might then encounter an item which they judge incorrect beyond doubt. This fact brings it about that the correctness of the former item is likely to be evaluated relative to that of the latter. A second instance of relational influence in acceptability judgements can be found when participants, instead of reporting their intuitions on one sentence at a time (or on two sentences, in the case of magnitude estimation), start relating the judgement of a sentence to all of the preceding sentences they can remember. This moves participants' focus from the task of judging the acceptability of the sentence at hand to the consistency of the unit of measurement they are using in the process of error detection. This unconscious activity is so pervasive that participants in acceptability studies often find and evaluate errors that do not exist, but are an artefact of the experimental conditions. The flaw of false negatives occurs because, in sentence acceptability tasks, the units of measurement used to judge the grammaticality of sentences are likely to change across sentences. Participants in fact fine-tune their current judgements based on previous ones. The fact that acceptability judgements cannot be shielded from the influences of relational reasoning pushed some researchers to admit that a psychological theory of linguistic judgements is still missing and is needed (see Cowart, 1997; Fanselow & Frisch, 2006; Fanselow & Weskott, 2009; Featherston, 2009; Gibson & Fedorenko, 2013. For a different position see Sprouse, Schütze & Almeida, 2013; Sprouse & Almeida, in press). Quantum cognition theory can provide SLA researchers with a framework that can explain relational reasoning and other phenomena that surface when we think about language. For instance, the quantum principle of complementarity – when applied to the field of human beliefs and decision-making processes – states than when one is asked a question, information carried over from the preceding question provides a context for the construction of the second and influences the subsequent response (Wang & Busemeyer, 2015: 2). The kind of relational reasoning that originates from acceptability/grammaticality judgments can be considered an example of human cognition where such a principle applies.

2.1 Measuring 'interlanguage'

The interlanguage (IL) is the language produced[7] by second language learners. The construct of IL has been one of the most important concepts in SLA since the birth of the discipline. As it is known, it was proposed by Larry Selinker more than 40 years ago (Selinker, 1972). The Interlanguage Hypothesis (IH) states that a learner's language is a linguistic system in its own right and that L2 development is driven by internal laws – and not by the L1 (Tarone, 2014: 9). The proposal was originally meant to overcome the predominance of errors in the analysis of learners' language production and to substitute the notion of *divergence* (between the IL and the TL) with that of internal consistency in a learner's language. There are important differences between the original statements of the IH and how it evolved over the course of forty years of SLA research. In fact, IL continued to be studied mostly as a transitional competence system[8] (for an overview see Han & Tarone, 2014). Discontinuities, jumps, U-shaped trajectories, and backsliding in this transitional phase were taken on board in the IH and

[7] With the words "production" and "produce" I mean whatever utterance is generated or comprehended by learners. It is of no relevance here whether the message that is generated at a mental level is actually produced or heard at phono-articulatory level. Magrassi et al. (2015) recently presented findings from their study using electrocorticograms recorded during awake neurosurgical operations in Broca's area. These results suggest that the activity of the language areas is already organised by sound when language is generated, even before any actual utterance occurs.

[8] This definition was proposed by Corder (1967).

handled with statistical models appropriate to deal with non-linearity, e.g. in Dynamic Systems Theory[9] (see de Bot, Lowie & Verspoor, 2005, 2007; van Dijk & van Geert, 2007; van Geert, 2003; Caspi, 2010). Nevertheless, the assumption that the IL is a transition with an endpoint (however the path may be shaped) has reigned over SLA research from the beginning and has never substantially changed since then.[10] The imposition of the endpoint perspective on the IL brought it about as a natural consequence that the procedure of error analysis – apparently dismissed in the theory after the anti-behavioural, Chomskyan whitewash – regained its attraction. Currently, error classifications (e.g. overuse, omission, substitution, distributional errors, etc.) and the comparison of error percentages over time as a way of tracking L2 development continue to be used in both cognitive and generative-oriented SLA, even though very few researchers would explicitly defend the view that learners' errors provide insights into internal processes. Many – also in the generative field – accept the seemingly unavoidable paradox that learners' competence is judged based on the evaluation of their performance (e.g. Hawkins, 2001). This reversion to error analysis was made possible because the original promises of the IH were difficult to keep without a preliminary theory of measurement and error classification. Since this theory is still lacking, error analysis continues to fill a theoretical void in SLA theory.

In quantum cognition terms, the situation described above can be summed up by saying that the IL is (wrongly) assumed to be an observer-independent system. The TL is used as an absolute (non-relational) measure and the IL is assessed in terms of its distance from it. Periodically,

[9] This theory accounts for "developmental noises", that is, the non-linearity and variability that accompany linguistic performance at any proficiency level of individual learners over time (the intra-learner variability). The theory gives an explanation of the fundamental non-linearity of language development, but does not discuss the point that second language acquisition is 'transition with an endpoint' (however non-linear this transition might be).

[10] Of no importance here are discussions about the nature of the endpoint or the reasons employed to explain the variability in learning outcomes. The only point that counts is that the learner's language is compared to a native speaker's, however the latter may be represented and thought of in sociolinguistic, pragmatic, and sociocultural terms.

SLA researchers become cognisant of the fact that measuring the IL in terms of distance from the TL represents a methodological as well as an epistemological flaw. Some SLA researchers seem aware that the issue of measurement should be addressed prior to any investigation. For instance, at a certain point, the question "how do we know what a learner knows?" was echoed in the SLA literature, giving rise to a point-counterpoint debate (Lardiere, 1998; Lakeshmanan & Selinker, 2001; Lardiere, 2003; Shirai, 2007). This debate could have been the occasion to reflect on the problem of IL measurement. Instead, the non-relational nature of IL measurement was re-affirmed by statements such as: (1) all we can know about what learners know is only derived indirectly from performance data (Lakshmanan & Selinker, 2001: 393; Tarone, 2014: 15); (2) all we can know about what learners know is to what extent the IL differs from the TL (Lardiere, 2003: 140). So the issue of measurement underlying the question "how do we know what a learner knows?" remained unanswered and did not reach its theoretical potential.

2.2 The comparative fallacy

There is a cost SLA researchers must pay every time they avoid answering that question and continue to use the TL as a benchmark for IL measurements. This cost is known as the 'comparative fallacy'. In SLA research, the comparative fallacy (CF) is the mistake of studying the systematic character of one language by comparing it to another or by super-imposing the TL standpoint and the TL categories on learner data (Huebner, 1979; Bley-Vroman, 1983; Klein & Perdue, 1992; Cook, 1997, 1999; Lardiere, 1998; Lakshmanan & Selinker, 2001; Lardiere, 2003; Selinker, 2014: 233). The CF is more likely to be committed every time one uses the notion of accuracy percentages in obligatory context and every time one tries to reconstruct what a learner intended to mean. In the former case, the researcher counts the suppliance rates of – for instance – correct and incorrect past tense marking in obligatory finite past contexts, that is, in all contexts where finite past verbs are expected to occur in the TL. In the latter case, the researcher assumes native speaker intuitions about the meaning of verb stems by assigning semantic categories (such as activity, achievement, etc.) to the data, and by applying diagnostic tests for those categories (Lardiere, 2003: 136). Yet, it is possible that initial L2 learners do not have any concept of a verb being telic or atelic and that they are only worrying about the general meaning of that L2 verb (Giacalone Ramat & Rastelli, 2013). Two ways of avoiding the CF were proposed in the late eighties and in the nineties. The first one was put forth in the framework of cognitive linguistics. It was suggested that one should not look at errors but at the dynamics of form-meaning mappings over time. For example, researchers who want to investigate the acquisition of the past tense under the concept-oriented approach should not look at suppliance rates of correct past tense marking, but at how learners express the concept of temporality – regardless of the fact whether these expressions are TL-appropriate or not (Bardovi Harlig, 2000). Note that in this framework, the study of variability in a learner's production is still not considered "in its own right". Rather, the focus of research has shifted from accounting for mere errors to accounting for the different (non-target-like vs. target-like) linguistic means that learners use to encode supposedly cognitive universals. Once and again, L2 competence is seen as a non-relational notion. The second way of avoiding the CF was put forth by Manfred Pienemann in his Processability Theory framework. In this theory, the construct of emergence was meant to substitute that of accuracy in mandatory contexts. Emergence is "the first systematic use of a structure, so that the point in time can be located when a learner has—in principle—grasped the learning task" (Pienemann, 1984: 191). In this case, the supposed theoretical advancement is that error classifications are different and that errors are defined and counted in more sophisticated ways. This does nothing to avoid the CF as the standpoint from which the IL is measured is still the TL. As a matter of fact, as long as IL is treated as being an observer-independent, non-relational notion, the CF simply cannot be avoided. As it happens, one will always measure the IL in terms of distance and differences from the TL, it does not matter whether these differences are conceived and operationalised in terms of errors or in terms of the cognitive operations that are underlying language functions.

3. Quantum effects

Quantum cognition is a new research programme that uses mathematical principles from quantum theory as a framework to explain human cognition, including judgement and decision making, concepts, reasoning, memory, and perception (Busemeyer & Wang, 2015). As Aerts et al. (2016: 1) put it, in quantum cognition theory, empirical phenomenology in the mind-related domains (states, measurements, and probabilities of outcomes) are represented by utilising the same quantum-theoretic probabilistic model which is utilised in the phenomenology of microphysics. In this section, I describe some quantum cognition studies that could be relevant when one attempts to treat language competence as a relational notion. These studies describe the effects of measurement on the phenomenon that is measured, that is, how an observer system actively interacts with an observed system to the extent that some of the measured values are a direct consequence of the very measurement. In the next sections, I look at three of these quantum effects: non-commutativity, the 'Guppy' effect, and superposition.

3.1 *Values do not predate measurement*

Behavioural experiments in the last twenty years have shown that when cognitive phenomena (such as human judgement and preference) are evaluated, the act of measuring may create some values that did not exist prior to measurement (Busemeyer & Wang, 2015). Aerts & Aerts (1995) investigated what kind of knowledge can be gained when participants in a survey are asked two different questions. The first question is "are you a smoker?". The second question is "are you for or against the use of nuclear energy?". If 21% answered yes to the first question, the 0.21 probability can be interpreted as something that fills a void in the authors' knowledge about the values that are real and independent of being measured. This void is called lack of knowledge of actualities (in this case: the actuality of being a smoker). If 21% of the sample answered yes to the second question, however, this probability cannot be referred to a situation that predates the experiment (the survey) because some participants might not have had an opinion on nuclear energy before they were asked about it. The survey itself (the way the question is asked) has contributed to the manifestation of the values that are being investigated. Probabilities that emerge from the latter question are referred to as ‘genuine potentialities’ and should be analysed in a non-deterministic framework as if values in a responder’s mind were all overlapping (superposing) before the question was asked. It is important to clarify that quantum theory does not make the trivial point that all measurements are relative or arbitrary and impose an artificial perspective on the observed phenomena. Quantum cognition theory rather updates the old philosophical stance that the reality of a phenomenon is not independent of its observability. In any measurement there are indeed two systems interacting: the observer system and the observed system. An interaction means that the information an observer system can get from the observed system is always relative to their reciprocal position and configuration.

3.2 *Quantum effects: Non-commutativity*

Not only does the nature of the question have an impact on the knowledge that is elicited, but the order of survey questions has been found to affect the hierarchy of values in the participant’s mental state. The feature of non-commutativity of variable values in human judgements means that a preceding question creates a relation of entailment or causality that works as an abstract benchmark against which a subsequent question is evaluated. Considering the first question would therefore alter the participant’s state of mind, which in turn affects the consideration of the second question. For instance, if a person is asked whether she is happy and whether she is employed, the order in which the questions are asked has an impact on the potential answers. Pothos & Busemeyer (2013: 259) formally consider the comparison between first asking “are you employed?” and then “are you happy?” versus first asking “are you happy?” and then asking “are you employed?”. In classical Kolmogorovian probability theory, it is predicted that the probability (Prob) is commutative, that is, it is order- and context-independent, as is formalised in [1]:

[1] Prob (employed $\wedge$ happy) = Prob (happy $\wedge$ employed)

However, in quantum probability theory, conjunction of incompatible questions fails commutativity because the action of clarifying a mental state elicited by one question interferes with the measurement of a subsequent one. Therefore it could be either that having elicited first the certainty about one's happiness makes the same person fearful and uncertain about – say – the stability of her employment over time or, conversely, that being certain about one's employment casts doubt about the fact that a person's happiness entirely depends on her job, and so on. The formalisms in [2] and [3] describe in elementary manner how this situation of conditional probability is represented, respectively, in Kolmogorovian classical probability theory and in quantum probability theory for the conditional probability of "happy" under the condition "employed":

$$[2]\ \text{Prob (happy|employed)} = \frac{\text{Prob (employed} \wedge \text{happy)}}{\text{Prob (employed)}}$$

$$[3]\ \text{Prob (happy|employed)} = \frac{\text{Prob (employed} \wedge \textit{then}\ \text{happy)}}{\text{Prob (employed)}}$$

In technical terms, it is said that in [3] the events *happy* and *employed* are not stochastically independent because the occurrence of employed is thought by the responder as affecting the probability of *happy*.[11] Quantum cognition studies propose that the context of measurement (here: the order in which the questions are posed) has elicited a quantum effect. Before the first question is asked, different variable values (different answers to questions about happiness) may coexist in a person's mind. After the question is posed, the values in the observed system (human judgements) collapse (that is, change) in such a way that the probability of being happy given the occurrence of being employed must also consider all possible entailments as well as the resulting causal directionality triggered by the order of questions in the computation. The non-commutative feature of the observer system would therefore affect the overall outcomes of probability measurement.

[11] The events *happy* and *employed* are independent if their joint probability equals the product of their probabilities.

3.3 *Quantum effects: the 'Guppy' effect*

Quantum cognition studies in the last three decades observed that people conjoin and disjoin concepts in ways that violate the rules of classical logic. The overextension of conjunctions of concepts is referred to as the 'Guppy' effect. This effect surfaces – for example – when one measures the category membership of exemplars. According to classical probability theory, the overall probability of conjunct terms (such as something's probability of belonging to the category of "Food and Plant") derive from the additivity of their single probabilities (of being Food and of being a Plant, separately). Aerts et al. (2013) report that this additivity is captured by equation [4] (where μ stands for membership, Δc is the minimum difference between conjunct terms and terms taken singularly, and A and B are two concepts such as "Food" and "Plant"):

[4] $\mu\text{ (A and B)} - \min(\mu(A), \mu(B)) = \Delta c \leq 0$

Hampton (1988) first considered the concepts "Food", "Plant", and their conjunction "Food and Plant" in a famous experiment. He measured how participants decided whether a certain exemplar was or was not a member of each concept. When participants were asked about the exemplar MINT, the relative frequency of membership was 0.87 for the concept "Food" and 0.81 for the concept "Plant". For the conjunction "Food and Plant", the relative frequency was unexpectedly higher (0.9). These results violate the equation [4] because they show that participants perceived the exemplar MINT more distinctly as a member of the conjunction "Food and Plant" than they perceived it to be a member of either of the two component concepts "Food" or "Plant" separately (Aerts et al., 2013: 242). This category membership measurement elicited the existence of a mental state where the concepts of "Food" and "Plant" are conceived as an inseparable whole and not as a sum of their sub-components.[12] The Guppy effect can also be

[12] An anonymous reviewer pointed out that it is currently debated whether the effect described here is really due to entanglement (as hypothesised by Aerts & Sozzo, 2011) or just to interference (the phenomenon by which a new concept can emerge out of two constituent concepts, Aerts et al., 2012).

relevant for SLA research because it shows that concepts and categories are not just "containers" according to which items are labelled and stored, but they can be viewed as entities in a specific state that is changing under the influence of a context and that is affected by measurement (Gabora & Aerts, 2002). In this paper, the validity of this statement is extended to rules in an L2 learner's competence. As in the case of concepts, L2 rules are not static. They can combine and interact in a learner's performance. When this happens, it can be totally irrelevant for the purposes of a scientific description of an L2 competence whether the outcome of this interaction is target-like or not.

3.4 *Quantum effect: Superposition*

Let us suppose that you are asked whether God is male or female. You could be personally convinced that God is both male and female, or neither male nor female, and this uncertainty (which in its technical sense is conceived as a "superposition" of states) can be accepted as being a state of your mental observing system without being further questioned. In different circumstances (say, during a conversation), your language may force you to abandon and resolve uncertainty and to represent God in the speech system by using – say – the masculine gender, like in Italian. That does not necessarily mean that you also think that God is male.

Ambiguity (technically defined as "uncertainty", see section 4.4) is indeed a legitimate state in your mental observing system (which is different from the speech observing system). So while in the mental observing system God is described as being both masculine and feminine without contradiction, in the speech observing system, *he* cannot be described in the same way. By using either gender in conversation one does not mean to deny the existence of a mental superposition state. Since we think of God always relative to a mental viewpoint (the observing system) and speak about God always relative to the speech viewpoint (another observing system), there is no contradiction in saying – for instance in Italian – *il buon Dio è madre* ('the good Lord is a mother') where the phrase *il buon Dio* is masculine. This statement and the utterance that conveys it are in fact measured simultaneously from two different viewpoints or observing systems. Rovelli (1996) proposes that many scientific observations result from the interaction between the observer (any measuring device, not necessarily a human being) and the object being observed (physically interacted with) and are not a property of the object, nor are they due to any technological limitations of measuring instruments. Rovelli (1996) claims that the notion of an observer-independent state of a system (or observer-independent values of its physical quantities) is no longer tenable. In his view, any theory can only describe the information (the knowledge of the number of different possible states) that different systems have about each other. In the following section, I will apply this first to language in general and then to SLA. In the latter case, I will be considering the native speaker's competence as the observing system and the learner's IL as the observed system.

4. Quantum L2 competence

4.1 *Deterministic and probabilistic approaches*

In this section, I will describe and contrast the deterministic and the probabilistic[13] approaches to SLA research. SLA theories that rely – more or less explicitly – on accuracy data to draw conclusions on L2 competence all have in common a deterministic idea of the learners' competence. According to this idea, if a learner knows a rule, s/he is expected to apply it in most of the appropriate contexts. The more often they apply it, the better they know it. If learners fail to apply the rule, then performance factors (elicitation circumstances, processing difficulties, working memory limitations, etc.) are invoked to explain deficiencies.

According to a probabilistic approach of L2 competence (see section 4.3), however, the investigation of L2 acquisition must include preliminary information about the intended measurement ("how can we know what

[13] The term "probabilistic" has nothing to do with a statistical model of language learning and use which denies the existence of any innate, genetically endowed Language Faculty or grammatical categories. The term "probabilistic" is used here as in the quantum theory and does not describe the nature of the described system (L2 competence), but our way of gaining knowledge about it.

learners know?"). Asking this preliminary question prevents us from concluding that a learner knows a rule simply because we counted enough instances of the potential applications of that rule in the IL. Probabilistic quantum SLA instead tries to determine how likely it is that a learner applies any kind of rule in a given context (sentences, utterances, etc.). To illustrate the advantages of a probabilistic approach, I will present a probabilistic description of a trait of L2 competence (see section 4.3) where the idea of developmental uncertainty[14] and the quantum effect of superposition account for cases of apparent unattained acquisition.

4.2 *Deterministic L2 competence*

Let us consider a TL structure such as gender and number agreement between the determiner and the noun in the Determiner Phrase (DP). Let us suppose that a learner's suppliance of this structure in mandatory contexts is around 50%. Given that this percentage is below even the lowest cut-off point for attainment proposed in the literature so far (see Pallotti, 2007), many would conclude that this learner has not acquired this TL structure yet. A further step would be to analyse all mandatory contexts in which this TL structure is expected to be supplied. One can in fact suspect that the TL structure is provided correctly or incorrectly by the learner as a function of the backward transition probability score (BTP)[15] (calculated backward between the noun and the preceding article). One can hypothesise that – if a certain noun is frequently preceded by an article in the TL input

---

[14] In the quantum theory, the term "uncertainty" does not mean that we are unsure about something. The term defines instead the situation of phenomena before they are measured, since any measurement restricts the field of possibilities for some aspects of the phenomenon to be captured. Therefore uncertainty is the sum of all probable states that a phenomenon can be in when it is not being measured. If L2 development is one of these uncertainty states, all we can know about a learner's competence (e.g. the application of a rule) is probabilistic.

[15] Transition Probability (TP) is the probability that Y follows X in the sequence XY or that Y precedes X in the sequence YX. In the former case we have a forward TP (FTP); in the latter case we have a backward TP (BTP). Both FTP and BTP are relevant for the detection of regularities.

– the learner is more likely to get it right most of the time. If a certain noun is not frequently preceded by an article in the TL input to which the learner is exposed, then the learner is likely to perform more poorly. Let us suppose that a linear regression analysis confirms our hypothesis: The higher the BTP score, the more likely is the learner to choose the right form. The information we have now is the probability ($p \leq .05$) that the learner gets it right – given that she has already encountered that article + noun combination in the input. At this point, one would be tempted to conclude that the only rule at work in a learner's competence is a statistical (frequency-driven) rule and that any other abstract grammatical rules should be excluded. This is an example of a deterministic fallacy that occurs when interpreting accuracy percentages and regression data. The fallacy is that a statistical rule – which can in fact account for only 50% of occurrences – is invoked to explain all of them. It is supposed that the application of the (statistical) rule will yield 50% of correct forms, while the lack of its application explains the remaining incorrect 50%. Neither of these statements are correct though. What we can say at best is that we have found a number of contexts (50%) in which it is likely that a statistical rule was applied by a learner. We cannot say anything about the remaining 50%. This argument is valid for grammatical rules as well. Let us imagine that the TL structure is provided correctly or incorrectly by the learner as a function of the degree of entropy of the grammatical rule (Avrutin, 2006a, 2006b, 2012; de Lange, 2008; van Ewijk & Avrutin, 2010). Some features of the rule (e.g. gender agreement between determiner and noun in the DP) are easier to learn because the amount of information to be handled is reduced. Let us suppose that a simple chi-square test reveals that percentages are significantly different for the feminine *la* with respect to the masculine il because *la* has higher entropy (in fact, *la* is also a clitic pronoun in Italian). One could be tempted to conclude that the only rule at work in a learner's competence is a grammatical rule and that any statistical rules should be excluded. However, one would commit the same deterministic fallacy in this case as in the case of the statistical rule described above.

4.3 *Probabilistic L2 competence*

The alternative quantum interpretation of the data above is the following. In 50% of cases, the application of a statistical or grammatical rule is likely. The remaining 50% wait to be explained though. The probabilistic approach does not accept the explanation that – in 50% of circumstances – learners failed to apply the rule because they either have not met enough instances of the right article + noun combination in the input or because the entropy of the grammatical rule is too high. The quantum SLA approach explains the remaining 50% as a quantum effect of superposition of different rules that coexist in a L2 learner's competence. These superposing rules are statistical and grammatical rules. Although taken in isolation these rules are target-like, the outcome of their superposition might not be such. Superposition of rules brings it about that a learner can do, and actually does, the same thing in different ways simultaneously. A learner can process separately and put in a hierarchy the article and the noun because her mental grammar enables her to merge the two and project the DP head feature. At the same time, the learner can remember which nouns are preceded by which articles because she has already met this combination in the input. Let us suppose that the first capacity does not exclude the other, but rather they interact, compete with, or support one another. Most deterministic approaches and theories (with relevant exceptions though[16]) exclude the possibility of any interaction between two competences. Instead, we imagine how the grammatical rule and the statistical rule interact in probabilistic terms. For instance, we can assume that a learner perfectly knows the grammatical rule of gender agreement between the article and the noun and applies this rule in 50% of contexts. As to the remaining 50%, however, the learner might struggle to generate the correct form – not because she ignores the rule, but because a statistical rule covers the same range of words to which the grammatical rule would apply. This statistical rule could be that the noun in the DP occurs much more frequently in the feminine singular than in the feminine plural in the learner's input. So this statistic bias would simply break the grammatical rule down in two sub-parts, the one that concerns the

[16] E.g., Townsend & Bever, 2001; Yang, 2002; Newport, 2011; Nooteboom et al., 2002..

feminine singular being more likely to be acquired than the one concerning the feminine plural. Both accuracy percentages and the emergence criterion would miss the existence of this interference of a statistical rule. The former would interpret the 50% as incomplete attainment, thus underestimating the presence of grammatical representations in the learner's competence. The latter would instead overestimate the importance of the emerging half-rule (only feminine singular) because it would see a prospective TL rule when there are two (statistical and grammatical) rules already at work and interacting. As an opposite case, we can assume that a learner might already have encountered a given combination of article + noun in the input many times (higher than average FTP), but the presence of an overlapping grammatical rule with identical scope might drive her to keep the components of this combination dissociated in her competence for a long time. There are many reasons why this can happen. One of these reasons is that statistics can be modelled by different factors. Ellis & Cadierno (2009: 118) list a number of variables whose values can make input frequency more or less influential for language acquisition: type-token frequency, Zipfian distribution, recency, salience and perception of form, prototopicality of meaning, redundancy, etc. It is believed that storage of this combination in memory depends on token frequency, while schematisation depends on type frequency. In fact, a combination becomes a more abstract and therefore more productive construction if it is associated with multiple items rather than with a single lexical entry. Let us suppose that the collocational head (the noun) in our combination has a very low type frequency (it occurs in very few different combinations) but it has a very high token frequency (there are many instances for each of these few combinations). According to Ellis & Cadierno (2009: 119), this is a situation where irregular forms are very likely to be entrenched and stored in memory as a whole by L2 learners. We have a concrete case of this situation in Italian. The noun *problema* ('problem') is exceptionally frequent in both written and oral contemporary Italian.[17] It is also an irregular form, because it is masculine even though it displays the feminine ending *-a*. The article + noun combination *il problema* ('the problem')

[17] The noun problema is ranked 87th in the Lessico dell'Italiano Parlato LIP corpus

has a very high token frequency in both oral and written Italian because it covers one third of all occurrences of the word *problema* alone.[18] Following Ellis & Cadierno (2009), we would expect that this irregular combination is easily acquired by L2 learners of Italian and easily stored as a whole as well. This is not the case though. Different learner corpora of Italian[19] show in fact that even advanced learners of Italian continue to alternate the TL and the non-TL combination as in [5]:

[5] *il problema* | **la problema*

In the latter combination, learners seem to have dismantled and analysed the components of the combination regardless of its frequency as a whole. The use of the feminine article *la* instead of *il* is the consequence of a grammatical misanalysis that results in a wrong feature [+feminine] being projected by the phrase head. This non-TL combination neither stems from the lack of a grammatical rule, nor does it come from low frequency. Rather, it derives from the interference between two target-like rules: a statistical rule and a grammatical rule. This interference prevents the learner from delivering the article + noun combination in real time speech, even though this combination (*il problema*) is among the most frequent in modern Italian. Combinations with high token frequency and a collocational head which is irregular and has low type frequency might represent a leak in a learner's statistical competence where a grammatical analysis infiltrates and interferes with statistics to yield a non-TL combination.

4.4 Developmental uncertainty

The two cases above were chosen to show what developmental uncertainty consists of. It is the theoretical attitude of considering the IL a measurement-dependent system (see section 4.1). Before competence is measured probabilistically in context, we cannot know anything about either the presence of rules or whether these rules will be applied. In fact,

[18] See for instance the CORIS/CODIS corpus of written Italian at http://corpora.dslo.unibo.it/TCORIS/.

[19] See for instance the ISA (Italian writings of American students) corpus (Rastelli, 2007).

knowledge of a rule does not mean that the expected TL form will be supplied and – in contrast – suppliance of a TL form does not guarantee the presence of a corresponding rule. If L2 competence is dealt with as a relational concept, we can only predict the probabilities of a given state's occurrence (e.g. that one or more rules are applied) when taking a certain measurement (in certain contexts). Developmental uncertainty is the default way to refer to a learner's mental competence in probabilistic terms. Not all measurements are probabilistic or able to reveal superposition of a statistical and a grammatical rule. As we have seen, developmental uncertainty remains invisible to accuracy measures. In fact, I have shown that both correct and incorrect suppliances can derive from the interference between different rules coexisting in a learner's competence. Overall, this interference could not affect the 50% percentage of accurate suppliances seen in the last section. So one can be tempted to believe that a learner who fails to deliver 50% of accurate forms does not know the rule while she actually applies two different rules. In quantum SLA, L2 competence is seen as a word-by-word affair (a few-milliseconds affair) where learners can apply different overlapping statistical and grammatical rules simultaneously. Any string of words is therefore a context for a probabilistic measurement of SLA. In this measurement, the TPs of words in the string, their projected grammatical features, and the hidden variables form a unique vector state. In the next section, I will describe how this vector visualises geometrically the probabilities of a learner applying a given rule in a given context (under certain measurement conditions).

5. Computing the probability amplitude of quantum L2 rules

5.1 *A quantum L2 rule*

A quantum L2 rule is not a set of defining grammatical features to be learned, as it is proposed in generative oriented approaches. Nor is it an analogy-based generalisation that learners abstract from the concrete instances in the input they are exposed to (like in the constructionist or emergentist approaches). A quantum rule is a set of defining abstract grammatical features that cannot be separated from a set of instances of these features stored in a learner's memory. One can imagine a quantum rule as an abstract algorithm whose record of previous applications is one of the factors that affect the probability of its next applications. The quantum rule accounts probabilistically for a dual relationship that keeps words together in a sentence. This relationship can be both hierarchical and statistical. Any sentence (uttered or comprehended) is a probabilistic measurement of the balance between two forces that are simultaneously at work. These forces may head in different directions or in the same direction. These forces are the phrase-head that projects its features, and the collocational head that decides which is the most likely word that will precede or follow. This tension between forces holds regardless, whether words are adjacent or not (an important exception being non-combinatorial grammar, see Rastelli, 2014, chapter V). Unlike other approaches that recognise vaguely that language phenomena can be accounted for by "the right mixture of general principles governing cognition and statistical biases" (Benítez-Burraco & Boeckx, 2014: 123), the quantum SLA approach makes the distinct point that our knowledge of whether L2 learners are driven by any rule (and how) can only be probabilistic. A quantum rule is a rule that *might* be at work under a specific measurement. Researchers can only measure the probability amplitude that an interaction between statistical and grammatical rules is at work.

5.2 *Probability amplitude of an L2 rule*

I will now partially follow Feynman (1985, chapter I) for a naïve, informal way of representing graphically the vector state of an L2 rule when trying to compute the probability amplitude for an L2 rule to occur under two different measurements. My representation is not meant to be alternative to that representing formally event projections in subspaces (Busemeyer & Wang, 2015). It only has the limited purpose of visualizing how an L2 rule can be described in probabilistic terms. This description is alternative to any deterministic, distance-based metrics of measurement (see section 1). Let rule *x* be noun-article agreement in the DP under a grammatical measurement and let rule *y* be the same rule, but this time measured under a statistical measurement (in the following discussion, a 'measurement' is equal to a 'context'). We want to measure the probability that the event 'noun-article agreement' occurs in a sentence by a L2 learner. This means that we want to know how it is likely that the L2 learner has applied the noun-article agreement rule in a sentence. Since we have two different measurements or 'contexts' {x,y}, we can imagine that the probabilities for these two events to occur is equal to the square of the length of two arrows which represent our rule vectors {x,y}[20]. The length of each arrow {x,y} is not solely determined either by the grammar or by statistics, but it is an interaction of two vectors. in our example, this length is determined by the entropy of the grammatical rule together with the statistical biases (token and type frequency, TPs, etc.) that keep together the combination article + noun. Therefore we must draw two arrows for each rule {x,y}, one for each way the event 'apply a rule' will happen (the statistical and the grammatical one) and then combine them in a final arrow whose square represents the overall probability magnitude derived from the interaction of events under measurements {x,y}. Let us start with rule x as it is measured by the context *il problema* 'the problem' (see section 4.3). Let us decide that the relative probability that an L2 learner uses the grammatical rule of agreement in this DP depends on the degree of entropy of the noun-article channel.[21] Let the degree of entropy of this channel be calculated as in van Ewijk & Avrutin (2010).[22] Since the lower the entropy, the higher the probability for a rule to be applied, we can use for instance the inverse function f(x) = 1/x to plot values in inverse proportion to their

[20] In this and in the following paragraph, the value corresponding to the square of length is not taken as an absolute probability value (that must be comprised between 0 and 1) for an event to occur, but as an indication of the probability amplitude. This probability amplitudetude is exclusively relational. this means that it can be used to the only purpose of allowing comparisons between different values.

[21] There might be other deciding factors at work of course, but the procedure will be the same.

[22] In information theory, entropy is the degree of uncertainty of a message. In language studies, the entropy of a morphosyntactic rule is defined as the inverse function of the capacity of the channel existing between two or more items through which the information is transmitted. For instance, the information or 'message' between items such as articles and nouns in the DP is all feature information (gender, noun, case, etc.) sent by a noun (which is to be produced) to the set of potential articles it can be preceded by in

magnitude[23] and then have them normalised to range from 0 to 1. Let us suppose that for the rule *x* we got a very high (inverse) value of entropy = 0.3. Let us also suppose that such a purely speculative value is caused by an information overload due to a serious mismatching between the information carried by the absolute token frequency of the masculine article *il* (which is the most frequent word in both spoken and written Italian) and the information carried by the morpheme *-a* (which in Italian is basically always feminine). This value of entropy corresponds to an arrow whose length is 0.3, and to a square of relative probability of 9% (0.32 = 0.09). Let us then decide that the relative probability that an L2 learner uses a statistical rule in the same article + noun combination depends on the BTP between *problema* and *il*. Let this BTP be calculated as the normalised (from 0 to 1) ratio of the occurrences of the whole combination *il problema* and the noun *problema* alone in spoken and written Italian corpora. Let us suppose that for this statistical rule we obtain a very high BTP value of = 0.7. This statistical value corresponds to an arrow whose length is 0.7, and to a square of relative probability of 49% (0.72 = 0.49). We must then draw two arrows, one for each way the event 'apply a rule {x,y}' will happen. In order to combine {x,y} vectors (arrows), we must put the head of x against the tail of y and draw the final arrow from the tail of x to the head of y. My purpose is to combine {x,y} with a final arrow whose square represents the overall probability magnitude that an interaction of different rules will be at work in the combination *il problema*, like it is represented by Figure 1.

5.3 *Overall probability for an L2 rule*

A problem arises because – in order to draw the final arrow in the

a given language. Too much information or too little information both represent difficult cases to handle for speakers and learners. Grammatical items have different individual information loads and therefore different degrees of entropy. Frequency of a word and the number of functions a particular inflected word form can perform are taken into account in measuring the information load. De Lange (2008) found that the information load for individual articles played a role in the acquisition process of those articles in Dutch, Italian, and German preschoolers.

[23] This excludes the possibility that entropy = 0.

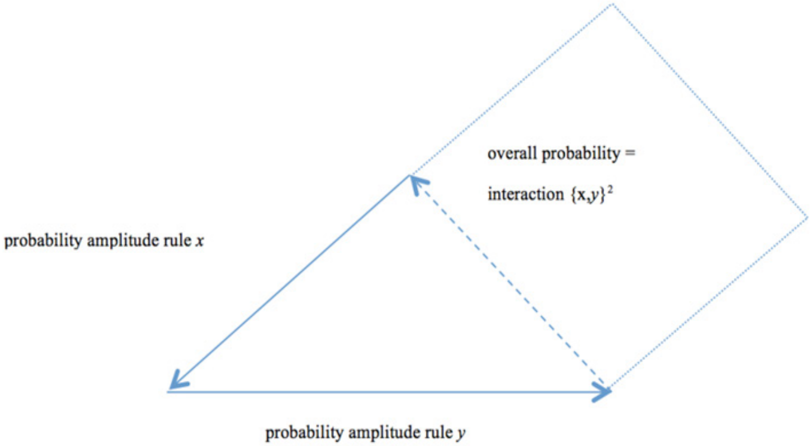

Figure 1. Probability amplitude for rules {x,y}

triangle and to compute the correct square value of the interaction {x,y} – we need to know the orientation (direction) and the angle of arrows *x* and *y*. As to the angle of vectors, similarly to what happens when computing the probability that photons are absorbed or reflected by a mirror, this information is a function of time. Unlike Feynman (1985: 27), we cannot use a stopwatch that can time arrows as they move across time. Since we are dealing with SLA, we can interpret the variable TIME as "length of acquisition" to represent the amount of time flowing from the Onset of Acquisition (OoA) – which is the point zero – up to a given moment in time which is represented by a value on the X axis. In our fictitious case, let this value be = 6 (months). Let us plot the values we obtained for *x* (entropy) and y (BTP) on the Y axis. In order to calculate the angle of the vectors, I create an intersection between the orthogonal projection of TIME and the orthogonal projection of {x,y} on Y. In this way, the arrow for rule *x* will point to one intersection while the arrow for rule *y* will point to the other. This will determine (precisely as it can be in a naive representation like this) the angle of each arrow.[24] Once both the length and the angles of

[24] We have assumed that arrows representing rules are iso-oriented because they move on the same developmental dimension on the X axis. It is of course possible that one

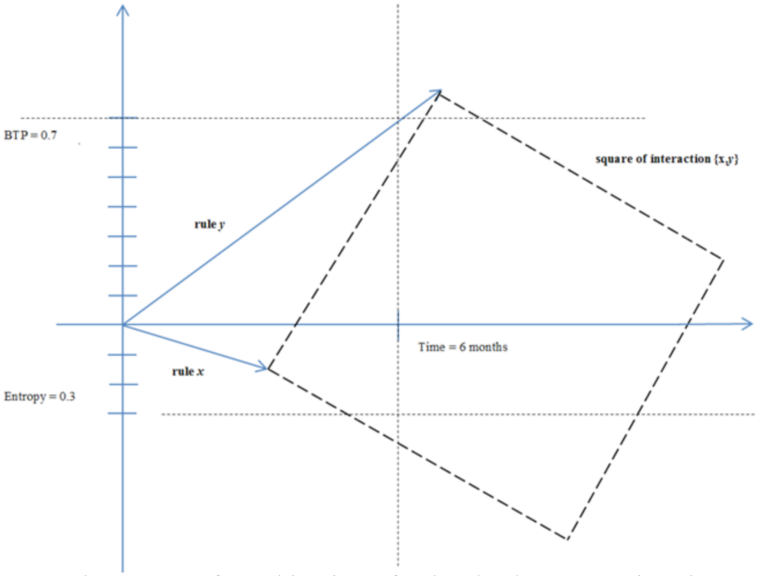

Figure 2. The square of combination of rules {x,y} representing the overall probability

the arrows are set, we can combine {x,y} with a final arrow *z* whose square represents the overall probability magnitude (which is only a *relational* value, see footnote 20) that an interaction of different rules will be at work in the combination il problema, as it is represented by Figure 2.

The value z^2 – once normalized – could work as a primitive assessment of the overall probability magnitude that an interaction of different rules (different measurements) has given shape to the unique, target-like outcome *il problema* as a function of both the rule features and of time (the developmental dimension). This z^2 value is purely relational because it only makes sense if it is compared to other values that are derived in the same way. For example, if we compute the probability amplitude for the similar – albeit non-target-like – context **la problema*, we would probably discover that the corresponding square of the final arrow does not differ much from our z^2. It could be in fact that a very low value for entropy

disregards the variable of TIME. In this case, the angle will be chosen arbitrarily and will be the same for each arrow.

(and a very long corresponding arrow) would compensate a very low BTP value, so that the final area of the square is roughly similar. This would explain why the alternation [5] is so frequent in Italian learner corpora. The purely speculative computation of the probability amplitude of quantum L2 rules explained in this paragraph offers an abstract way of looking at developmental data, regardless of the accuracy (similarity vs dissimilarity to the TL) of the form in question. Such quantum computations make it possible to shift from the study of the IL to the study of 'intra-language'

6. A learner's intralanguage

'Intralanguage', not IL, is the object of enquiry in quantum SLA. Intralanguage is the study of how L2 forms are generated, regardless of whether they change or remain identical over time, and whether they are target-like or not. In the study of intralanguage, errors and TL forms are treated on a par so that the comparative fallacy is avoided. Treating errors and TL forms on a par means that a form is not defined according to its face value, but according to how it is generated, represented, and processed in a learner's mind. Both correct forms and errors can be generated and processed either statistically or grammatically. Statistically generated errors, for example, comprise incorporation, reprise, and repetition of another's speech turn. Grammatically generated errors are classical, well known overgeneralisations of a rule (*goed instead of went). One purpose of quantum SLA research could be to ascertain whether an L2 form is grammatically or statistically generated/processed, but not whether a form is TL-appropriate or not. To do that, quantum SLA uses a probabilistic model of developmental L2 competence. Probabilistic modelling means that we can never make absolute (non-relational) predictions about which form a learner will use in a given context. We can only predict the probability that they use either under certain measurements. An L2 learner's competence is conceived as being a superposition of mental states3 whose uncertainty is decided a posteriori, in an utterance-sentence. The context is what decides (resolves) the inherent uncertainty resulting from a superposition of states in a learner's competence. L2 competence is therefore a word-by-word affair, where statistical or grammatical forms are measured (that is, decided) on the spot, depending on the context (the preceding and the following words), and given the availability or lack of abstract algorithms. This amounts to saying that quantum SLA treats the L2 competence as a relational concept and that it analyses a learner's intralanguage as the probabilistic counterpart of a learner's IL.

7. Conclusion

In this paper, I tried to describe a novel approach to the study of SLA. I proposed that the quantum-cognition approach can be fruitful for research on language because it incorporates the measurement issue into the developmental theory. The measurement issue stresses the need for a theory in which the measuring system and the measured system are predicted to interact. Quantum cognition theory suggests that this interaction should be modelled in a quantum probabilistic manner rather than in deterministic or classical probabilistic terms. If language development is seen as superposition of different linguistic modalities that can be accessed only probabilistically, one could consider for example that a developmental steady-state is attained not only when a learner can merge items in a phrase properly, but also when a learner can switch effortlessly between statistical and grammatical processing within the array of a single phrase and within a few milliseconds.

References

Acacio de Barros, José & Suppes, Patrick. 2009. Quantum mechanics, interference, and the brain. *Journal of Mathematical Psychology* 53(5): 306–313.

Aerts, Diederik & Aerts, Sven. 1995. Application of quantum statistics to psychological studies of decision processes. *Foundations of Sciences* 1: 361–411.

Aerts, Diederik, Gabora, Liane & Sozzo, Sandro. 2013. Concepts and Their Dynamics. A Quantum-Theoretic Modeling of Human Thought. *Topics in Cognitive Science* 5(4): 737–772.

Aerts, Diederik, Broekaert, Jan, Gabora, Liane and Sozzo, Sandro. 2016. Editorial. Quantum Structure in Cognitive and Social Science. *Frontiers in Psychology*.

doi:10.3389/fpsyg.2016.00577

Avrutin, Sergey. 2012. An Information-theoretic View on Language Processing as the First Step towards Physical Linguistics. *Invited talk at IGG 39*. Verona (Italy), February 23.

Avrutin, Sergey. 2006. The usability of syntax. *Applied Psycholinguistics* 27(1): 43–46.

Avrutin, Sergey. 2006. Weak Syntax. In *Broca's Region*, Yosef Grodzinsky & Katrin Amunts (eds), 49–62. Oxford/New York: Oxford University Press.

Bardovi-Harlig, Kathleen. 2000. *Tense and Aspect in Second Language Acquisition. Form, Meaning, and Use*. Oxford: Wiley-Blackwell Publishing.

Bickerton, Derek. 2009. Recursion. Core of complexity or artifact of analysis? In Syntactic Complexity. *Diachrony, acquisition, neuro-cognition, evolution [Typological Studies in Language 85]*, Talmy Givón & Masayoshi Shibatani, 531–544. Amsterdam/Philadelphia PA: John Benjamins.

Bickerton, Derek. 2003. Symbol and structure. A comprehensive framework for language evolution. *In Language Evolution*, Morten Christiansen & Simon Kirby, 77–93. Oxford/New York: Oxford University Press.

Bickerton, Derek. 2009. Syntax for Non-syntacticians. A Brief Primer. In *Biological Foundations and Origin of Syntax*, Derek Bickerton & Eörs Szathmáry, 3–13. Cambridge MA: The MIT Press.

Bley-Vroman, Robert. 1983. The Comparative Fallacy in Interlanguage Studies. The Case of Systematicity. *Language Learning* 33(1): 1–17.

Bruza, Peter, Kitto, Kirsty, Ramm, Brentyn, Sitbon, Laurianne, Blomberg, Simon & Song, Dawei. 2011. Quantum-like non-separability of concept combinations, emergent associates and abduction. *Logic Journal of IGPL 3* doi: 10.1093/jigpal/jzq049.

Bruza, P., Wang, Z. & Busemeyer, J. 2015. Quantum Cognition: a new theoretical approach to psychology. *Trends in Cognitive Sciences*, 19/7, 383–393.

Busemeyer, J.R. & Wang, Z. 2015. What is Quantum Cognition, and How Is It Applied to Psychology? Current directions in *Psychological Sciences* 24(3), 163–169.

Casasanto, Laura Staum, Hofmeister, Philip & Sag, Ivan. 2010. Understanding acceptability judgments. Additivity and working memory effects. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, Richard Catrambone & Stellan Ohlsson, 224–229. Austin TX: Cognitive Science Society.

Caspi, Tal. 2010. A Dynamic Perspective on Second Language Development. Groningen Dissertation in Linguistics 85. University of Groningen.

Chomsky, Noam. 2002. *On Nature and Language*. Cambridge: Cambridge University Press.

Coecke, Bob, Sadrzadeh, Mehrnoosh and Clarck, Stephen. 2010. Mathematical Foundations for a Compositional Distributional Model of Meaning, arXiv:1003.4394v1

Conte, Elio, Khrennikov, Andrei, Todarello, Orlando, Federici, Antonio Mendolicchio, Leonardo & Zbilut, Joseph. 2009. Mental states follow quantum mechanics during perception and cognition of ambiguous figures. *Open Systems & Information Dynamics* 16: 85–100.

Cook, Vivian. 1999. Going beyond the native speaker in language teaching. *TESOL Quarterly 33*(2): 185–209.

Cook, Vivian. 1997. Monolingual Bias in Second Language Acquisition Research. *Revista Canaria de Estudios Ingleses* 34: 35–50.

Corder, Pit. 1967. The significance of learners' errors. *International Review of Applied Linguistics 5*: 161–170.

Cowart, Wayne. 1997. *Experimental Syntax. Applying Objective Methods to Sentence Judgments*. Thousand Oaks, CA: Sage Publishers.

de Bot, Kees, Lowie, Wander & Verspoor, Marjolijn. 2005. *Second language acquisition. An advanced resource book*. London: Routledge.

de Bot, Kees, Lowie, Wander & Verspoor, Marjolijn. 2007. A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition* 10: 7–21.

de Lange, Joke. 2008. Article Omission in Headlines and Child Language. *A Processing Approach*. Utrecht: LOT.

Ellis, Nick & Cadierno, Teresa. 2009. Constructing a Second Language. Introduction to the Special Section. *Annual Review of Cognitive Linguistics 7*: 111–139.

Fanselow, Gisbert & Frisch, Stefan. 2006. Effects of Processing Difficulty on Judgments of Acceptability. In *Gradience in Grammar. Generative perspectives,* Gisbert Fanselow, Caroline Fery, Matthias Schlesewsky & Ralf Vogel, 291–316. Oxford/New York: Oxford University Press.

Fanselow, Gisbert & Weskott, Thomas. 2009. Scaling issues in the measurement of linguistic acceptability. In *The Fruits of Empirical Linguistics I: Process*, Sam Featherston & Susanne Winkler, 229–246. Berlin/New York: Mouton de Gruyter.

Featherston, Sam. 2009. A scale for measuring well-formedness. Why syntax needs boiling and freezing points. In *The Fruits of Empirical Linguistics I: Process*, Sam Featherston & Susanne Winkler, 47–73. Berlin/New York: Mouton de Gruyter.

Feynman, Richard. 1985. QED. *The Strange Theory of Light and Matter*. Princeton NJ: Princeton University Press.

Gass, Susan & Polio, Charlene. 2014. Data then and data now. In *Interlanguage. Forty years later [Language Learning & Language Teaching 39],* ZhaoHong

Han & Elaine Tarone, 147–171. Amsterdam/Philadelphia PA: John Benjamins.
Giacalone Ramat, Anna & Rastelli, Stefano. 2013. The qualitative analysis of actionality in learner language. In *Research Design and Methodology in Studies on L2 Tense and Aspect*, Rafael Salaberry & Llorenç Comajoan, 391–422. Boston MA/Berlin: de Gruyter Mouton.
Gibson, Edward & Fedorenko, Evelina. 2013. The need for quantitative methods in syntax and semantics research. *Language and Cognitive Processes* 28(1/2): 84–124.
Hampton, James. 1988. Disjunction of natural concepts. *Memory & Cognition* 16: 579–591.
Han, ZhaoHong & Tarone, Elaine. 2014. *Interlanguage. Forty years later. [Language Learning & Language Teaching 39]*, Amsterdam/Philadelphia PA: John Benjamins.
Hawkins, Roger. 2001. *Second Language Syntax. A Generative Introduction*. Oxford: Wiley-Blackwell Publishing.
Heunen, Chris, Sadrzadeh, Mehrnoosh, & Grefenstette, Edward. 2013. *Quantum Physics and Linguistics. A Compositional, Diagrammatic Discourse*. Oxford/New York: Oxford University Press.
Huebner, Thom. 1979. Order-of-Acquisition vs. Dynamic Paradigm. A Comparison of Method in Interlanguage Research. *TESOL Quarterly 13(1)*: 21–28.
Klein, Wolfgang & Perdue, Clive. 1992. Utterance structure. In *Adult Language Acquisition. Cross-Linguistic Perspectives – The Results, Volume II*, Clive Perdue, 3–38. Cambridge: Cambridge University Press.
Lakshmanan, Usha & Selinker, Larry. 2001. Analysing interlanguage: how do we know what learners know? *Second Language Research 17*(1): 393–420.
Lardiere, Donna. 1998. Case and Tense in the 'fossilized' steady state. *Second Language Research 14(1)*: 1–26.
Lardiere, Donna. 2003. Revisiting the comparative fallacy: a reply to Lakshmanan and Selinker, 2001. *Second Language Research 19(2)*: 129–143.
Magrassi, Lorenzo, Aromataris, Giuseppe, Cabrini, Alessandro, Annovazzi-Lodi, Valerio & Moro, Andrea. 2015. Sound representation in higher language areas during language generation. *PNAS 112(6)*: 1868–1873.
Newport, Elissa. 2011. The Modularity Issue in Language Acquisition: A Rapprochment? Comments on Gallistel and Chomsky. *Language Learning and Development 7(4)*, 279–286.
Nooteboom, C., Weermar, F. & Wijnen, F. (eds.) 2002. *Storage and Computation in the Language Faculty*. Dordrecht: Kluver Academic Publishers.
Pallotti, Gabriele. 2007. An Operational Definition of the Emergence Criterion. *Applied Linguistics 28(3)*, 361–382.
Pienemann, Manfred. 1984. Psychological Constraints on the Teachability of

Languages. *Second Language Acquisition 6(2)*: 186–214.

Pothos, Emmanuel & Busemeyer, Jerome. 2013. Can quantum probability provide a new direction for cognitive modeling? *Behavioral and Brain Sciences 36(3)*: 255–274.

Rastelli, Stefano. 2014. *Discontinuity in Second Language Acquisition. The Switch between Statistical and Grammatical Learning*. Bristol: Multilingual Matters.

Rovelli, Carlo. 1996. Relational Quantum Mechanics. *International Journal of Theoretical Physics 35(8)*: 1637–1678.

Schwartz, Jeffrey, Stapp, Henry & Beauregard, Mario. 2005. Quantum physics in neuroscience and psychology: a neurophysical model of mind-brain interaction. *Philosophical Transactions of the Royal Society Biological Sciences 360*: 1309–1327.

Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics 10*: 209–231.

Selinker, Larry. 2014. Interlanguage 40 years on: Three themes from there. In Interlanguage: *Forty years later [Language Learning & Language Teaching 39]*, ZhaoHong Han & Elaine Tarone, 221–246. Amsterdam/Philadelphia PA: John Benjamins.

Shirai, Yasuhiro. 2007. The Aspect Hypothesis, the comparative fallacy and the validity of obligatory context analysis: a reply to Lardiere, 2003. *Second Language Research 23(*1): 51–64.

Slabakova, Roumyana, Leal, Tania & Liskin-Gasparro, Judith. 2014. We Have Moved On: Current Concepts and Positions in Generative SLA. *Applied Linguistics 35(5)*: 601–606.

Sprouse, Jon & Almeida, Diogo. (in press). *Power in acceptability judgment experiments and the reliability of data in syntax*. Accessed August 27, 2015. ling.auf.net/lingbuzz/001362/current.pdf.

Sprouse, Jon, Schütze, Carson & Almeida, Diogo. 2013. A comparison of informal and formal acceptability judgments. *Lingua 134*: 219–248.

Sprouse, Jon, Schütze, Carson & Almeida, Diogo. 2013. The empirical status of data in syntax: A reply to Gibson and Fedorenko. *Language and Cognitive Processes 28(3)*: 222–228.

Tarone, Elaine. 2014. Enduring questions from the Interlanguage Hypothesis. In Interlanguag: *Forty years later [Language Learning & Language Teaching 39]*, ZhaoHong Han & Elaine Tarone, 7–26. Amsterdam/Philadelphia PA: John Benjamins.

Towsend, D.J. & Bever, T.G. 2001. *Sentence Comprehension. The Integration of Habits and Rules*. Cambridge. MIT Press.

van Dijk, Marijn & van Geert, Paul. 2007. Wobbles, Humps and Sudden Jumps: A Case Study of Continuity, Discontinuity and Variability in Early Language

Development. *Infant and Child Development 16*: 7–33.
van Ewijk, Lizet & Avrutin, Sergey. 2010. Article Omission in Dutch Children with SLI: A Processing Approach. *Entropy 12*: 798–817.
van Geert, Paul. 2003. Dynamic Systems approaches and Modeling of Developmental Processes. In *Handbook of Developmental Psychology*, Jaan Valsiner & Kevin Connolly, 640–672. London/Thousand Oaks CA/New Delhi: Sage Publications.
Wang, Zheng & Busemeyer, Jerome. 2015. Reintroducing the concept of complementarity into Psychology. Frontiers in Psychology doi: 10.3389/fpsyg.2015.01822
Yang, C. 2002. *Knowledge and Learning in a Natural Language*. Oxford-New York. Oxford University Press.