Accepted, to appear in K. Kersten, A. Winsler (eds.), *Understanding Variability in Second Language Acquisition, Bilingualism, and Cognition – A Multi-Layered Perspective.* London, New York: Routledge. (peer rev.)

Investigating cognitive-linguistic development in SLA – theoretical and methodical challenges for empirical research[1] (PREPRINT)

Kristin Kersten
Dpt. of English, Hildesheim University
kristin.kersten@uni-hildesheim.de

Werner Greve
Dpt. of Psychology, Hildesheim University
wgreve@uni-hildesheim.de

0. Introduction

It should be general consensus among researchers in the SLA field that the acquisition of a second language cannot be explained without taking a multitude of developmental conditions into account. Possible influencing factors are manifold and highly diverse in nature: Individual factors (such as, e.g., brain structures, working memory, intelligence or motivation), social factors (e.g., parent-child interactions or the family's socioeconomic background), or institutional factors (e.g. a specific school program, the intensity and duration of L2 contact, a submersive or immersive school context), have all been identified as influential for (S)LA by numerous studies (for an overview, see Larsen-Freeman & Cameron 2008, N. Ellis 2007, Dörnyei 2009).

The contribution of these various (potential) influencing factors to predict and explain (S)LA is an empirical question. This endeavor, however, is challenging not only for methodical reasons, but also because the relations and interactions between these potential factors need to be accounted for theoretically: An appropriate *explanation* (here: of SLA) does not only require the identification of potential factors but rather the identification of their causal and conceptual interrelations in the form of a model – in other words, it requires a *theory.*

The first step in theory formation is often to find empirical support for a factor's prognostic or explanatory power, and for its relative or partial weight in relation to other factors; this also includes the question of whether a certain factor does *not* have any predictive value, and could thus be neglected in the analysis or even in the explanation. We do not want to discuss the fundamental questions pertaining to an inductive proceeding of theory development. Rather, we attempt to outline some conceptual problems with analyzing the predictive value of several factors at the same time – which can be done both from an inductive and a deductive point of view.

Such investigations have generally been carried out with the help of multivariate analyses, and lately increasingly with structural equation models and multilevel modeling. Based on the argumentation outlined in this paper, however, we hold that these types of analysis are often not sufficient to discover and/or confirm the *causal* effects of factors on different hierarchical levels, because they remain "blind", so to say, for a number of theoretical problems outlined below. In particular, it has to be taken into account that factors may be related to each other in different *conceptual* relations. If these conceptual relations are disregarded, a prediction or a causal explanation can be incomplete or, at worst, downright wrong.

[1] Although the term "methodical", in everyday language, refers to a "systematic" performance (in a broader sense) of any given task or action, we will use it in a more narrow sense as pertaining to scientific methods (such as multilevel analyses). In distinction to this, we will use the term *methodological* as pertaining to the theoretical (meta-)discussion of research methodology, i.e., to the suitability of certain methods to certain presuppositions (e.g., deductive vs. inductive approaches).

It is, therefore, the goal of this contribution to discuss the various difficulties which are connected to the investigation of influencing factors on different conceptual levels for developmental processes.[2]

0.1 Line of reasoning

To this end, we will discuss six types of problems. The structure of the paper will then follow this line of reasoning depicted here.

(1) Explanation vs. prediction: As a prerequisite of each investigation, studies have to clarify whether they intend to generate a *causal explanation* or whether they just state a statistical *prediction*. While, to add to the confusion, predictions or relations derived from statistical models are often called "explanations", it is crucial for the current line of argumentation to differentiate between both concepts because the conceptual problems outlined in this paper may impair an explanatory approach, while a predictive approach is often not affected by these problems in the same way.

(2) Conceptual layers ("vertical perspective I"): The next section, which we call a "vertical perspective" on the issues in question, discusses the difference between conceptual layers of both internal (individual) and external factors. We use the term "vertical" as referring to the theoretical architecture of inter-factorial relations without a reference to their development over time, because this differentiation becomes relevant even in a synchronic perspective in which factors are assessed at the same time. (The time-dependent perspective will be discussed under (4) and (5), "horizontal levels I and II.) In both the internal and external layers we will differentiate between macro-, meso-, micro-, and nano-levels.

(3) Qualities of relations ("vertical perspective II"): Regarding factors on different conceptual layers, a number of different types of relationships can be distinguished. We call these the different "qualities of relations". In particular, they comprise *causal* and *constitutive* relations. In addition, we will discuss *emergent* relations, which might go beyond these two types of relations. When we describe these different types and the role they play, the individual development of a learner (of a language) is used as *explanandum*, i.e. as the dependent variable to be explained. The careful distinction of these types of relations has consequences both for the explanatory and the predictive perspective of an investigation.

(4) Development across time ("horizontal perspective I"): A further complication is added when the interrelation of factors is regarded "horizontally", i.e. across time. Here, we will discuss time-dependent relations in general. These include in particular mediating and moderating effects.[3] A prediction and, even more so, an explanation will be incomplete if such interactions

[2] The conceptualization of relations as suggested below is the product of yearlong discussions in an interdisciplinary research project on influencing variables in regular and immersive L2 programs. The results of these discussions as presented here helped the authors to disentangle some of the various problems we were trying to come to terms with. We do, however, neither claim our reasoning to be complete nor final; actually, we do not even agree with respect to every single point of the following discussion. It is simply intended to serve as a stimulation for further reflection, and we will be grateful for any hints concerning flaws in our argumentation.

[3] Of course, mediator and moderator relationships can be tested in synchronic cross-sectional analyses as well; the theoretical intention of these hypotheses, however, is a causal one and thus a relation across time.

are not taken into account. The sheer number of possible relations and interactions renders both statistical predictions as well as causal models (and their falsification) complex and confusing; however, this complexity is merely a practical problem, not a theoretical one.

(5) Dynamic processes across time ("horizontal perspective II"): A further problem arises when we take into account that the nature of these interrelations between factors is, as a rule, dynamic. In SLA research, longitudinal processes have, for the longest time, been investigated with the help of linear statistical models. Over the last two decades, however, researchers have started to point out that linear models are not sufficient to depict the complexity of individual developmental processes. One strand of discussion has, therefore, started to focus on describing properties of SLA development across time with the help of *dynamic systems* (DS) theory. Since the statistical modeling of non-linear dynamic systems is comparatively sophisticated because it requires huge data sets and sufficient accuracy of measurement, mathematical modeling of non-linear dynamic systems has rarely been attempted (Dörnyei 2009, for an exception see van Geert 1994, 2008, 2014). While a thorough discussion of these problems goes beyond the scope of this chapter, it has to be pointed out that it is necessary for an explanatory theory as well as for a predictive perspective: Even if a statistical *approximation* of a linear model yields a *practically useful* prediction, a mathematically more correct modeling of a dynamic progression would increase the accuracy of the prediction.[4]

(6) General methodical challenges: These "vertical" and "horizontal" layers of problems are connected to a number of methodical problems of empirical investigations in a more technical sense, which are logically independent from the conceptual problems (2) to (5). While it is not our intention in this paper to discuss the challenges of numerous statistical approaches, we will reflect on three "practical" aspects of empirical investigation used to investigate human development resulting from the previous discussion. Firstly, all constructs used in a model require suitable methods of assessment: For an explanatory as well as for a predictive perspective, each construct has to be validly (and hence reliably) captured. Secondly, certain methodical approaches (e.g., analysis of variance) entail specific methodological assumptions (e.g., an independent, "additive" impact of each factor on the phenomenon to be explained). Thirdly, for the test of causal hypotheses (theories), correlational approaches (in contrast to experimental approaches) are generally weak and, as a rule, insufficient, while they can be sufficient for predictive approaches. In other words: Each approach requires of course an appropriate research design.

0.2 What is an "influencing factor"?

It is helpful for the purpose of this text to introduce/clarify/define our use/definition of "influencing factor" in advance. To this end, Mackie's (1965) concept of *INUS conditions* has proven helpful: He conceptualizes a "cause" as an ***I****nsufficient but* ***N****ecessary part of an* ***U****nnecessary but* ***S****ufficient* constellation: That means that the aspects (variables) in question, such as the intelligence of the child, the climate of the class, the teaching approach of the teacher, the educational background of the parents, the language of the environment, etc., are part of a

---

[4] This also means that the authors do not follow those DS approaches which altogether deny the general possibility of modeling (and thus predicting and explaining) a process based on its assumed "dynamicity": Complexity certainly does not necessarily entail unpredictability (comp. Lenzing 2015).

constellation which, *in its entirety*, is *sufficient* to explain a phenomenon (here: SLA). In this constellation, each of the variables is *necessary* to explain the phenomenon, that is, contributes to the explanation of SLA *beyond* the contribution of all others influencing factors considered – this is exactly why it is as *necessary* condition of *this sufficient* constellation (of factors). Whether or not a particular factor is indeed a necessary part of this constellation remains an empirical question. This does not preclude, however, that *other constellations* of (partly or entirely) different INUS conditions might result in the very same outcome (i.e., language competence); this is why this particular constellation is called *sufficient but unnecessary*.

1. Explanation versus prediction

In the investigation of influencing factors in human development, two (scopes of) intentions need to be differentiated, which are affected by the problems outlined here in different ways: The aim of a scientific *explanation* for a developmental phenomenon in question (here: SLA) is to identify conditions and causes which – in specific combinations or configurations – are *sufficient* to explain the given phenomenon. In other words, the aim of the explanatory perspective is to formulate an (approximately) true (developmental) theory. The basis of this perspective is a scientific understanding of causality which holds that natural phenomena have natural causes which are, in general, accessible for empirical research. [5]

The aim of a *predictive* (prognostic) *approach*, on the other hand, is just to predict the phenomenon (here: language development) in question.[6] A good prediction, or the occurrence of a predicted effect of an intervention, however, does not necessarily imply that the underlying theory is true. Consequences which are effective in practice can also follow from incorrect theories: In medieval times, for instance, pest infections were effectively reduced through isolation, although the underlying theory that infections were transmitted by contact with another person's "aura" was, in all probability, incorrect. Good predictions or successful intervention effects are therefore no "proof" for a theory. For the same reason, successful interventions (e.g., taking an Aspirin tablet often times reduce headaches) do not offer a promising way to infer explanatory (causal) theories (e.g., "headaches are caused by a lack of Aspirin"). On the other hand, incorrect predictions can falsify causal hypotheses (under certain conditions) (Jordan 2004). The predictive and the explanatory approach converge only in the sense that a correct causal explanation leads to a correct prognosis and an effective intervention. (Predictions based on incorrect theories will, in the end, only be partially successful.) Both approaches, however, need to be empirically tested (Shmueli 2010).

In the following, we will first discuss a number of theoretical (conceptual) problems which present challenges especially for an explanatory perspective (sections 2 and 3), followed by

[5] The discussion of causality (as a theoretical concept) is beyond the scope of this paper (for an overview see Mackie 1974). In the following, we will rely on an intuitive understanding of "cause" (for the independent variable) and "effect" (for the dependent variable). Neither will we enter the fundamental epistemological discussion of whether an objective "reality" can be captured by scientific methods at all, let alone the question of whether such a reality actually exists. We assume that there is a reality – without claiming that we will ever be able to fully detect, let alone explain, it. (For those who do not believe in reality, you might want to stop reading right here: We're not writing for you (because, of course, you're not real)).

[6] It is irrelevant for the current discussion whether the prediction is based on data collected in a cross-sectional or longitudinal design (*synchronic* versus *diachronic* prediction). Both approaches test whether and to what degree a dependent variable can be predicted by independent variables. Often times, a predictive perspective has the second aim to change a current state with the help of an intervention, and, thus, to predict that change. However, a longitudinal (diachronic) perspective is, as a rule, superior to cross-sectional designs for developmental processes as the latter are cannot depict the chronological sequence of the events under investigation.

conceptual problems concerning interactions and dynamics across time, which present (different) challenges for both an explanatory and a predictive approach (sections 4 and 5).

2. Developmental conditions on several conceptual layers ("vertical perspective I")

Variables identified in the research literature as relevant for SLA range from factors with narrowly defined functions (such as, e.g., functional components of the working memory, the heart rate of a person) to factors which comprise or imply a number of components (such as the socioeconomic status (SES) of a person, or the culture of a person's environment). We will refer to these different types as "molecular" vs. "molar" variables: "molecular" refers to elements on a lower hierarchical level, while "molar" refers to a higher hierarchical level which is composed of several lower-level constituents – in other words: we use these concepts in a relative manner. In this sense, we will try to outline a hierarchical structure of factors which describes the different levels and their (possible) interrelations in the following discussion.[7] (Ushioda 2015)

For this purpose, we will differentiate between internal (individual and intraindividual, or subpersonal), and external levels (e.g. social structures), and we will refer to these levels as macro-, meso- and micro-level respectively (cf., figure 5 below), and we will add a nano-level for reasons which will become clear later. As the same type of hierarchy can be applied to the external as well as to the internal factors, we will use the index "e" (e.g. $micro_e$) for the external levels, and the index "i" (e.g. $macro_i$) for the internal hierarchy.

2.1 Individual and intraindividual conditions of SLA: Conceptual levels "within" the person

Regarding internal factors, we can differentiate between attributes of a person, as for example a person's motivation or intelligence, and functional or physiological factors within a person, as e.g. functional components of the working memory or neural processes in the CNS.

Individual factors

When investigating the (linguistic) development of an individual, it seems natural to firstly identify other aspects or characteristics of the person in question, such as his or her motivation to learn, intelligence, or language aptitude, among other factors (Dörnyei 2009). Even though these aspects to not describe "the person" in her/his entirety, they are "personal" constructs in the sense that they are ascribed to the person, as opposed to single parts of it: It is the person who is motivated or intelligent, not his or her brain or memory. Language proficiency, which is in the focus of the contributions of this volume, can be understood as a personal variable in this sense. Lower levels are referred to as intraindividual levels (see below).

The differentiation between the individual ("personal") and intraindividual (i.e., "subpersonal") levels combines (and thus possibly confounds) two "kinds" of conceptual leveling. Of course, "the" individual is to be viewed as a complex system of interacting (sub)systems: Organs and muscles, veins and neurons, bones and liquids constitute "the" individual. As these subsystems are systems themselves (cells are complex organs, chemicals in the muscles make

[7] Note that here and in the following, we do not intend to come up with an exhaustive, flawlessly logical taxonomy of levels and factors – this would require a theoretical paper in its own right. The very examples used here show that for such an endeavor much more reflection is needed. Therefore, they are really solely intended as illustrations of a more abstract line of argumentation.

them move, etc.), "the" individual can be seen as a nested hierarchy of subsystems. Certainly, the conceptual relations between these different hierarchical layers are therefore confronted with the very same conceptual difficulties as the differentiation between different social (interpersonal) layers (e.g., family – neighborhood – city – country – culture) which we will discuss in the following section (2.2) more thoroughly. Viewed from this angle, then, this multilevel differentiation just continues "within" the person. We call it "within" just because our interest focuses on phenomena (in particular: language) that can be observed at the level of persons only (neither brains nor societies talk). As a consequence, "internal" vs. "external" is used throughout this paper in relation to a person who acquires a language. This boundary is, of course, arbitrary as it only reflects our present focus of research. Certainly, the arguments discussed here could be applied to other segmentations as well (e.g., "within" a body cell vs. "outside" of it but still within the human body). A more complete theory of explanatory layers would have to take this relativity into account.

This leads to a second perspective at the individual layer. The fact that only persons (not brains) talk reminds us that certain phenomena "exist" only given a certain "phenomenological" perspective. Language, intentions, consciousness, etc. are restricted to the level of persons (hence the term "personal"). If we investigate these *mental* phenomena, we have to apply what Dennett (1990) has termed the "intentional stance". With "intentional stance", he means a perspective in which we perceive human beings as intentional beings with mental phenomena, and in which we describe these phenomena with the help of a language which uses mental terms (the "mental idiom"). Wanting, perceiving, thinking, talking: These activities can only be carried out by mental beings, i.e., persons.[8] Conceptually, the explanation of mental phenomena entails the usage of the mental idiom. Referring only to "functional" or "physical" concepts leaves the gap between functional and physical entities on the one hand and mental phenomena on the other hand unabridged (Bieri 1981, quoted in Fuchs 2018). The decisive question, however, is whether this unabridged gap is more than a different linguistic way of describing the phenomenon (i.e., more than switching between "stances", in Dennett's words). In other words: The question is: If we want to explain mental phenomena, is it necessary to presuppose that "mental entities/processes" (such as the consciousness or will) actually exist, or is it possible to explain them just by referring to physical entities/processes (such as electro-chemical transmission in neurons)?

We certainly cannot discuss, not even aptly describe the mind-body problem (this is exactly what we are touching upon with this terminological distinction) in this paper; however, one has to keep in mind (well, yes …) that this is a theoretical problem to be treated whenever we attempt to explain mental phenomena (such as thinking or talking). This conceptual gap has to be taken into account even if we attempt to subdivide a certain phenomenon into components or subsystems. For instance, intelligence does not consist of neurons, even if one component of intelligence is memory; and memory – among us human beings – is realized in a brain consisting of neurons. This is why it is necessary (for an explanatory approach) to differentiate between various conceptual levels even within the individual.

[8] This perspective is, of course, the one which we take intuitively in everyday life – but for the sake of completeness it has to be mentioned that the debate of whether it denotes "real" phenomena *different* from, say, neurological ones or just expresses a view on these phenomena has been unresolved for more than two millennia (Metzinger 2013).

Intraindividual factors

Intraindividual factors are not ascribed to the entire person but denote functional components (units) of larger (molar) factors. Processes carried out by the working memory are, for example, constitutive for the person's intelligence. For the sake of organization, we will conceptualize individual factors such as motivation, aptitude, or intelligence on the $macro_i$/$meso_i$-level, and intraindividual factors such as working memory (WM), phonological short-term memory (PSTM), phonological awareness on the $micro_i$-level.[9] Smaller sub-components such as neurons or neural complexes are referred to as $nano_i$-level.[10]

This distinction illustrates the complication already addressed above. On the one hand, "the person" is sort of an aggregation of different (sub(sub))systems, but on the other hand, a person can "do things" (can be conceptualized as doing things) which the subsystems are not capable of alone. Moreover, these personal activities may have causal consequences for the subsystems of the system that performs these activities: To use an illustration, components of the language faculty ($micro_i$-level) cannot, for instance, decide to drink a bottle of red wine for dinner; only the person ($macro_i$-level) can do so – however, a relatively large amount of alcoholic intake can indeed have an influence on the functioning of the language faculty during dinner conversations.

2.2 External conditions on SLA: Conceptual levels "outside" of the person

As mentioned above, the corresponding problem can be described on the level of "external" conditions for SLA. Enumerating these external (context) factors presents a complex challenge not only in the face of the sheer number of possible influences but also because of the innumerable interactions between them. In the following, we will conceptualize the (external) $micro_e$-level as the immediate interaction of the individual with its direct environment (e.g., with a communication partner, in a didactic intervention in the classroom, but also with a (linguistic) medium such as a book or a computer). The $meso_e$-level describes the environment for these $micro_e$-level interactions (such as the culture of a family, the rules in a school etc.), whereas the $macro_e$-level refers to the wider social context such as the laws of the given country, the cultural background of the language community (which language/s?), or informal social norms for behavior (rules of politeness) (cf. figure 5; see Bronfenbrenner 2005 for a classical approach, Lerner 2013 for a more differentiated extension).

Such a complex hierarchy of a multitude of factors prompts the question of whether and how they interrelate with each other, and how this conceptual hierarchy can affect the prediction and/or explanation of a dependent variable. We will try to systematize some important aspects pertaining to this question in the following section.

---

[9] PSTM is, for example, part of the working memory, which, in itself, is a part of the memory system, which constitutes part of intelligence … and so on. As stated above, the conceptualization presented here is used simply in an abstract, theoretical way to provide a linguistic tool for the authors to illustrate the characteristics and differentiate between different types of factors.

[10] It seems worth mentioning once more that the inclusion of the $nano_i$-level into this hierarchy (i.e., to conceptualize neurons as a sublevel of, say, motivation or even memory) ignores the mind-body problem mentioned above. (This ignorance is statistically harmless, but theoretically inappropriate.) Of course, this hierarchical structure repeats itself on many different levels. The neural level itself can, again, be subdivided into different cellular – or even atomic – components.

3. Qualities of relations ("vertical perspective II")

3.1 Factors on different levels influence the dependent variable independently of one another

One option to capture the influence of several factors (statistically modeled as *independent variables*, henceforth IV) on a developmental phenomenon proceeds on the assumption that they impact the dependent variable (DV) independently of one another – that is to say in an *additive* way. This approach does not take into account, however, on which level these factors are located. Rather, in multivariate analyses (e.g. multiple regression analysis) the factors under investigation are analyzed with respect to their respective predictive contribution, that is, their partial contribution, i.e., the contribution of each particular factor given the contribution of the other factors in this analysis (figure 1).

Although in these analyses the partial contribution takes the statistical contribution of the *other* variables (located on the same or on different conceptual levels) into account, their *conceptual* relation is ignored. For instance, the individual's subjective motivation to learn might contribute statistically to the prediction of his/her language development *beyond* the socio-economic status of the family – but since the child is a part of the family and motivation to learn might be seen as one aspect or indicator for SES, the child's motivation can be seen as conceptually related to SES. Consequently, it would be misleading (in an explanatory approach) just to add the contribution of motivation to learn to the contribution of the SES. The finding that the "addition" of both variables in the analysis shows a unique (additional) statistical contribution might just be the consequence of an insufficient assessment of (all facets of) SES. We will return to this problem in the following sections.

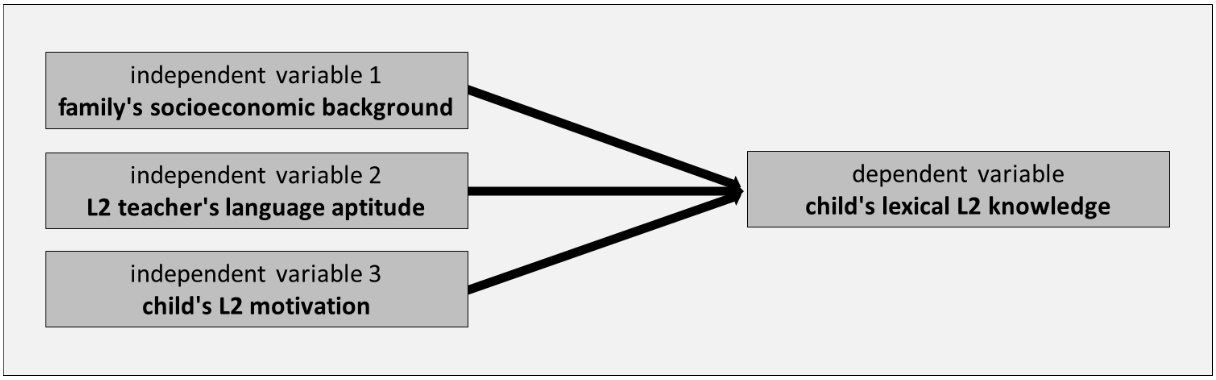

Fig. 1: Examples of independent variables on different hierarchical levels which are conceptualized as influencing the dependent variable independently of each other. (Example of analysis: multiple regression)

In some analyses, such IVs are grouped according to larger concepts and analyzed in a stepwise manner (figure 2). As this kind of grouping still assumes an independent influence of each IV on the DV, this does not change the disregard of (possible) conceptual relations between variables and levels (actually, of the mere fact that there *is* more than one level).

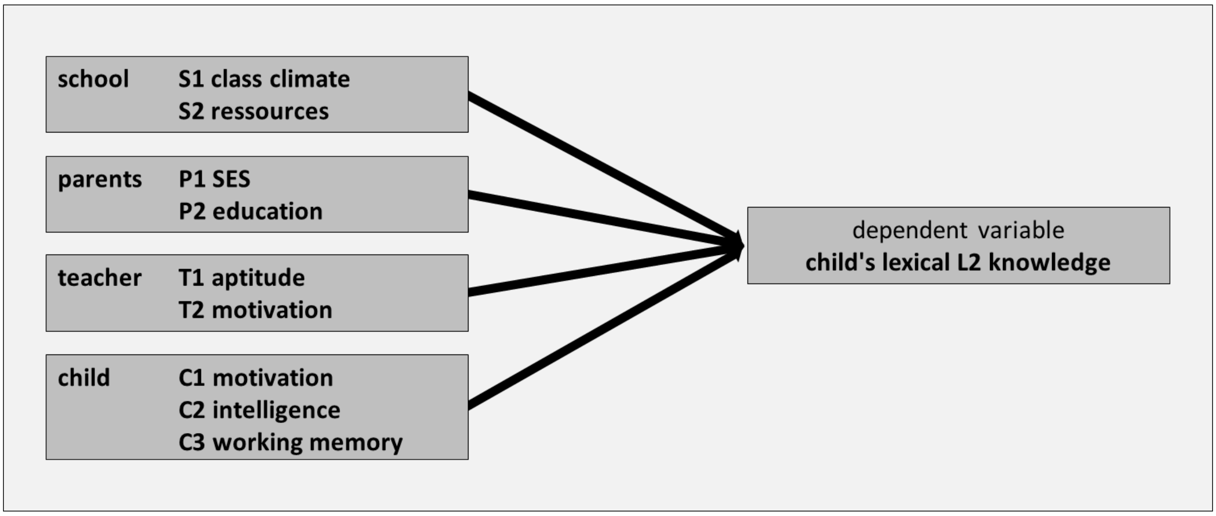

Fig. 2: Examples of independent variables on different hierarchical levels which are grouped and analyzed in a stepwise manner, and which are conceptualized as influencing the dependent variable independently of each other. (Example of analysis: stepwise multiple regression)

Nowadays, the fact that variables are “located” on different conceptual levels is increasingly accounted for with the help of multilevel analyses (Cunnings 2012, Windzio 2008). Such analyses *statistically* take the hierarchy of levels into account by incorporating intra-class-correlations (i.e. if one factor is similar for a number of different individuals, e.g., the teacher’s language aptitude is equal for all pupils of the class). As will become clear in the following sections, however, they also remain *theoretically* insensitive to the *reasons* for intra-class-correlations, that is for the (causal or conceptual) quality of the interrelations between variables or levels.

To sum up, this type of conceptualization is based the presupposition that the impact of the IVs is *independent* of other factors, *direct*, and *additive*.[11] In other words, this type of analysis does not differentiate between internal and external levels, nor between $\text{nano}_{i/e}$-, $\text{mikro}_{i/e}$-, $\text{meso}_{i/e}$- or $\text{makro}_{i/e}$-levels, nor does it take into account different conceptual relations between these factors and/or levels. If an analysis pursues an *explanatory* approach, we claim that simple multivariate analyses of factors on different conceptual levels (e.g., the impact of the school climate, of the pupil’s L2 motivation, of the pupil’s working memory capacity) on the pupil’s language competence do not do justice to the complexity of relations between variables, even if the variables are *measured independently* of each other (see section 3.2.1). Predictive goals, on the other hand, can often be achieved; as always, the interpretation of the data and the conclusions drawn from them need to be carefully limited to what is actually investigated in the analysis.

3.2 Factors on different levels do not influence the dependent variable independently of one another

If the influencing factors (IVs) are to be seen as interrelated, we can think of several types of relations. Firstly, higher level IVs can exert their impact on a DV *via* lower-level IVs (or vice versa); “via” means that the impact under investigation is a mediator relation (cf. 3.2.1).[12] It is

[11] In many analyses of this type, of course, *interactions* of (some of) these variables are taken into account as well (we will discuss this aspect in section 5).

[12] Certainly, there are many other forms of interrelations. For instance, an IV (meso_e) might have an influence on *the influence* of another IV (micro_e) on the DV (moderator effect, see section 4, figure 7).

important to note that this relation (a *causal* dependence) presupposes a *conceptual independence* of the factors (levels) involved as conceptually related concepts cannot have a causal relationship (see below).

Secondly, lower-level factors can (conceptually) partly or fully *constitute* a higher level factor (cf. 3.2.2). In this latter case, the dependency between factors (levels) is a conceptual one and thus precludes a causal relation. It might be worth stressing that causal and constitutive relationships are sometimes confused because not only causal, but also conceptual relationships imply correct *predictions*: We can, for example, predict that all bachelors we will meet tomorrow will be unmarried – because this is what bachelors are. However, being unmarried does not *cause* the status of a bachelor – it rather *entails* it, and thus denotes a conceptual relationship. With respect to relationships across levels, it would be equally wrong to say that the living room, bathroom, and bedroom *cause* an apartment – they *constitute* it. This is what Ryle (1949) referred to as *category mistake*. In the same way it is incorrect to claim that the phonological loop, the visuospatial sketchpad, etc., *cause* the working memory – the molar structure of the working memory is *constituted* by its molecular components. Assuming a causal relationship in a multivariate analysis in the sense that the "impact" of the phonological loop on the working memory is measured would thus be theoretically misleading. The following sections will elaborate on these distinctions.

3.2.1 A factor on a higher level influences factors on a lower level, which influence the dependent variable

In figure 3, examples of indirect (mediator) relations are depicted. Note that these relationships can occur both between and within levels. For the sake of illustration, in both cases a *complete* mediator effect is shown in contrast to a partial mediation, which will certainly be the more typical case. (For instance, the intelligence of a person might influence both his or her L2 motivation and language aptitude. Since L2 aptitude influences the language motivation as well (Dörnyei 2010), the effect of intelligence on language motivation is partially mediated by L2 aptitude.)

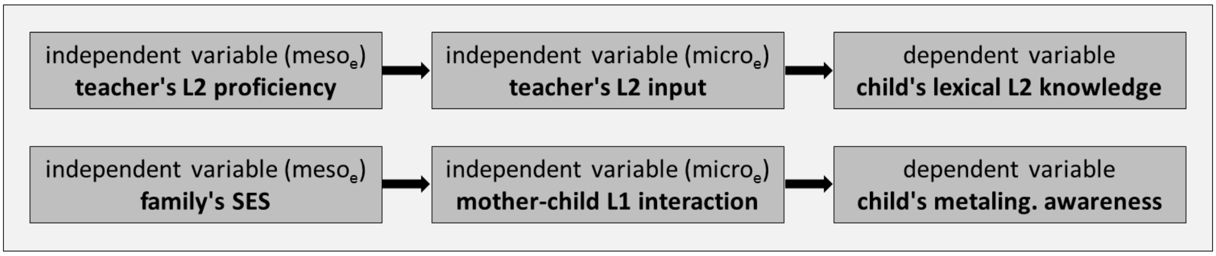

Fig. 3: Influence of a higher order independent variable on DV is mediated by a lower order independent variable (two examples). (Example of analysis: mediator analysis via multiple regressions)

As mentioned above, in all cases the *conceptual* independence of the IV's involved is presupposed (if the analysis pursues an explanatory purpose). It is important, moreover, to distinguish the conceptual (in)depedence of two factors from the (in)dependence of the assessment (method) of these factors: If two factors are confounded in the assessment, e.g., by using some identical items for both factors, any correlation that was found in the analysis would be (partly) artificially created: If we find a correlation between working memory capacity and intelligence, and if the assessment of intelligence includes the assessment of the working memory capacity, this correlation is not based on an actual relation but is (partly) created artificially through the selection of the measuring instrument.

But even if the (method of) assessment of two constructs is independent, e.g. by using one specific test for working memory and a completely different one which does *not* include a direct assessment of working memory for intelligence, a conceptual relation might still exist. For instance, it could be argued theoretically that (almost) all subtests of any given intelligence tests entail an *indirect* assessment of working memory capacity (since WM is a *component of* intelligence). In other words: Even if we *measure* factors independently of each other, this does not imply *conceptual independence*. Even if we check, for instance, in the official record whether Peter is unmarried (one type of measurement), and then assess his status as a bachelor by asking several of his friends to report on his behavior, e.g., his changing relationships with many women (another type of measurement), this does not alter the fact that “being unmarried” and “being a bachelor” are conceptually related (even if both measures would (predictably) not correlate perfectly). This conceptual relation would preclude a causal relation.

Of course, the conceptual independence of the variables (concepts) included in an analysis cannot be taken for granted, in particular if the analysis comprises variables on different levels. Actually, causal relationships (such as mediator relationships) are easy to confuse with conceptual relationships, in particular with a “container relationship” between variables (section 3.2.2).

3.2.2 A factor on a higher level is constituted by factors on a lower level

While in the cases discussed so far high-level “molar” variables are treated as independent variables with a direct impact on the developmental phenomenon (figure 4A below, compare also figure 1, SES: macro$_e$-IV and L2 motivation: meso(macro)$_i$-IV) and are ranked as equally important, such contributions of (independent) variables on different levels are further complicated by the fact that often these independent variables on different levels are *not conceptually independent* of each other, which precludes both independent direct effects on the DV (figure 1) and an indirect (mediator) effect (figure 3).

It is important to differentiate between two constellations here: Often, a molar variable is *constituted* by a number of molecular variables, and has no independent effect of its own (figure 4B). We will refer to this type of molar variables as *container variables*. The other option is that a molar container variable, even though it is constituted by a number of lower level variables, exerts a *causal effect of its own* which surpasses the added effect of the molecular ones (see figure 6 below). This phenomenon is referred to as *emergence* (e.g., Noordhof 2010). Both possibilities will be discussed in the following sections.

3.2.2.1 Container variables without independent effect on the dependent variable

The problem that relations between different conceptual levels can obscure the (causal) explanation of the dependent variable is most obvious when elements on the higher level are *constituted by* elements of the lower level. A prototypical example for a container variable constituted by a number of lower level variables in developmental research is age. Age can be seen as textbook predictor for L1 competence (at the age of 2 months, it is commonly extremely restricted, while at the age of 20, it’s usually very good). However, in the explanation of various developmental L1 acquisition trajectories, age functions as an “umbrella term” for more specific processes, such as the developmental state of the working memory, the developing effectiveness of the individual's executive functions, the change of complexity of the neural networks of the brain, etc.

Actually, it is not only exceptionally the case that variables of different levels are considered in one analysis without taking their relationships into account – even in cases of obvious constitution (container variables). Another example for this is the (routinely pursued) "control" for gender in social science. Obviously, "gender" is a container comprising a huge variety of heterogeneous elements such as social expectations, physical strength, self-esteem, to name but a few. If in a given analysis in order to "explain", say, aggressive behavior all these aspects (strength, self-esteem, expectations, gender, age) are introduced (comp. figure 1), it is less clear whether the authors have taken the conceptual relations into account. figure 4B depicts such constitutive relationships.

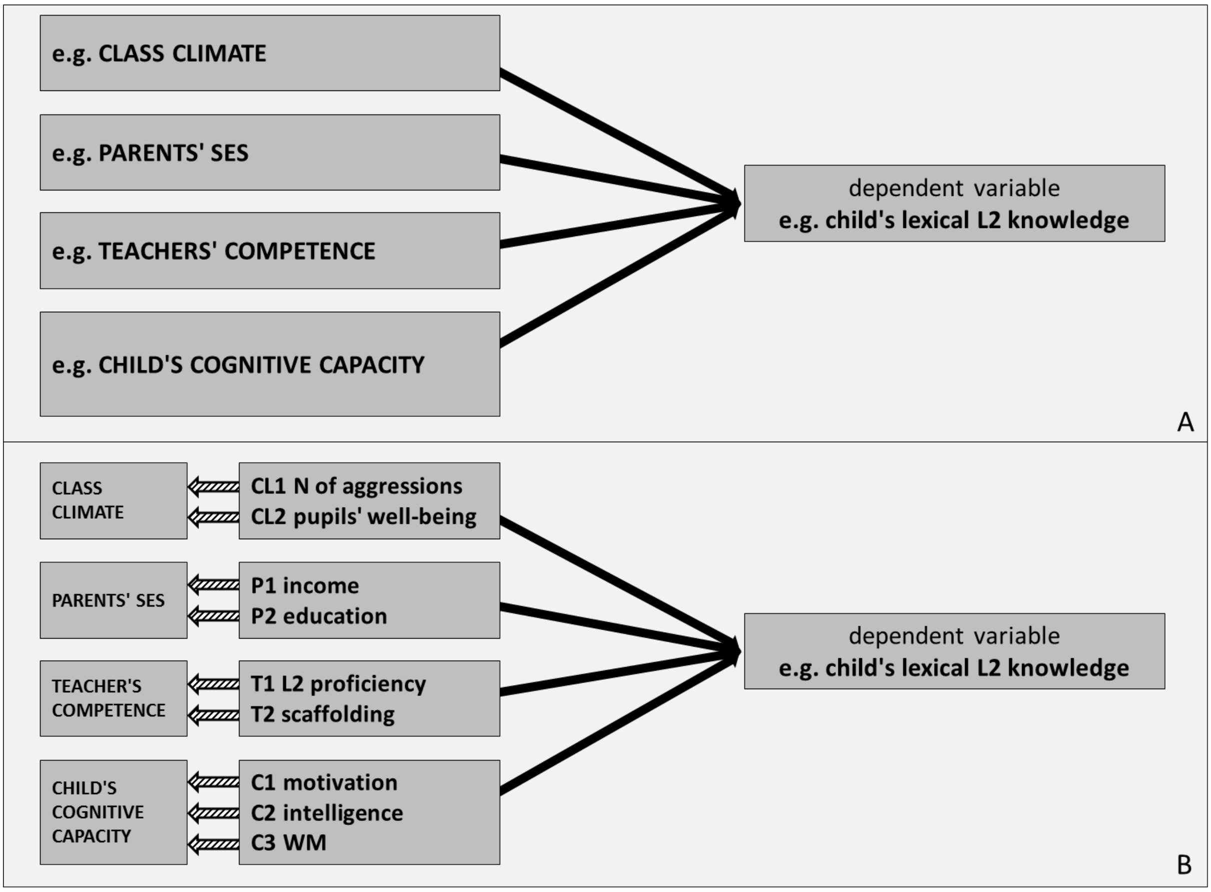

Fig. 4: Examples of higher order (molar) variables (A) which are constituted by lower order (molecular) variables, but which do not exert an independent influence on DV (B). The relationship between the molar variables (here: "container variables") and the molecular variables is that of *conceptual entailment* rather than of a causal (mediator) relationship: The respective container variable is *constituted* by the molecular elements (thus, the relation between the molar and the molecular level is conceptual, not causal; this is why the direction of the shaded arrows (4B) is reversed). This conceptual entailment can exist on all levels. (Example of analysis: multiple regression, multilevel analysis, structural equation model (but see section 6.1))

As the age-example suggests, containers can, and usually will, be large. Certainly, "class climate" ($meso_e$-level) comprises – apart from individual perceptions (emotions and cognitions) of many persons – social rules and, in particular, social practices, i.e. concrete social micro-interactions, which the individual not only experiences but actively shapes, influences, or co-creates.[13] That means that not only (aspects of) other IVs are entailed in the IV “class climate” but (at least partly) even aspects of the DV (i.e., the pupils’ behavior or competence).

Unfortunately, it is not always that easy to recognize containers as mere conceptual umbrellas. An instructive sample case in point is intelligence. Several theoretical approaches (traditions) conceptualize intelligence differently (for an overview on different approaches see, e.g., Gardner 2012). For instance, according to a Thurstonian view (first outlined in Thurstone 1938), intelligence is to be seen as comprising several *independent* basic components (such as verbal fluency, conclusive thinking, memory, etc.); in other words: intelligence is just the *container* of these components. From a Spearmanian perspective, however, the successful performance of any given intelligence-related task is – beyond a task-specific competence – explained by a *common competence* (termed the g[general]-factor of intelligence by Spearman 1904). In the latter theoretical tradition, (g-)intelligence, thus, is mainly a factor of its own potency, so to speak, *not* just an umbrella (container) term for independent factors.

Obviously, then, it depends on the theoretical perspective whether or not a certain variable is seen (conceptualized) as container. Whether or not the particular theory (claiming or denying the container-status of a certain construct) is true is, of course, an empirical question (although most difficult to solve). We will return to this aspect in the following section.

Such constitutive relations as described in these examples, in which complex levels of variables are embedded in a higher one, can be conceptualized on all different levels and with manifold interrelations. Thus, they often occur in *complex nested* structures of variables. These *nested structures* of hierarchical levels (Linck & Cunnings 2015) are depicted in figure 5.

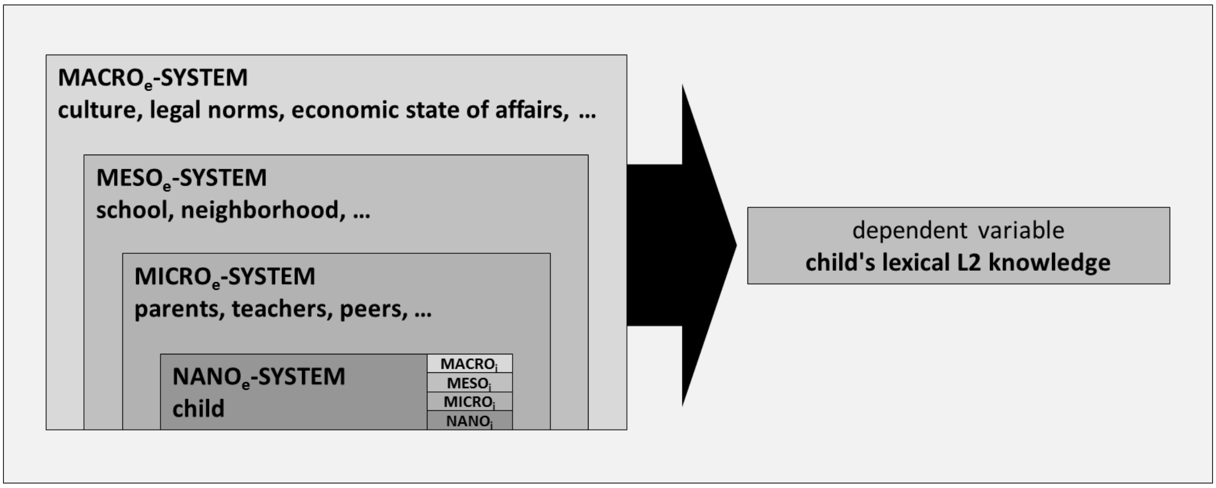

Fig. 5: Nested structural depiction of variables of different levels. (Example of analysis: multilevel analysis)

For instance, the parents’ socioeconomic status (SES) is a highly complex configuration of two people, their personal and intrapersonal structures, their belongings, their cognitive and material resources, their social interactions both with each other and with many other people

[13] This means also that molar factors such as class climate include not just molecular constituents (e.g., class climate *consists of* individual perceptions of the climate in the class etc.) but real interactions between individuals. For instance, a teacher might react differently to a motivated student than to an inattentive one, and, dependent on that, the quality of her input and her interactions might change (comp. de Bot 2008, Ushioda 2015). The teacher’s reactions, in turn, are part of (i.e., co-constitute) the class climate. As a consequence, if we are to analyze the impact of the class climate, the teacher’s behavior, and the pupils’ motivation (IVs) on the pupils’ competence (DV), class climate (partly) comprises many other variables (including the DV) and their interactions. This argument entails, however, that the quality of these interactions have to be understood. We will discuss this “horizontal perspective” in greater detail in section 4.

(family, friends, colleagues) etc. – which means that SES is constituted by these structures which it entails; likewise, the class climate is not just constituted by the "elements" referred to above but also by the child's behavior, the teacher's behavior, the school's climate, etc., and each of these constituting structures are, again, composed of a variety of lower-level constituting structures (the cognitive resources of the parents, for instance, by their intelligence, their education, their motivation etc., which, again, are constituted by their intraindividual (neural) processes; the teacher's behavior depends on a large variety of different factors, as well.) Of course, there are manifold interactions between these different factors changing over time, which will be discussed in chapter 4 below.

The insight that there are nested dependencies among and across levels has led to the development of more complex statistical models such as multilevel analyses or mixed method approaches (compare e.g. Linck & Cunnings 2015, Windzio 2008). For instance, if one molar variable affects several molecular elements in the same manner (e.g., the language competence of the teacher is *identical* for all pupils in the classroom), this will result in high intra-class-correlations. This is what multilevel analyses in particular take into account. However, although these approaches do account for mutual interdependencies, even such complex multilevel approaches are "blind" to the problem of the *conceptual nature of the relations* between the different levels, i.e., they do not depict the *reasons* for a given intra-class-correlation, which can be a conceptual or a causal one.

Consequences for statistical models

The idea of a container variable assumes that, if all relevant factors it contains are identified, the container variable, e.g. "age", would not have an additional effect of its own – for the exact reason that it was just an umbrella term.

If the molar construct (variable) functions just as a container (conceptual umbrella) an explanatory analysis must *either* calculate the *global* influence of the container (figure 4A), *or the singular effects* of its lower-level components (figure 4B). As mentioned above, it can be an important and helpful strategy to decompose (and, hence, eliminate) the higher-level structure (in particular if an explanatory approach is pursued). This holds both for factors outside and within the person. For instance, instead of measuring the global working memory capacity of an L2 learner, it could be useful to analyze the components of the WM (such as the phonological loop, the visuospatial sketchpad, etc.) separately to arrive at a more precise understanding of mechanisms of how working memory components affect the acquisition of L2. To give another example, working memory, neural networks and executive functions can be introduced into the analysis *in order to* "eliminate" the age effect (rather than to "investigate" whether age or working memory is the "stronger" predictor). It is our understanding that this procedure, i.e. the replacement of global molar predictors (e.g., SES, migration background, etc.) by their more specific components, leads to a deeper understanding of developmental phenomena and should, therefore, be the goal of theoretically informed empirical (explanatory) research. It is a heuristically fruitful procedure if these conceptual relationships are taken into account.

This procedure, however, may prove problematic in several respects. If we analyzed, in a single multivariate analysis (comp. figure 1), the variables class climate, connectivity of pupils, number of aggressive actions per year, parents' SES, parents' education, parents' income, teachers' educational competence, teachers' language aptitude, variety of media applied in the classroom, the child's cognitive capacity, working memory capacity and lexicon in order to eliminate those variables that do not significantly contribute to the prediction, we would ignore that *both conceptual* and *causal* reasons might explain why a certain predictor (IV) would show an effect or not.

A conceptual reason can be described, for example, as follows: If the distinction of constituency is disregarded, if for example the influence of SES is analyzed *concurrently* with the L1 lexicon of the mother in a multivariate analysis (compare figure 1), and if the L1 lexicon of the mother can be understood as a component of the family's SES, the molar variable will always "win" (i.e. have a stronger impact than the lower-level variable): Firstly, because SES contains many other component variables whose power is added up and included in the analysis (if they are not separately identified), resulting in a much higher predictive power; and secondly, because the factor L1 lexicon is actually measured *twice* in the analysis: once as a separate variable, and a second time *conceptually "hidden" within* the container of SES. For the same reason, to give another example, if the working memory capacity ($meso_i$-level) and the capacity of the phonological loop ($micro_i$-level) are measured in the same multivariate analysis, the working memory would probably absorb all variance and remain as the only predictor in the analysis.[14] If this happens with a number of loosely defined, fuzzy constructs and their constituents within a statistical analysis, the predictive impact of the container, here SES or working memory, is blown out of proportion, and it is not possible to arrive at an exact explanation: It would be incorrect, indeed, to conclude that the L1 lexicon of the mother or the phonological loop had no causal contribution to SLA.

In many studies, however, even after all of these attempts to eliminate a molar factor and to replace it by its constituent molecular factors, often some variance of the DV remains unexplained. In such cases, however, it is unclear whether a molecular constituting factor remained undetected in the analysis at hand, or whether the conceptual relationship between the molar and the molecular structures is not only that of constitution, but is of a different nature. We will come back to this possibility in the following section when we discuss the phenomenon of *emergence*.

Another problem could be that an effect remains invisible – and the IV would be excluded/show no effect – because there is not enough variation of the factor within a specific set of data. We will return to this issue in section 6.2.

Consequences for prediction and explanation

The differentiation between cause and constitution does not necessarily need to be taken into account in a predictive approach: The size of the living room may statistically predict the size of the apartment, in the same way as the capacity of the working memory may statistically predict the intelligence. If it is the goal to identify factors which statistically (cross-sectionally or longitudinally) best predict, for example, the (language) competence of a child, it might not be necessary for that particular purpose to disentangle the different relations between these factors: If factors are only regarded competingly the most potent predictor will "win".

[14] For a brief discussion of the problem that the distributions (variances) of the variables in a given set of data have implications for the possible results *independent* from their actual causal relevance see section 6.2.

From an explanatory perspective, however, the equal treatment of factors on different levels which might account for their statistical interaction but which disregards their conceptual relationship, risks confounding a constitutive relation with a causal relation – and thus contaminates the theoretical explanation (Smedslund 2016). Studies which regard factors on different conceptual levels as equal and which do not analyze their conceptual relations will remain theoretically incoherent, and their interpretation will become misleading or, at worst, incorrect. It is therefore essential for an appropriate theoretical *explanation* of a developmental phenomenon to carry out a conceptual analysis of these relations in advance.

Thus, if we are to explain (not just to predict) a certain developmental phenomenon, and if we include molar and molecular variables at different conceptual social and individual levels in our analysis, we have to be aware of the "nested" order of these variables in order to avoid explanatory mistakes. It follows that we have to have a more systematically ordered theory *in advance*. If we are to analyze a given set of variables in order to explain L2 motivation, we have first to "locate" them at their proper conceptual level, e.g., according to the nested multilevel structure introduced above (see figure 5). Secondly, we have to clarify their conceptual relations *before* we pursue any analysis using these variables and discuss its results.

3.2.2.2 Emergence

Even if a molar variable is constituted by its (molecular) elements, it can have causal effects of its own that *go beyond the singular (added) effects of all of its constituents* (figure 6). This phenomenon (or, to be cautious, the claim) that "the whole is greater than the sum of its parts", which is often attributed to an Aristotelian treatise, is called *emergence*[15] (Noordhof 2010). Can, to take up the example discussed in the previous section, the class climate be something *more* than the sum of perceptions of the climate of the class members?

It is at least possible that the class climate has a causal impact on an individual learner's language development *above and beyond of what this child perceives* of the class climate.[16] While it is not self-evident that "the climate" actually is more than (or something beyond) the singular perceptions of the class' pupils, it is easier for us to imagine such an emergent effect on the personal level: It is *the person* who talks, who is motivated, who has a certain intelligence and a state of consciousness, and not the person's brain – even though we believe that neurophysiological processes constitute what we call "motivation", "language", or "consciousness". The person, her thoughts, her motivation, her language, can therefore be regarded as *emergent* phenomena which seem to be more than just neural activities in the brain. In other words: A person can do things which each of his components can't do, but which are nevertheless relevant for these components (a person can decide to drink a bottle of red wine – an action which will not only influence his decisions but also the functioning of his central executive, his motor control, etc.).

The central idea of emergence, thus, means that in addition to the two possibilities discussed so far (i.e., (a) a mediator relationship, and (b) a constitutive relationship between hierarchical factors), there might be a third possibility: (c) the higher-order molar structure could have *emergent* characteristics *with an independent causal effect that goes beyond the impact*

[15] Note that the term *emergence* in this context is not identical with the emergence of a linguistic structure in the sense as it is used in the framework of Processability Theory (Pienemann 1998, Pallotti 2007).
[16] Again, this difficulty is not related to the methods of assessment. Even if we assess the class climate via the teacher's rating, it may be (conceptually) nothing beyond the sum of the pupils' perceptions.

of its constituents (figure 6). The phenomenon of emergence thus assumes that the impact of the whole is systematically different from the impact of the sum of its part, *without* any additional elements or influences from the outside.

Fig. 6: Higher order (molar) variables exert an influence on the DV which goes beyond the added impact of the lower order (molecular) variables which it entails *(emergence)*. (We cannot insert an example of analysis for the emergence phenomenon since to establish an emergent relationship is a question in its own right. In any statistical model with unexplained variance it is unclear whether a molecular constituting factor remained undetected, or whether the conceptual relationship between the molar and the molecular structures is not that of constitution but of emergence.)

A thorough theoretical discussion of emergence is out of the scope of this paper. With regard to the conceptual relation between personal vs. subpersonal phenomena (see above, section 3.2.2.1), the topic has been discussed as an essential facet of the mind-body problem for over 2000 years (Is the personal realm really different from the subpersonal, physical (i.e., physiological) realm, do personal phenomena *emerge* from physical processes, or are personal phenomena reducible to physical ones? If both realms are fundamentally different: Do personal phenomena exert causal influences on physical phenomena?) – with manifold variations and trends, but without a consensus. Moreover, the phenomenon of emergence (if it actually *is* a phenomenon) is, as our present discussion illustrates, not confined to the relationship between mind and body, but refers (possibly) to all cross-level relationships and explanations.

Ultimately, the existence of emergent phenomena is – in a broader sense – an empirical question: They either exist, or they do not exist. At present, however, it is undecidable in almost any particular case whether or not the phenomenon at hand actually is an emergent factor of its own right, as it were. If they exist, however, such phenomena need to be accounted for within the conceptual definitions of influencing variables.

Consequences for prediction and explanation

Apart from the question of whether emergence actually exists, the possible impact of emergent factors is of particular importance for explanatory approaches. If a container such as parents' SES is found to have an impact beyond other variables on the dependent variable, how can this impact be interpreted? Do we find it *because there is an emergent relation*, i.e. because SES has an additional independent contribution, or do we only witness an addition of the effect of its constituents? If we find a contribution of SES that goes beyond the addition of what we measure concerning its constituents, how can we be sure that the reason for this is an emergent phenomenon and not only our failure to identify all relevant constituents – let alone measure them adequately? This question is crucial for the theoretical modeling of developmental phenomena in the explanatory perspective, because its answer tells us whether we can interpret the relations between the different levels as *causal effects*.

In contrast to section 3.2.2.1, however, this problem does concern the predictive perspective as well, at least to a certain degree. On the one hand, statistical prognoses can of course model the effect of the working memory *and* the SES *and* the phonological loop as (statistically) independent of each other. This statistical prognosis, however, is, to stress this again, only a necessary but not a sufficient condition for a *causal explanation*. On the other hand, *if* a certain factor actually *is* an emergent one, any prediction leaving it out (resting on the assumption that it is just a container) will be less successful.

If our line of reasoning does not go astray it is thus a central demand for theory construction (in the explanatory perspective) to discuss the conceptual classification of the involved factors.

4. Development across time ("horizontal perspective I")

So far, we have discussed the conceptual relation of factors on different hierarchical levels in what we have called the "vertical" perspective. Another complication of theoretical modeling is, however, that such relations manifest themselves, of course, *over time*. Ultimately, they can best be modeled in longitudinal designs, which can test chronological sequences (e.g., in "cross-lagged-designs"). Although cross-sectional designs – which are regularly used for pragmatic reasons or to avoid repeated measure effects – often use the same terminology (independent variables are said to "predict" or "explain" the dependent one/s), causal interpretations of synchronic relations are difficult and vulnerable. At best, cross sectional designs can *test* causal relations postulated by theoretical models: If a particular relationship between factors is not found in a synchronic measurement, a longitudinal effect would be unlikely.

In section 3.2.2.1, we have already touched upon the fact that molar variables (e.g., class climate) may comprise not only molecular factors (e.g., pupils' emotions or perceptions), but real social ($micro_e$) interactions as well. As a consequence, the interactions between factors *over time* have to be taken into account – both for predictive and explanatory purposes. Interrelations such as depicted in figure 7 are what we are looking at in a "horizontal" perspective.

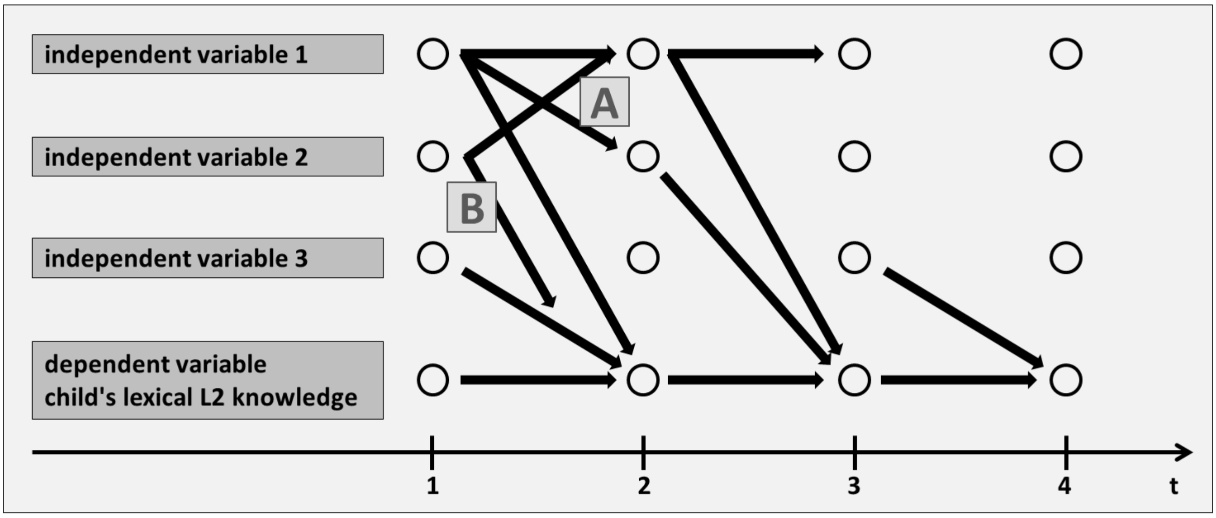

Fig. 7: The interplay of different variables over time. (Example of analysis: multiple regression, structural equation, multilevel analysis)

A: *Mediator effect* – an independent $variable_1$ influences another independent $variable_2$, which influences the dependent variable ($variable_2$ mediates the relation between $variable_1$ and DV)

B: *Moderator effect* – an independent $variable_2$ influences (moderates) the impact of another independent $variable_3$ on the dependent variable

We can theoretically differentiate between several types of impact effects across time. Particularly important for our discussion are mediator effects (already introduced in section 3.2.1 above, compare figure 3) and moderator effects. A mediator relationship identifies sequences of effects: V1 influences V2, and V2 influences the DV (figure 7, relation A). In this case, V2 is the mediator of the effect of V1 on the DV. Again, in a predictive approach, the acknowledgement of this mediator effect is not necessary: The prediction of DV by V1 holds independently of the causal (explanatory) V1-V2-DV-chain.

In contrast, a moderator relation denotes the "qualification" of an effect (the effect on an effect, so to say), i.e. the influence of a factor *(moderator)* on the *effect* of another factor (IV) on the dependent variable (comp. figure 7, relation B).[17] To illustrate this: The effect of a certain teaching strategy on the L2A of a student could be dependent on his attention or motivation to actually follow the classroom interaction.

Certainly, more interrelations than these are possible and possibly important. Beyond self-stabilizing effects (V1 at t1 predicts V1 at t2), more complicated relations (e.g., recursive effects) might occur. For instance, V1 might influence V2, which has an effect on V3's relation to V1 (e.g., the teacher influences the pupil's motivation, which, in turn, increases the effect of the teaching materials on the pupils' L2A, which, in turn, increases the teacher's motivation to increase the pupil's motivation, etc.). Such types of effects are, of course, possible on and between all (intra)individual and social levels.

Obviously, then, beyond the acknowledgement of the conceptual relations of the factors involved, any coherent explanation needs to take into account the interactions of these factors across time. In fact, this is what any theoretical (explanatory) model aims at. In a predictive perspective, however, the acknowledgement of mediator effects does not necessarily increase the predictive value of the predictors, while the acknowledgement of a moderator effect might increase (the precision of) a prediction.

5. Dynamic processes across time ("horizontal perspective II")

When focusing on the interaction between different factors across time, another complication is the (often) *dynamic nature* of this interaction. In the academic SLA discourse, it has recently been argued with increasing rigor that language development needs to be described as a dynamic system (DS) with all characteristics pertaining to it (comp. de Bot 2008, 2015, de Bot et al. 2007, Dörnyei 2009, 2010, N. Ellis 2007, Larsen-Freeman 2007, Larsen-Freeman & Cameron 2008, Lenzing 2015, to name but a few). These characteristics are controversially debated in SLA especially with regard to non-linearity, stability, and predictability vs. non-predictability of DS. Indeed, the characteristics of DS, as we will discuss below, pertain to both the explanatory and the predictive perspective on influencing factors in SLA.

Mathematically, the modeling of the dynamical nature of cross-temporal relations between factors is independent of the conceptual relations between these factors (as discussed in the section 3). However, the claim that a relation between factors explaining SLA actually is

[17] Obviously, a simple moderator effect can only capture a small part of reality: In fact, there are countless interactions (i.e. moderator effects of a second, third, and even higher order) between countless variables. The "academic enterprise" boldly presupposes that a theoretical model with a manageable number of factors and their interactions allows for an adequate empirically testable depiction of reality.

a dynamical one presupposes, of course, that this very relation is not a mere conceptual but rather a causal one.[18]

5.1 Dynamic systems and SLA

Language, speech communities, individuals, and sub-components of individuals such as their brains, have all been described as DS (N. Ellis 2007), “from the social level to the neurological levels” (Larsen-Freeman & Cameron 2008:201). As a consequence of the discussion of the previous sections (in particular section 3), any modeling of DS has to take into account the conceptual level of the variables included into the model. A DS-model is as “blind” for the conceptual level of its variables as any other model (such as regression analyses or structural equation models). Hence, the problems discussed above can equally occur and have thus to be taken into account within this framework as well.

A controversy in SLA derives from different views on language development which “range from approaches that see L2 development as a highly variable and nonpredictable process […] to those that view L2 development as both dynamic and rule-governed” (Lenzing 2015:91). Proponents of the first position claim, for instance, that universalities (i.e., developmental stages and sequences) do not exist and hence cannot be aptly modeled, but rather occur as artificial consequences of the models or methods applied (Lenzing 2015:100). According to this perspective, developmental sequences (such as language acquisition) are highly, and unreducible, individualistic (Lenzing 2015). Although according to the mathematical definition DS are deterministic, one important implication of such a DS approach in SLA is that for the explanation of variation and change the focus shifts from variables (causal factors) to processes: “… it is not the possible causes but the degree of variability in itself (which may include systematic, free and unsystematic variation) that is taken as providing insight in the developmental process” (de Bot et al. 2007:53).

Other positions focus more strongly on the regularities of patterns observed in SLA. Certainly, in a DS any variation always occurs within the constraints of the system. Beyond the formal constraints of a given *model* (here, of course, is everything “fixed”), any real process to be modeled in the DS is restricted by certain constraints: Our physical environment, for example, is constrained by gravity – thus, variation in the paths of falling raindrops in a DS limited by gravity will not include infinite upward movement. Accordingly, with respect to (second) language development, communication would break down if everything would vary (at the same time). Thus, communicative rules that govern speech communities need to be stable to a sufficiently high degree to ensure mutual comprehension (albeit non-static and subject to change). Likewise, language production of humans is constrained (among other things) by the structure of the articulatory apparatus – thus, variation in the production of sounds will not include sounds beyond the possible spectrum.

According to this view, variation (within the learner’s language or between learners) does not occur randomly but follows certain laws and is constrained, at least to some extent, by universal developmental pathways a learner takes (and has taken in the past, limiting the possible trajectories in the future): The nature and number of variable linguistic items at a certain stage of interlanguage is not random in the absolute sense of the word, but is restricted by

[18] Of course, DS as *mathematical* models are conceptual models. Within the SLA discourse, however, the hypothesis to be discussed claims that the empirical dynamics of (second) language acquisition (development) can be aptly described (albeit not always aptly measured) by a DS model. *This* claim entails a causal relationship between the factors involved.

the learner's cognitive apparatus:[19] "[I]ndividual learner variation occurs *within the overall constraints of the developing L2 system*" (Lenzing 2015:100, italics by the authors), with "the range of structural options [being] … restricted by the processing constraints of a given stage of development delineated in Hypothesis Space" (Lenzing 2015:113; cf. Plaza-Pust 2008:266).

Each of these theoretical perspectives has been discussed within the framework of DS.[20] A DS theory of SLA, as any other theory of SLA, however, needs to be able to account for both phenomena, variability as well as systematic regularity, including phases or states of stability (de Bot et al. 2007:53). The question within our current topic arises as to what the dynamic nature of developmental processes implies for a predictive and explanatory perspective. We will try to shed some light on these questions with the following reflections on characteristics of DS.

5.2 Terminological distinctions

On (non-)linearity and determinism

DS are systems with two or (usually) more elements that are, in a systematic way, related to each other such that both the value of these variables and their interrelation change from step to step – i.e., over time (for a similar definition see Dörnyei 2010:260).[21] DS can be linear or non-linear. Linear DS have been well researched in mathematics, and statistical methods in SLA include predominantly linear models, while the same is not true for non-linear systems. It can be assumed that SLA, like many if not most other developmental phenomena, are non-linear in nature (de Bot 2008, Dörnyei 2009). All DS – as mathematical models – are, by definition, "deterministic": The knowledge of any state of the system (theoretically) entails the knowledge of any other state, both prospectively and retrospectively, so to say (cf. Lenzing 2015:97).

On stability

Non-linear DS can be chaotic or non-chaotic. The term 'chaotic' in DS theory means that trajectories from two similar initial states can lead to very different future states. Thus, small initial changes may lead to large differences in effects (*butterfly effect*, comp. Verspoor 2015:38), while in non-chaotic (sub)systems small changes have small effects.[22] Thus, if a chaotic DS is applied to an empirical phenomenon, future states seem almost unpredictable given the limited accuracy of measurement (see below). DS applied to human phenomena (such as

---

[19] A linguistic theory which takes this into account is, for example, Processability Theory (Pienemann 1998; cf. Lenzing 2015; comp. also Plaza-Pust 2008 for an application of DS to a UG perspective).

[20] Many publications on DS in SLA have had a strong theoretical focus, while empirical studies concerning these discussions are currently still relatively scarce (Schwartzhaupt 2014).

[21] For the remainder of this section, we will restrict our discussion to discrete (as opposed to continuous) dynamical systems for several reasons: In discrete systems, the measurement of its variables takes place at discrete points in time and is not carried out continually (even in dense longitudinal studies). At present, truly continuous measurements are empirically impossible (although they would sometimes be more appropriate from a theoretical point of view).

It is worth noting, that even if we restrict our models to discrete DS, the "temporal resolution", i.e., the narrowness of observation, the frequency of measurement, the magnitude of the focused time-frame, remains arbitrary: Do we model a dynamic developmental process over seconds or minutes, over hours or days, over weeks or months, over years or decades, or over generations? The acquisition of a language consists of years of influences, these in turn consists of weeks of schooling (among other things), these weeks consist of days, of seconds etc. (de Bot 2015, Larsen-Freeman & Cameron 2008). We cannot delve into the details of this problem, but the nestedness of processes (within processes etc.) seems to horizontally mirror the vertical nestedness of factor levels discussed above (section 3).

[22] For a more thorough overview see Lenzing (2015).

SLA) deal with *finite* phenomena, while DS in mathematical theories are not limited by time or space and, hence, are conceptualized as infinite. Such stable end states towards which a system evolves are referred to, in mathematic terms, as *attractors* (Dörnyei 2009, 2010, Hiver 2015, Plaza-Pust 2008). Such attractors are robust states which resist change and which are recursively stable (i.e., they just (re-)produce their own state, as it were) so that the state of the system remains unchanged once the attractor state is reached. Attractors are, thus, "a relatively stable state of a system … [which] remains fixed for a long time (or forever, for that matter). An attractor state is insensitive to (small) perturbations" (van Geert 2007:47; see Ruhland & van Geert 1998 for an application to language development).

In the context of SLA, or of any limited human development for that matter, however, it would be inappropriate to talk about indefinite stable states. In the social sciences it thus makes sense to define DS terms such as *stability* and *attractor states* as finite concepts, and regard DS as comprising *temporary* stable and instable states. In this latter sense, attractor states are discussed within SLA discourse (see, e.g. Chan et al. 2015:256, Larsen-Freeman & Cameron 2008). Here, some authors use the term "*attractor states"* as indicating "stable *tendencies*, *solutions* or *outcomes* for dynamic systems" (Hiver 2015:20), or as "preferred patterns" that resist change over a sufficiently long period of time (Dörnyei 2010). Dörnyei 2010:261) suggests that higher-order variables of individual differences such as aptitude, emotional intelligence or possible selves "can be seen as powerful attractors that act as stabilizing forces" (Dörnyei 2010:261), and Plaza-Pust, trying to reconcile UG and DS theories, proposes that UG parameter settings function as attractors and limit variability, while variation, according to her, is an indicator for a grammatical system in an unstable state in between different attractors (Plaza-Pust 2008:255f).

To our understanding the terminology used in the SLA discourse deviates in some respects from the technical terminology in mathematical DS theories. Vvan Geert 2007:47) calls this "the possible tension between formal and informal dynamic systems models", elaborating that "a formal model has a specific mathematical format … [while] an informal model is one that applies certain dynamic systems properties by analogy or similarity." In order to reconcile, for instance, UG and DS theories, however, it might be acceptable to apply DS terminology in a metaphorical manner as an illustrating characterization of the phenomenon even though the linguistic discussion is still far from being able to model language development geometrically (comp. van Geert 2007:49). However, the implications of this conceptual distinction are, unfortunately, out of the scope of this paper and beyond the expertise of the authors.

5.3 Consequences for prediction and explanation of influencing factors in DS

Concerning the question of whether explanation or prediction in the horizontal dynamic interplay of factors is possible, we have then to differentiate between instable phases of the system, and stable ones (in the "informal" usage of these terms, as outlined above). While *theoretically possible* for all parts of a deterministic system, prediction of *applied* DS will *in practice* be limited and depend on the characteristics of the specific constellation of empirical factors (i.e., on the appropriateness of the description of the empirical constellation *as a DS*), and in particular on the precision of the measurement instruments available. Thus, they will be particularly difficult for instable parts and chaotic subsystems, as small differences in the initial state can lead to very different pathways and end states: If our knowledge of initial states is incomplete, and if measurement instruments are not precise enough to capture the relevant small initial differences (both of which is highly probable for the complex study of human development), prediction becomes a difficult endeavor. Moreover, any modeling of a highly complex reality (such as the empirical conditions of language development or SLA) through a DS will necessarily remain incomplete: The number of variables to be integrated into this model will be limited both by the boundaries of our knowledge and by the boundaries of the limited possibility to integrate a large number of variables into one model (given the sample sizes available). This incompleteness further restricts the precision of any empirical prediction.

For non-chaotic subsystems and for stable parts of non-linear DS, however, this might be different. Firstly, imprecise measurement of initial differences is much less relevant than for chaotic systems as they lead to similar consequences (compare Lenzing 2015:99; figure 2). While empirical measurement of initial states is always limited by our research instruments, the precision of prediction will be much higher and closer to the theoretical prediction than for chaotic parts. Secondly, as stable parts endure over time (at least to some extent in human development), the chance for imprecise instruments to capture the main characteristics of the observed state is higher than in instable parts and chaotic subsystems. (It is a question which needs to be empirically corroborated (or falsified) which parts within a DS under investigation are stable, and which parts are not.) We would therefore assume that prediction is easier (or: more precise) for the stable aspects of a DS within the practical constraints of a given research instrument, but less so for chaotic DS.

In summary based on this discussion, the cognitive (and other) influences on and constraints for language development can *theoretically* be described, as is being attempted in many different frameworks throughout SLA research. Such theories need to systematize and order what at a first glance seemed to look like random variation (reduce the “degree of randomness”, so to say). It is theoretically possible to claim these elements to be predictable, as argued above, *within the constraints of the context, the state and nature of the DS, and the limitations of the research instruments*. On the other hand and in the same vein, one can make a case for *practical* unpredictability of variation during an instable phase of restructuring in which small and undetectable differences lead to large differences which cannot be accounted for empirically. How to model such complex dynamic processes in a predictive perspective statistically is, however, far beyond our expertise. (Some researchers have proposed alternative, e.g. qualitative methodological approaches (Larsen-Freeman & Cameron 2008, Dörnyei 2009, 2011, Chan et al. 2015), but this discussion is out of the scope of this paper.)

From an explanatory point of view, however, it is probably essential to incorporate a DS approach into future theories of SLA. If the assumption that the interaction between factors within and between different levels is (as a rule) dynamic by nature should proof defendable, no explanation ignoring that can be true. Even complex (multi-level) models of interactions between a multitude of variables (see figure 7) will remain insufficient to truly depict the complex, dynamic nature of (second) language development.

6. Methodical challenges: Measurement, variance and design

Independent of the conceptual problems discussed so far, an empirical investigation faces a number of methodical problems in a narrower, more technical sense. In the following sections we will restrict ourselves to two basic practical aspects of empirical studies (of cognitive-linguistic factors in SLA) pertaining to the issues in question: Problems of measurement, statistical premises, and research designs. We will not focus, however, on technical challenges of specific research instruments (e.g. for the measurement of neural processes) or specific statistical analyses (e.g. technical presuppositions of multilevel analyses).

6.1 Problems of measurement

Each construct which is part of a theoretical model requires an adequate model of measurement: In both vertical and horizontal perspectives, the construct needs to be reliably and validly captured, i.e. either directly measured or latently estimated (i.e. estimated from a number of manifest indicators).

In a first step, as argued above, this is a theoretical challenge: What "is" SES, what is intelligence, the climate of a class, etc.? What is meant by these terms, what aspects do they include, and which aspects are no part of it? Is the parents' education, for instance, constitutive for their SES, or only correlatively associated to it? Etc. These conceptualizations need to be clarified theoretically before a measurement theory is applied.

In a second step, these theoretical constructs need to be operationalized. In the empirical practice, that means that certain aspects or components of the construct need to be selected, along with an adequate measuring instrument. More often than not due to state of affairs in research and to practical restrictions, we will not be able to assess the construct to be investigated entirely (without measurement errors), but will have to contend ourselves with procedures of estimation (e.g., by proxy-variables). For instance, the number of books might be useful as a proxy of the economic resources and the education level of a household. Naturally, each proxy requires an empirical corroboration of validity, which ultimately is based on the theoretical conceptualization of the construct in question.

Independent of whether a macro-construct such as SES serves just as a container or is assumed to possess an emergent causal function there are distinct problems connected with the indicators which are selected to represent the macro-construct more or less directly. An indicator could, for instance, theoretically indicate other constructs in addition to the selected one: The number of books, which is often used as in indicator of SES, could, e.g., also represent the *cultural background* of a family or a social group ("Do members read?"). This would have repercussions for the validity of the indicator since "books" would not only measure SES und would thus contain partly invalid (but still possibly predictively relevant) information. In a statistical analysis, this could lead to methodical artefacts, if e.g. the indicator "number of books" could then be more powerful than a separate indicator for *culture*, so that the influence of the latter is underestimated or becomes invisible. These aspects would render the theoretical interpretation of the results difficult (or impossible), but this only results from the operationalization (not from the theoretical problems discussed in this chapter).

The current methodical answer to this challenge is structural equation modeling (SEM). The central point of this approach is to *estimate* (the value of) a latent factor by calculating several manifest variables (in order to extract, in a way, measurement errors and other factors that reduce the latent factor's reliability and, hence validity). Certainly, SEM is helpful and necessary both for the predictive and the explanatory paradigm (independent of the number and type of relation/s of the levels involved in the study). SEM could (and sometimes will) be used to model "container"-relationships as well (the container can be modeled as the "latent" variable "behind" its "manifest" constituents). We have not referred to this approach in our discussion because all examples discussed so far (from figure 1 to figure 7, actually) can (and perhaps should) be calculated in a SEM frame *even though* a SEM approach is as insensitive (and, hence, no solution) for the problems discussed in this chapter as other statistical models mentioned so far.

Independent of these indicators and their problems of validity, the types of tools used to capture the respective variables differ widely (for theoretical and/or practical reasons): self-report data (interviews, questionnaires), physiological measurements of brain structures or processes, functional subpersonal data (measurements of memory capacity, semantic associations, etc.), observations (field notes, semi-structured and structured observations, e.g. of classroom interactions), or a combination of several of the above (for example, in order to measure emotional reactions, self-reports, physiological measurements, and observation of behavior are often used in conjunction). Hence, data are available in numerous qualitatively different formats, with a different degree of objectivity, and different types of errors, all of which have a different effect on reliability. To give an example: "Social desirability" does not play a role when measuring reaction times, but it certainly does in interviews or questionnaires.

Since the quality ("success") both of predictions and explanations depend on the quality of measurement of the variables included in a given analysis, predictions will often fail not because of a wrong theory but because of insufficiently precise measurement. Although this problem is a serious (perhaps even unsolvable) obstacle for research, it is independent from the problems discussed in this chapter.

6.2 Sources of variance

Throughout this chapter we have followed the (tacit) assumption that dependent variables can be predicted and/or explained by identifying the (partial) contributions of several IVs (and their interactions), i.e. that factors have independent contributions to the variance of a DV (this assumption is also entailed in the INUS conception of a causal factor). This, however, i.e. the central idea that the contribution of each factor is *independent* and that their causal contributions are *added up*, is not necessarily correct. It is just a theoretical assumption (presupposition) and certainly not self-evident. We cannot delve into the epistemological debate of whether this idea of independent and additional causality (INUS) is defendable. It seems worth noting, however, that certain statistical approaches (resting on analyses of variances, as do almost all approaches mentioned in this paper, albeit in different ways) entails methodological and epistemological presuppositions.

Moreover, the identification of a relative (partial) weight of a certain variable in a given (statistical) prediction depends on the distribution of variances in the actual sample. As a consequence, the actual (causal) contribution of a certain variable may remain invisible (i.e., the IV would show no effect) if there is not enough variation of the factor within a specific set of data. An illustrative example would be the question of whether bipedalism in humans is genetically explainable. As there will be very little variation with respect to the relevant part of the genotype in humans, but a broad variety of "environmental" influences (e.g., the loss of a leg due to injuries, accidents, etc.), analyses of variances will find (almost) no genetic effect on bipedalism. Yet, the conclusion that bipedalism is not genetically influenced would be plain wrong. Analyses of variance can, thus, only detect the contribution of variance which is present in a data set. However, this problem is certainly independent from the theoretical problems discussed in this chapter (actually, it is no theoretical problem at all, just a methodical one).

6.3 Implications for design

In an explanatory approach, the actual aim of modeling is to investigate the causal impact on the developmental phenomenon in question (here: SLA), in other words, to *explain* the developmental stages or trajectories. Since this impact (causality) cannot be measured directly, however, its investigation is not only a question of how to capture a certain variable (see above) but of how to set up the research design. Some limitations of certain research designs (e.g., the impossibility to control certain variables such as education or gender in an experimental manner) can be reduced by using specific longitudinal analyses, e.g. cross-lagged designs. In addition to the operationalization of the variables it is, thus, essential to clarify which *type of research design* is necessary to test the postulated causal relations in the explanatory model.

The “royal road” to test causal relations is the randomized experiment. This is often possible for the study of causal effects on the subpersonal level (e.g., the impact of cognitive factors on the (development of) L2 attainment). When conducted well, experiments increase the internal validity of the design and, thereby, support the causal (explanatory) interpretation of the effect under investigation. For practical purposes, however, it has certain shortcomings. For one thing, an increase in internal validity is as a rule accompanied by a decrease of external validity: It is precisely the systematic control of conditions which leads to an artificial constellation since it is impossible to include in an experiment the complex factors and their interactions from real contexts.

Secondly, the number of systematically controllable (and thus measurable) variables is inevitably limited in an experimental design. (Randomization controls *all* other variables, of course, but at the same time it renders their impact “invisible”.) If a large number of variables are to be investigated at the same time this cannot be modeled realistically in an experimental study. The alternative is to control them, not experimentally, but statistically, that is, to use a correlational design that just measures (instead of manipulates) the IVs to be investigated. This, however, leads to the confounding of variables which is inevitable in quasi-experimental designs: Children, for instance, are seldom distributed randomly across different types of EFL programs – their parents’ motivation, their parents’ own linguistic and educational background, their mobility, the schools’ catchment area, and may other (potentially effective) aspects will be confounded. Testing complex causal theories, therefore, yields considerable problems of design. For predictions, and even for interventions, this does not necessarily need to be an obstacle (precisely because interventions can be successful even if they are based on (partly) incorrect theories).

As a consequence, a combination or triangulation of several research designs might be helpful to capture the complexity of the phenomenon under investigation (i.e., (second) language development): For the investigation of a particular causal process (“*How* does this particular IV influence the DV?”), an experimental approach might be most appropriate; to gain a broader picture of the dynamic interplay of a multitude of factors in the explanation of the

DV, "correlational" approaches (with limited or nor manipulation of the IVs) will be more suitable (see also Dörnyei 2011 for the suggestions of an alternative qualitative research program).

7. Conclusion

The development of a learner's linguistic systems is intricately intertwined with the individual cognitive and emotional development, and is embedded in his or her individual developmental process in general and in the interactions with the environment. The conduct and interpretation of empirical research to explain cognitive and linguistic development, however, proves problematic in many respects and for different reasons. In this contribution, we tried to systematize two important groups of problems: First, the differentiation of conceptual layers into which factors are to be categorized, and the differentiation of their mutual relations within and across layers, are a necessary prerequisite of any theoretically coherent (causal) explanation, but not necessarily of a successful (statistical) prediction. We referred to these types of problems as the "vertical" perspective. Second, it is a necessary prerequisite both for coherent and true explanations (theories) and for successful predictions (and, hence, interventions) to understand the dynamic interaction of conditions across time. We called this the "horizontal" perspective. While predictions are *theoretically* possible for such dynamic systems, they are practically limited for several reasons, in particular for instable parts and chaotic subsystems of the individual's developmental trajectories.

For some of these problems, statistical models provide solutions (for instance, structural equation models enhance the exactness of measurement, multilevel models are able to capture several facets of a nested structure of variables) if they employ a sufficiently differentiated measurement approach. Yet, a statistical analysis of a model cannot test its *theoretical* coherence: Predictions, even if successful, do not "verify" a theory. Consequently, even advanced statistical methods, which may lead to successful cross-sectional and/or longitudinal predictions, are, in some respects, "blind" for differences between types of relations (i.e., causal vs. logical) between conceptual layers, and thus may lead to misleading or even wrong interpretations. In other words: The adequacy of a given theory cannot be detected empirically.

This asks for greater accuracy in theory development, for more investment into the development of dynamical modeling of developmental trajectories, as well as for a precise construction of the measurement applied to each variable in the model to be tested. This holds especially for the explanatory perspective, but, if the complexity of (dynamical) interactions between developmental conditions is not sufficiently taken into account, both prediction and explanation will, as a rule, remain insufficient or fail altogether. If our arguments do not go wrong, a more careful consideration of the theoretical premises of our research on the one hand and progress in the application and refinement of (dynamic) models for developmental processes will advance the understanding of SLA in particular and individual development in general. Finally, this might also widen the potential of interventions to facilitate the progress along developmental pathways for each individual learner.

References

Bronfenbrenner, U. (2005). *Making human beings human: Bioecological perspectives on human development*. Thousand Oaks: Sage.

Chan et al. (2015). Learner archetypes and signature dynamics in the language classroom: A retrodictive qualitative modelling approach to studying L2 motivation. In: Dörnyei, Z.; MacIntyre, P.D.; Henry, A. (ed.), *Motivational dynamics in language learning*. Bristol: Multilingual Matters, 238-259.

Cunnings (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research* 28/3, 369-382.

de Bot et al. (2007). A dynamic view as a complementary perspective. *Bilingualism: Language and Cognition* 10/01, 51-55.

de Bot (2008). Second language development as a dynamic process. *The Modern Language Journal* 92/2, 166-178.

de Bot (2015). Rates of change: Timescales in second language development. In: Dörnyei, Z.; MacIntyre, P.D.; Henry, A. (ed.), *Motivational dynamics in language learning*. Bristol: Macmillan, 29-37.

Dennett (1990). *The intentional stance*. Cambridge, Mass.: MIT Press.

Dörnyei (2009). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning* 59/1, 230-248.

Dörnyei (2010). The relationship between language aptitude and language learning motivation: Individual differences from a dynamic systems perspective. In: Macaro, E. (ed.), *Continuum companion to second language acquisition*. London, New York: Continuum, 247-267.

Dörnyei (2011). Researching complex dynamic systems: 'Retrodictive qualitative modelling' in the language classroom. *Language Teaching*, 1-12.

Ellis (2007). Dynamic systems and SLA: The wood and the trees. *Bilingualism: Language and Cognition* 10/01, 23.

Fuchs (2018). *Ecology of the brain: The phenomenology and biology of the embodied mind*. Oxford: Oxford University Press.

Gardner (2012). *Theories of intelligence*. Oxford: Oxford University Press.

Hiver (2015). Attractor states. In: Dörnyei, Z.; MacIntyre, P.D.; Henry, A. (ed.), *Motivational dynamics in language learning*. Bristol: Macmillan, 20-28.

Jordan (2004). *Theory construction in second language acquisition*. Amsterdam: Benjamins.

Larsen-Freeman (2007). On the complementarity of Chaos/Complexity Theory and Dynamic Systems Theory in understanding the second language acquisition process. *Bilingualism: Language and Cognition* 10/01, 35.

Larsen-Freeman & Cameron (2008). Research methodology on language development from a complex systems perspective. *The Modern Language Journal* 92/2, 200-213.

Lenzing (2015). Exploring regularities and dynamic systems in L2 development. *Language Learning* 65/1, 89-122.

Lerner (2013). *Concepts and Theories of Human Development*. Hoboken: Taylor and Francis.

Linck & Cunnings (2015). The utility and application of mixed-effects models in second language research. *Language Learning* 65/S1, 185-207.

Mackie (1965). Causes and conditions. *American Philosophical Quarterly* 2/4, 245-264.

Mackie (1974). *The cement of the universe: A study of causation*. Oxford: Clarendon Press.

Metzinger (2013). *Grundkurs Philosphie des Geistes (Band 2): Das Leib-Seele Problem*. Paderborn: Mentis-Verl.

Noordhof (2010). Emergent causation and property causation. In: Macdonald, C.; Macdonald, G. (ed.), *Emergence in mind*. New York: Oxford University Press, 69-99.

Pallotti (2007). An operational definition of the emergence criterion. *Applied Linguistics* 28/3, 361-382.

Pienemann (1998). *Language processing and second language de velopment: Processability theory*. Amsterdam: John Benjamins.

Plaza-Pust (2008). Dynamic Systems Theory and Universal Grammar: Holding up a turbulent mirror to development in grammars. *The Modern Language Journal* 92/2, 250-269.

Ruhland & van Geert (1998). Jumping into syntax: Transitions in the development of closed class words. *British Journal of Developmental Psychology* 16/1, 65-95.

Ryle (1949). *The concept of mind*. London: Hutchinson.

Schwartzhaupt (2014). Conceiving language as a complex adaptive system: A problem for SLA researchers? *Signo* 38/65, 261-272.

Shmueli (2010). To explain or to predict? *Statistical Science* 25/3, 289-310.

Smedslund (2016). Why psychology cannot be an empirical ccience. *Integrative psychological & behavioral science* 50/2, 185-195.

Spearman (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology* 15/2, 201-292.

Thurstone (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Ushioda (2015). Context and complex dynamic systems theory. In: Dörnyei, Z.; MacIntyre, P.D.; Henry, A. (ed.), *Motivational dynamics in language learning*. Bristol: Multilingual Matters, 47-54.

van Geert (1994). *Dynamic systems of development*. New York: Harvester Wheatsheaf.

van Geert (2007). Dynamic systems in second language learning: Some general methodological reflections. *Bilingualism: Language and Cognition* 10/01, 47-49.

van Geert (2008). The Dynamic Systems Approach in the atudy of L1 and L2 acquisition: An introduction. *The Modern Language Journal* 92/2, 179-199.

van Geert (2014). Dynamic modeling for development and education: From concepts to numbers. *Mind, Brain, and Education* 8/2, 57-73.

Verspoor (2015). Initial conditions. In: Dörnyei, Z.; MacIntyre, P.D.; Henry, A. (ed.), *Motivational dynamics in language learning*. Bristol: Macmillan, 38-46.

Windzio (2008). Social structures and actors: The application of multilevel analysis in migration research. *Romanian Journal of Population Studies* 2, 113-138.