Title

Multiple factors in second language acquisition: The CASP model

Journal

Linguistics, 51(1)

ISSN

0024-3949

Authors

Filipović, Luna

Hawkins, John A

Publication Date

2013-01-04

DOI

10.1515/ling-2013-0005

Peer reviewed

Multiple factors in second language acquisition: The CASP model*

LUNA FILIPOVIĆ AND JOHN A. HAWKINS

Abstract

We propose a new model of second language acquisition consisting of multiple interacting principles and inspired by work on complex adaptive systems. The model is referred to as CASP, short for complex adaptive system principles for second language acquisition. It is informed by a broad range of linguistic and psycholinguistic research and supported empirically by recent second language research studies based on a learner corpus. The novelty of our model lies in the definitions that we propose for a number of general and specific principles of learning, in the interactions that we demonstrate between them, in the predictions that we make and illustrate empirically, and in our integration of research findings from numerous areas of the language sciences. The result is a broadly based theory of SLA, which can potentially solve some of the traditional puzzles in this field, e.g., involving when transfer from an L1 does and does not occur.

1. Introduction

It has been noted many times by different scholars that second language acquisition involves the interplay of a number of factors that can either facilitate or impede learning. These include the typological relationship between L1 and L2, general principles of learning and critical ages, general principles of language processing (production and comprehension), social factors involving the general environment for learning, as well as pedagogical factors including teaching methods and materials and types of assessment. Various proposals have been put forward in order to illustrate how some of these factors interact. In our opinion, these proposals have met with only limited success. The principal reasons are, first, that not enough of the relevant factors have been modeled at any one time. And second, empirical testing of the interacting factors that determine SLA has been limited.

Our goal in this article is to outline a broadly based set of principles for a multi-factor model of learning in second language acquisition. We have developed these principles by drawing on a rich set of theoretical insights from numerous branches of the language sciences, grammatical theory, typology, language processing, computational linguistics, first language acquisition and second language acquisition. We illustrate our findings based on data-driven studies in the field.

In an important study comparing different theories of SLA Long (1993) was concerned with the question of how to choose between them and he summarized a number of assessment strategies for doing this. We would argue that there is a more important priority for future research. Some SLA theories do indeed stand in opposition to others. But the data of second language learning are constrained by multiple factors, psychological, biological, social, grammatical, typological, educational and pedagogical, and many, arguably most, of the principles and generalizations that have been proposed in language acquisition research are simply attempting to describe different pieces of this complex reality. The challenge before us, we believe, is to show how they interact and work together to constrain different

L1-L2 interlanguages and the progression towards, and the limitations on, L2 mastery. That is the goal of this article and of the learning model we present in preliminary form, namely CASP, short for Complex Adaptive System Principles (of SLA).

CASP has been influenced by work on complex adaptive systems that originated at the Santa Fe Institute, New Mexico (see Hawkins and Gell-Mann 1992 for discussion of this general approach to the language sciences and Gell-Mann 1992 for the underlying theory). In Sections 4 and 5 we propose our interacting principles for second language acquisition. They provide a framework for better understanding the relative sequencing in the acquisition of the numerous properties that constitute knowledge of an L2. Some of these properties are what we call "criterial features", i.e., they are important and defining differences between different levels of proficiency as the L2 is gradually learned. A large number of these features for L2 English is discussed in Hawkins and Filipović (2012) (see also Hawkins and Buttery 2009, 2010).

One big unresolved issue involves transfer from the L1 into L2. Theoreticians are divided over the extent of its impact, and the data that have been reported in case studies are often puzzling. For some linguistic properties transfer seems to occur productively (phonetic properties, for example), for others it is less common or even unattested. Some L1-L2 pairs reveal it, others do not.

Another point of interesting disagreement within the SLA community concerns the respective roles of ease of processing and ease of learning in explaining second language acquisition. What is the precise relationship between the two? When do they overlap in their predictions and when do they not? Pienemann's (1998, 2003, 2005) Processability Theory reduces much or most of learning to processing, and more specifically to limited processing capacity.

We begin this article with a brief overview of selected relevant literature on multiple factors in language acquisition and their competition and co-operation (Section 2), followed by a section on issues and definitions of some key concepts (Section 3), as a background to our proposed model. We then provide our data and analysis, first some general principles of the CASP model (Section 4) and then more specific principles following from them (Section 5). The last Section (6) summarizes our conclusions.

2. Brief literature review

One of the earliest multi-factor models of direct relevance for SLA was the Competition Model originally proposed for first language acquisition by Bates and MacWhinney (1981, 1982, 1987). MacWhinney (2005) subsequently extended it to SLA. A key proposal in this model is that language learning is based on *cue validity*, i.e., on the roles that different linguistic forms play in signaling grammatical or semantic information. Cues can have different weights in different languages and they may compete with one another. For example, the dominant cue for recognizing a grammatical subject in English is preverbal positioning, whereas in Spanish and Italian word order is not such a strong cue. The object relation in Spanish is often explicitly marked by a preposition whereas the subject is a simple Noun Phrase or zero (Spanish being a so-called "pro drop" language). In German the case marking of the definite article is a powerful cue for the subject, while in Russian it is case marking on the noun. MacWhinney points out that it is crucial to establish the relative cue strength for a given form-property linkage and he observes how these cue strengths interact. He proposes a unified model of language acquisition whereby the

"mechanisms of L1 learning are seen as a subset of the mechanisms of L2 learning" (MacWhinney 2005: 49). Even though L1 mechanisms are less powerful in the L2 learner than they were in first language acquisition, they are still partially accessible to the L2 learner. His model relies heavily on transfer from the L1, the basic claim being that "whatever can transfer will" (MacWhinney 2005: 55).

We agree that L1 transfer is one of the central factors in SLA, but there are also many occasions in which it is impeded by another factor of relevance to learning, such as complexity, or infrequency in the input. The range and impact of such interacting and potentially competing forces needs to be made explicit.

Another early proposal for multiple factors in language acquisition can be found in Slobin's (1977) *operating principles* based on crosslinguistic comparison. These operating principles are relevant for the principles of second language acquisition to be proposed here. They are presented as general information processing constraints that impact language learning and linguistic communication in general. Slobin's four operating principles are (see Slobin 1977: 186–188):

(i) Be clear.
(ii) Be processable.
(iii) Be quick and easy.
(iv) Be expressive.

These principles are formulated in the form of imperatives to the speaker, "Be clear" etc., but this should not be taken too literally. Slobin intends them as descriptive statements about what native speakers and learners actually do and about the linguistic data they produce. The principles we shall propose in Sections 4 and 5, while not formulated as imperatives, will also be offered as descriptive statements of observed behavior, without any commitment being made to normative agents or intentionality or to underlying psychological or biological causes or mechanisms.

Slobin's principles can either cooperate or compete to produce the observable linguistic outputs. Commenting on this, Strömqvist et al. (1998: 206) observe that, for example, "Be quick and easy" can put "Be processable" at risk. The first and second operating principles, "Be clear" and "Be processable", are prioritized in early stages of language development, while the third and the fourth have "greater relative weight" in later stages of development (Strömqvist et al. 1998: 206). This is understandable because "mapping out form-content relations and making yourself understood" are, they point out, priorities in early language acquisition. The latter two principles can also come into conflict, as they show in their study of spoken vs. written language acquisition. Spoken language is constrained by principles of on-line interaction, while a high degree of expressivity demands more grammatical and lexical resources and also more processing effort and planning time. These constraints of on-line interaction are lifted when composing a written text.

Slobin's principles provide a moral and a precedent for the kind of model we propose in Sections 4 and 5. There are multiple factors at play in structuring the patterns we observe in second language data, which sometimes reinforce one another but on other occasions come into conflict. We must also ask what the key concepts "be processable" and "be expressive" mean exactly and how these factors interact? These are some of the questions that our model tries to clarify and answer, in outline form at least.

The most recent multi-factor approaches to SLA view it as a form of complex adaptive system in the sense of Gell-Mann (1992) in which multiple factors interact to

produce a range of observable outcomes and different kinds of interlanguages. Some of these models are referred to as *emergentist* (see, for example, Ellis 1998; Larsen-Freeman 1997; Mellow 2008; O'Grady 2005, 2008). Larsen-Freeman and Cameron (2008) is explicitly in this tradition. The empirical support for their claims is not compelling, however, and is based largely on the longitudinal study of a single learner. The Ellis and Larsen-Freeman (2009) contribution to SLA makes use of connectionist modeling techniques in psycholinguistics which complement the more symbolic approach to the formulation of SLA principles as proposed in Sections 4 and 5.

These works all recognize the need for multiple interacting factors in understanding SLA. They differ in the number and nature of the principles proposed, in their precise formulation, in the predictions they make for interlanguage data, and in the range of data on which they have actually been tested. The empirical support offered hitherto is encouraging but still limited.

Contrasting with these integrative and interdisciplinary approaches to SLA are the more traditional and often single-factor theories that have characterized most SLA research ~~hitherto~~. One of the most influential has been the Universal Grammar (UG) theory of generative linguistics that hypothesizes an innate language faculty to guide the learning of languages (Chomsky 1965; Hoekstra and Kooij 1988). The learner's access to UG is claimed to solve the negative evidence problem, i.e., the ability of learners to learn the limits of grammaticality and to develop intuitions about ungrammaticality without generally being exposed to or taught about ungrammatical sentences. The claimed success of UG in explaining this aspect of learning, which is a remarkable empirical fact that any theory must eventually account for, has been questioned by Bowerman (1988) and Hawkins (1988), and more recently by Culicover (1999) and Hawkins (2004: 273). In particular, the vast majority of negative evidence problems for learning involve grammatical and lexical idiosyncrasies of particular languages, to which any putative innate constraints would not even begin to apply to constrain learning. For example, the grammaticality of *John is likely to pass the exam* in English, and the ungrammaticality of **John is probable to pass the exam*, involves an idiosyncrasy of English raising rules (cf. Postal 1974): *likely* triggers Subject-to-Subject Raising, *probable* does not. The problem this poses for learners, namely the absence of ungrammatical sentences from the data, is of the same logical type as that discussed by Hoekstra and Kooij (1988) involving subjacency constraints on movement rules. These latter are claimed to be underdetermined by the data but known innately by virtue of their universality. For the raising contrasts, and countless others, there is no innate constraint to help learners out, and yet they figure out the limits of grammaticality on the basis of positive evidence alone. This argues against the role of any innate UG as a general explanation for the negative evidence problem.

More generally the precise content of UG has undergone considerable scaling back in recent generative theories (see e.g., Hauser et al. 2002). In the present article we remain agnostic about UG and we note only that most of this second language acquisition debate becomes moot if the innateness of language is scaled back to include little more than recursion. The descriptive machinery of generative linguistics, on the other hand, has provided a rich set of tools for the description of first and second language acquisition data (see e.g., Radford 1990 and Hawkins 2001) and we shall make use of these tools here, without any commitment as to their ultimate explanation (see contributions in Hawkins (1988) for the different possibilities). We also consider generative-inspired SLA studies to have contributed very valuable data and case studies (Mellow 2008; Hawkins 2001; White 2003 and many others) that we

can draw on. But our approach in this article, as explained in the Introduction, is a multi-factor and integrative one, not an oppositional one as in Long (1993).

Our main emphasis will be on syntactic and semantic phenomena, i.e., it will be linguistically based, and on their learning and processing, i.e., it is also psycholinguistically based. This is not to deny or lessen the importance of the social dimension in language learning, for example, in the form of what Larsen-Freeman and Cameron (2008: 126) call "co-adaptation", defined as the linguistic adjustment that each interlocutor makes to others; see also Atkinson (2002) for a "sociocognitive" approach to SLA, and Preston (1989) for an interaction/social model that accounts for certain aspects of interlanguage variation. Second language learning can also be impacted by pedagogical and assessment factors (see Hawkins and Filipović [2012] for summary and references).

We have offered here only the briefest of literature reviews, in the interests of space, and we refer the reader to Hawkins and Filipović (2012) for a more detailed summary of relevant research in SLA. We turn now to what we consider some key issues that need to be resolved in this area and that we believe can be resolved by a more integrative, multi-factor approach.

3. Some issues in SLA and some definitions

3.1. *Transfer and non-transfer*

One big issue that is currently unresolved in SLA research involves transfer. It seems to occur productively for some linguistic properties, whereas for others it is less common or even unattested. Some L1-L2 pairs reveal it, while others do not.

Transfer plays a central role in MacWhinney's competition model for L2 acquisition. For Pienemann (1998, 2003, 2005), however, it does not. Transfer from the L1 is constrained, according to his Processabilty Theory, by general processing principles, irrespective of the typological distance between L1 and L2. Order of acquisition in the L2 is guided by a processability hierarchy that determines if and to what extent L1 transfer can be expected. The competition model predicts that the learner acquires L2 structures directly in areas where transfer is poorly supported (MacWhinney 1992: 371). For example, there is a lack of negative transfer (i.e., relatively few errors) in the early acquisition of L2 Japanese SOV by English learners (Kawaguchi 1999, 2002) and this for MacWhinney is a consequence of the emphasis on the input that fixes the Japanese word order for learners in the very early stages of acquisition (MacWhinney 2005: 60).

As we see it, input may indeed play the crucial role here, but this does not explain why some L1 speakers do nonetheless transfer their word order preferences into L2 English whereas others do not. We return to this puzzle in Sections 4 and 5.

Similarly McDonald and Heilenman (1991: 331) point out that English learners of French L2 abandon their English word order strategies early, particularly when learning noncanonical orders. And according to Gass (1987, 1989) English learners of Italian switch to animacy as an early cue to subjecthood rather than sticking to their preferred L1 word order cue, whereas Italian learners of English seem to move more gradually to the dominant word order pattern of English (Gass 1989: 193). Japanese learners of English acquire SVO very rapidly (Rutherford 1983), just as English learners of Japanese produce SOV order and subject omission early, even though their own language is SVO and non-pro-drop, i.e., English does not permit

subject omission (Kawaguchi 1999, 2002). English and Japanese have diametrically opposed word orders, English being head-initial and Japanese head-final (Hawkins 1983, 1990). It is often claimed in the literature that pragmatic or discourse functions for word order can be transferred from the L1 into L2, but there is disagreement over whether grammaticalized word order patterns can also be transferred and under what conditions (cf. Ellis 2003: 316–317).

Hawkins and Filipović (2012) offer evidence that Spanish learners of English do transfer their pro-drop patterns along with a number of un-English word orders, including what we might call Light NP Shift (the type of *I like very much sweets*) and post-posed subjects (of the type *Yesterday came my boyfriend*), the motivation for which is arguably as much or more to do with processing efficiency than with pragmatics (~~cf.~~ Hawkins 2004, Wasow 2002). Similarly Chinese speakers transfer preverbal PPs as in structures like *You can by bus get there*. On the other hand, there are numerous examples of L1-L2 contrasts that do not result in transfer effects. These have been much less discussed in the literature than cases where transfer does occur, for understandable reasons, but the non-occurrence of transfer is just as revealing and as significant for theories of SLA, in some ways more so, as its occurrence. For example, Pienemann (2005) gives examples of different L1 learners of Spanish L2 all getting the Spanish pro-drop structure from the very beginning (studies cited in Pienemann 2005). Japanese learners of English and English learners of Japanese do not appear to transfer their L1 word orders into the L2, however. Spanish, Chinese and Japanese learners of English derive different benefits from direct learning and we need to specify why this is so. Clearly we need a model of learning and processing that can account for these differences and our model does this in many cases.

More generally, our model enables us to capture numerous phenomena pertinent to transfer. For instance, definite and indefinite articles provide good examples of L1 transfer, since many languages lack them and there are many language-specific contrasts here, and hence their presence or absence in an interlanguage can often be at variance with the L2 and convergent with L1 (see Section 5.1 below). Cross and Papp (2008: 68) show that Chinese learners of English make significant numbers of errors involving the appropriate co-occurrences of verbs with prepositions (they count 20% in their learner corpus). These are often lexically idiosyncratic in English. They argue that L1 Chinese is the reason behind the omission or inaccurate use of prepositions since Chinese has fewer words functioning as prepositions and they are used less frequently than in English.

Summarizing this purposefully brief account, the research literature on transfer in SLA provides us with some intriguing examples of transfers that occur productively, and of others that do not. There is much that is still unresolved at a theoretical level, therefore, regarding when it occurs, and when it does not, and with what L1-L2 pairs. In Sections 4 and 5 we incorporate and revise some of the ideas that we have summarized in this section and we propose a number of different factors that hold for second language acquisition in general. It is the interaction between them that determines possible versus impossible interlanguages and that predicts when transfer should, may or should not occur.

3.2. *Processing and learning*

Another point of interesting disagreement within the SLA community concerns the respective roles of ease of processing and ease of learning in explaining second language acquisition. More generally, what is the precise relationship between these

two theories? When do they overlap in their predictions and when do they not? Pienemann's (1998, 2003, 2005) Processability Theory reduces much or most of learning to processing, and specifically to limited processing capacity. He states (Pienemann 2003: 679) that the limited-capacity view of L2 processing constitutes a basic assumption in his work and in that of numerous other researchers (e.g., Clahsen 1984; Krashen 1982; VanPatten 1996; Bates and MacWhinney 1982). Pienemann argues further (2003: 679) that his Processability Theory makes "testable predictions for developmental routes across typologically different languages" and it "applies to L2 as well as to L1 acquisition". In other words, he claims that these central facts about learning follow not from a theory of learning , but from processability.

We share Pienemann's view that a lot of learning can and should be explained in terms of processing, but learning and processing are nonetheless distinct, in principle, and we do need theories of both, even though their effects may sometimes be hard to tease apart. In fact, equating ease or difficulty of learning with ease or difficulty of processing seems to fail in both directions.

For example, the learning of morphological irregularities is hard (sing/sang/sung vs. walk/walked, mouse/mice vs. house/houses, etc.) and children will go through a stage of regularizing them (~~cf.~~ Andersen 1992 for a literature review), but once learned they are often easy to process on account of the frequency and entrenchment of irregular forms (see Bybee 2007). Conversely, semantically transparent forms, with a one-to-one mapping between forms and their meanings (Slobin 1973, 1985; Andersen 1984), facilitate learning. Yet processing is often easier when surface forms are reduced and shorter and less transparent, namely when semantic interpretations can be readily inferred. *John is eager for John to please* is a semantically transparent structure with the subject of pleasing expressed explicitly and overtly in the infinitival phrase. *John is eager to please* is arguably easier to process because fewer words need to be processed and the assignment of *John* as a subject argument *to please* can be inferred clearly and unambiguously (~~cf.~~ Hawkins 2004: 162–164). The one-to-one mapping principle was introduced by Slobin (1977) in a first language context. He argued that the agglutinative nominal morphology of Turkish (one case form for each case, nominative versus accusative, etc.) is easier to learn than the many-to-many mappings of rich inflectional languages like Russian (one case property being conveyed by many forms, one form often conveying many cases). Evidence for the simplicity of one-to-one mappings has been replicated for second language acquisition by Andersen (1984) who states that "[a]n interlanguage system should be constructed in such a way that an intended underlying meaning is expressed with one clear invariant surface form" (Andersen 1984: 79).

A second point to mention in connection with Processability Theory involves the central notion of limited processing capacity. There almost certainly are such limits, but attempts to define exactly what they are have proven to be extremely difficult (~~cf.~~ Hawkins 1994, 2007 for critical discussion of proposed working memory limits for word order processing that are counterexemplified in languages other than English). Those who propose such theories generally end up defining a relative ranking of processing difficulty, among the permitted options, rather than making a concrete proposal for a well-defined upper limit that excludes certain structural types as unprocessable. Pienemann does this as well with his processability hierarchy.

Discussions of processing difficulty are further compounded by general issues in language processing research that have not yet been resolved. Hawkins (2011) provides a brief summary of some of these. They include the sometimes quite different empirical findings associated with different methodologies. On-line

experiments, acceptability judgments and corpus data reveal quite distinct patterns for the "weight"-based syntactic constructions discussed in Hawkins (2011); see also Konieczny (2000) and Francis (2010). Some models focus more on the integration of on-line material (Gibson 1998, 2000), others on the predictability of what lies ahead (Hale 2001; Levy 2008), some on the speaker's perspective in production (Wasow 2002), others on the hearer and comprehension (Hawkins 1994). These contrasting data and assumptions are discussed in Hawkins (2011), to which the reader is referred.

Until these larger issues are resolved it will be difficult to define with certainty what the predictions of "processability" and of ease and difficulty in processing are and to compare them with the predictions from learning and learnability, which also involve different perspectives and unresolved issues of their own (see Reali and Christiansen 2009; Stabler 2009). In the interim there do appear to be large areas of overlap in the mechanisms of learning and processing and in their respective predictions for SLA, but equally clearly there are some differences between them.

3.3. *Complexity and frequency*

Two factors that play a central role in any model of language use and language acquisition are simplicity/complexity and frequency/infrequency. Often they correlate. Simpler structures are generally used more frequently (Wasow 2002; Hawkins 2004) and are acquired earlier (Diessel 2004; Hawkins and Filipović 2012). Complex ones are infrequent and acquired later. There are cases, however, in which the correlation fails. For example, Diessel (2004) shows for first language acquisition that in a number of instances simplicity does not correlate with frequency, i.e., more complex structures are acquired earlier on account of their frequency.

As in the previous section on processing and learning, a prerequisite for assessing the relative weight of two major factors, such as complexity and frequency, requires a general theory of each. In the case of complexity, this means answering the question: what exactly does "complex" mean? For frequency, the situation is potentially easier, although the universe of comparison for linguistic items relative to which frequency is calculated is never straightforward. A key question is therefore what exactly is complexity and how do we define it?

Consider first a clear example. Tough Movement structures in English (such as *John is easy to please*) are complex in their grammar and are hard to process and learn. Children initially interpret the surface subject *John* as the subject of *easy* and not as the object of *to please*, moving gradually to adult-like competence. In other words, they first assign the more frequent, and simpler, and unmarked interpretation to this adjacent subject-predicate pair (Anderson 2005). For second language acquisition Callies (2008) has shown, on the basis of a study using a learner corpus, that Tough Movement structures are acquired late by German learners of English and are significantly underrepresented in the writings of even advanced German learners.

More generally Mellow (2008) describes a revealing case study of relevance to syntactic complexity in the learning of English. He shows how a 12-year old Spanish learner gradually acquires syntactic dependencies, including control structures in verb and adjective complements and many other structures, and argues that complexity results in later acquisition.

Some issues that complicate a general definition of complexity are discussed in Hawkins (2009), for example *trade-offs*. Simplicity in one part of the grammar is often matched by complexity in another. English transitive NP-V-NP sequences regularly have an Agent-Verb-Patient interpretation, as in *The king attacked the*

enemy, but often they must be mapped onto complex argument structures in ways that many (indeed most) languages do not permit, even closely related languages like German (cf. Hawkins 1986; Müller-Gotama 1994). Some less common theta-role assignments to transitive subjects that are grammatical in English but ungrammatical in most other languages are: *This tent sleeps four* [Subject = Locative]; *A few years ago a penny would buy two to three pins* [Subject = Instrument]; *The book sold 10,000 copies* [Subject = Theme]. There are more argument structure types to be linked to NP-V-NP than to the corresponding transitive clauses of languages with only Agent-Verb-Patient interpretations, which adds length and complexity to the grammar of English (as discussed in Dahl 2004).

This also makes processing more difficult. Assigning a Locative to a subject NP or an Instrument requires access by the processor to the verbs in these sentences, to their lexical semantics and co-occurrence structure, and also possibly to the post-verbal NP. More disambiguation must be carried out by the English processor, and greater access is needed to more of the surface structure for this disambiguation (Hawkins 2004).

It is trade-offs like this that make it difficult for us to give a definition of overall complexity for a grammar and that result in unresolvable debates over whether some grammars are more complex than others, when there is no clear metric of overall complexity for deciding the matter.

These difficulties in defining overall complexity should not prevent us from appealing to the complexity of certain individual constructions or meanings, however, when there is ample empirical evidence to back up the complexity claim, as there is in the Tough Movement example. This evidence can come from grammar alone, from crosslinguistic distributional evidence, from processing, and from acquisition. There is an emerging consensus in the literature that complexity does have measurable consequences in all these areas, in the form of fewer languages with the complex pattern in question, greater processing difficulty, and later and more errorful first and second language acquisition (~~cf.~~ Hawkins 2004; Hawkins and Filipović 2012) .

3.4. *Complexity and efficiency*

A final issue in this section that will be relevant for our model involves the relationship between efficiency and complexity. This is discussed in Hawkins (2009) (see also Hawkins 2011). He argues that metrics of complexity need to be embedded in a theory of efficiency, and that efficiency and not complexity is the larger and more inclusive notion. Efficiency relates to the most basic function of language, which is to communicate information from the speaker (S) to the hearer (H). He proposes that:

a) Communication is efficient when the message intended by S is delivered to H in rapid time and with minimal processing effort;
b) Acts of communication between S and H are generally optimally efficient; those that are not occur in proportion to their degree of efficiency.

Complexity metrics, on the other hand, are based on the grammar and structure of language. An important component of efficiency will often involve structural and grammatical simplicity. But sometimes communicative efficiency requires the use of structures that are more complex, for example when a hearer needs more detailed and explicit information about a referent or event. Noun phrases whose referents are unfamiliar or inaccessible to the hearer require longer and more complex structures

(e.g., *the professor that I bumped into this morning*) compared with those that are familiar and accessible (such as *he*); see Hawkins (1978, 1991), Ariel (1990). Efficiency also involves additional factors that determine the speaker's structural selections including: *speed* in delivering linguistic properties in online processing; *fine-tuning* of structural selections to frequency of occurrence, accessibility and inference; and *few online errors* or garden paths. These factors interact, sometimes reinforcing sometimes opposing one another (see Hawkins 2004, 2009 for details).

4. The CASP model: General principles

After this brief review of relevant research and of certain issues in SLA, first language acquisition, language processing, typology and complexity, we can now propose some principles of our model. Some of these principles have been discussed in previous corpus-based studies (e.g., Hawkins and Buttery 2009; Hawkins and Filipović 2012). What is novel to our approach is the breadth of factors to which we appeal, as well as details of the proposed formulations and the interaction between principles. As we pointed out in the Introduction we refer to our model as CASP, short for "Complex Adaptive System Principles of SLA", since it follows the general logic of complex adaptive systems (Gell-Mann 1992). The model consists of four very general principles that we propose in this section. More specific subprinciples following from them are discussed in the next (Section 5). They define possible versus impossible, and likely versus less likely, acquisition stages and interlanguages and they are offered here as a first attempt at a predictive model for the relative sequencing of learner data.

The first general principle we propose is principle (A):

(A) Minimize Learning Effort (MiL)
Learners of a second language (L2) prefer to minimize learning effort when they learn the grammatical and lexical properties of the L2.

Learning effort can be minimized in a number of ways. It is minimized when grammatical and lexical properties that are shared between L1 and L2 can be transferred directly into the L2, thereby exploiting pre-existing knowledge from the L1 (see subprinciple (1) in Section 5.1). It is minimized when properties of the L2 are frequently occurring in the L2 input, which increases their exposure and with it the ease of learning (subprinciple (2), Section 5.2). It is minimized when structural and semantic properties of the L2 are simple rather than complex (subprinciple (3), Section 5.3). These specific principles following from principle (A) will be defined and illustrated in the next section.

Note that principle (A) is intended as a descriptive statement that captures observed patterns and preferences in second language data. There is no suggestion here that learners are consciously or intentionally striving to comply with (A), or are taught to comply with it, for this principle or for others in this article.

The second general principle is (B):

(B) Minimize Processing Effort (MiP)
Learners of a second language (L2) prefer to minimize processing effort when they use the grammatical and lexical properties of the L2, just as native speakers do.

Even when more complex properties have been learned at an acquisition stage, L2 learners will still prefer to use simpler properties, just like native speakers do.

We have seen some real debate (Section 3.2) over the extent to which developmental stages in SLA are shaped by ease of processing or by ease of learning. This is a big and complex issue in psycholinguistics generally. Our view, as mentioned above, is that there is a clear need for both a theory of learning and a theory of processing, given certain dissociations between them, but that their predictions for the observable outputs of second language acquisition will often overlap. Infrequent items in the input will be harder to learn and also harder to access and process. Complex grammatical and lexical properties will also be harder to both learn and process.

Principles (A) and (B) are, in essence, principles of least effort. If these were the only principles determining learning and production, however, our learner corpora would reveal increasingly minimal outputs. Clearly, they do not. As Hawkins and Filipović (2012: Ch.2.2.) point out, MLUs (i.e., mean length of utterance figures) increase at higher proficiency levels, and greater use is made of less frequent and more complex structures and meanings. The reason, we suggest, is that second language learners are always striving to increase their expressive power in the L2, and to behave like native speakers, which means learning and using the mix of infrequent and frequent, and complex and simple, linguistic items, just like native speakers do. This is captured in principle (C).

(C) Maximize Expressive Power (MaE)
Learners of a second language (L2) prefer to maximize their expressive power, i.e., to formulate in the L2 whatever thoughts they would wish to express in the L1, and to perform the same language functions as L1 users.

We shall not delve into the thorny issue here of what the term “native speaker” can mean in SLA. We simply view it as an ideal towards which L2 learners can aspire even though not all native speakers have necessarily attained it, given their educational and social backgrounds (for an in-depth view and analysis see Davies 1991). As a result of principle (C) successive stages of acquisition reveal increasingly complex sentence structures for the expression of increasingly complex thoughts and language functions. This principle stands in partial opposition to principles (A) and (B), therefore, and results in increasingly more complex, less frequent and more native-like L2 outputs.

There is another general principle that is sometimes opposed in its predicted outputs to principles (A) and (B), and that involves not expressive power per se, i.e., semantics, but the efficient delivery of meanings in real time between given interlocutors. We propose principle (D) in this context.

(D) Maximize Communicative Efficiency (MaC)
Learners of a second language (L2) prefer to maximize their communicative efficiency in relation to the hearer and his/her mental model.

Our definition of communicative efficiency is based on Hawkins (2004, 2009), see Section 3.4 above. Communication is efficient when the message (M) intended by the speaker (S) is calibrated to the hearer's (H) mental model in such a way as to achieve accurate comprehension of M with rapid speed and the least processing effort

compatible with H's mental model. The need to communicate efficiently with a given hearer on a given occasion results in sometimes more, sometimes less processing effort, in partial opposition to principle (B). Recall the different referring expressions *the professor that I bumped into this morning* versus *he*. The result is sometimes more, sometimes less, complex expressions. Principle (D) is also evident in certain checks on negative transfer (see Section 5.5).

5. Some specific principles and supporting patterns

In this section we present and exemplify some more specific principles (building on those of Hawkins and Filipović 2012) and crucially, we explain how they derive from the general principles of the CASP model discussed in the last section.

5.1. *Maximize Positive Transfer*

Our first specific principle is Maximize Positive Transfer (1):

(1) Maximize Positive Transfer (MaPT)
Properties of the L1 which are also present in the L2 are learned more easily and with less learning effort, and are readily transferred, on account of pre-existing knowledge in L1.

When properties are shared between L1 and L2 we generally expect to see earlier L2 acquisition, more of the relevant properties being learned, and fewer errors, unless these shared properties involve e.g., high complexity and are impacted by other factors such as (3) below. Properties that are not shared will be harder to learn on account of the additional learning that is required, again in general.

Principle (1) (MaPT) derives from general principle (A) MiL, therefore. Positive transfers are advantageous for learning. We would argue that they are also good for processing since processing effort is minimized when transfers are positive. Processing mechanisms that are already being used for production and comprehension in the L1 can be applied directly to the processing of L2 in this way, reducing overall processing effort, in accordance with general principle (B) MiP. Furthermore, positive transfers enhance the expressive power of the non-native L2 user, in accordance with principle (C) MaE, by capitalizing on native speaker expressiveness in the L1. Communicative efficiency can also be maximized by positive transfers, see principle (D) MaC, in the event that the non-native L2 user can transfer L1 efficiency routines for the appropriate selection of these positive properties. In short, while Maximize Positive Transfer may be motivated primarily by general principle (A) MiL, it is also reinforced by general principle (B) MiP, by general principle (C) MaE, and even by general principle (D) MaC. Everything converges to encourage it.

For some relevant data, consider the learning of definite and indefinite articles in English (discussed in Hawkins and Buttery 2010 and Hawkins and Filipović 2012). The English articles are easier to learn when the L1 has the same grammar and usage. This is illustrated in Tables 1 and 2 (adapted from Hawkins and Buttery 2010) which give sample learner data from L2 English involving the presence versus omission of definite and indefinite articles by speakers of first languages with articles (French, German and Spanish) and without articles (Turkish, Japanese, Korean and Russian) at five successive stages of learning. The missing determiner error percentages are very

low at all proficiency levels for the L1s with similar article systems to English but they are significantly higher and show a general linear improvement from level 3 to level 6 proficiency for the L1s without articles.

Table 1. *Missing determiner error percentages for L1s with articles in L2 English*

Missing "the"	Level 2	Level 3	Level 4	Level 5	Level 6
French	4.76	4.67	5.01	3.11	2.13
German	0.00	2.56	4.11	3.11	1.60
Spanish	3.37	3.62	4.76	3.22	2.21
Missing "a"					
French	6.60	4.79	6.56	4.76	3.41
German	0.89	2.90	3.83	3.62	2.02
Spanish	4.52	4.28	7.91	5.16	3.58

Table 2. *Missing determiner error percentages for L1s without articles in L2 English*

Missing "the"	Level 2	Level 3	Level 4	Level 5	Level 6
Turkish	22.06	20.75	21.32	14.44	7.56
Japanese	27.66	25.91	18.72	13.60	9.32
Korean	22.58	23.83	18.13	17.48	10.38
Russian	14.63	22.73	18.45	14.62	9.57
Missing "a"					
Turkish	24.29	27.63	32.48	23.89	11.86
Japanese	35.09	34.80	24.26	27.41	15.56
Korean	35.29	42.33	30.65	32.56	22.23
Russian	21.71	30.17	26.37	20.82	12.69

These results support principle (1) MaPT. Speakers of French, German and Spanish already know many of the rules and conventions for using English definite and indefinite articles, because they overlap significantly with their own L1s. There is much less learning required for these speakers than for speakers of Turkish, Japanese, Korean and Russian, whose languages do not have definite and indefinite articles, and for whom learning is more effortful. This is shown in the error rates.

Principle (1) asserts that there will be learning advantages whenever L1 and L2 share grammatical and lexical properties. Sharing of grammatical and lexical properties between L1 and L2 does not automatically mean that there will be early transfer of these (positive) properties. This is because there are other factors with which principle (1) MaPT interacts, such as principles (2) and (3) below. Complex linguistic units and properties, for example, complex consonant phonemes or complex lexical meanings shared between L1 and L2 may not be transferred early on account of their complexity. Eckman (1984) provides relevant phonological data here showing that Persian learners of English did not initially transfer the voiced-voiceless opposition in word-final obstruents from Persian into L2 English, even though both languages share this crosslinguistically marked feature. Kellerman (1983) has shown that more complex and non-core lexical meanings of verbs like *break* are not initially transferred from L1 Dutch into L2 English, despite the fact that these two languages permit the same semantic extensions for this verb. The principles of a multi-factor

model can be in partial competition like this, and this can result in different possible outcomes depending on their relative strengths. The greater the complexity, for example, the more it will resist positive transfer.

5.2. *Maximize Frequently Occurring Properties*

Our second more specific principle is (2) (see Hawkins and Filipović 2012):

(2) Maximize Frequently Occurring Properties (MaF)
Properties of the L2 are learned in proportion to their frequency of occurrence (as measured, for example, in the British National Corpus): more frequent exposure of a property to the learner facilitates its learning and reduces learning effort.

More frequent properties in the input will accordingly result, in general, in earlier L2 acquisition, in more of the relevant properties being learned, and in fewer errors. Learning is more effortful for learners when properties are infrequent in the input, i.e., this principle derives from general principle (A) MiL. The learning of high-frequency properties also reduces processing effort when they are being used, however, in accordance with general principle (B) MiP.

In our summary of MacWhinney's (2005) competition model for first and second language acquisition (Section 2), we saw that grammatical properties are acquired early when they are used frequently and have high cue validity. This was exemplified for the expression of basic argument relations and thematic roles, agent versus patient, etc. Different languages use different strategies here, basic word order, agreement, animacy, etc., and they conventionalize them to different extents, with more or less cue strength and frequency. It appears that learners are sensitive to these cues and they acquire relevant properties rapidly when they are frequent. Not only does this reflect principles (A) MiL and (B) MiP, it also supports our communicative efficiency principle (D) MaC: learners want the meanings of their basic clause types to be readily understood by native speakers, and efficient communication requires rapid learning of cues that are appropriate for the L2.

Further support for principle (2) MaF comes from a study by Williams (2007), summarized and discussed in Hawkins and Filipović (2012: 89–90). This study is based on the Cambridge Learner Corpus (CLC), which contains over 40 million words of learner data at all levels of acquisition from over 130 typologically and genetically different languages and which is composed on standardized written exams by Cambridge ESOL. Williams examines the acquisition of new construction types in English, described and parsed in terms of verb co-occurrence or "subcategorization" frames, appearing at the levels 2, 3 and 4. The data were based on the first/earliest occurrence in the corpus. At the earliest level (2) we find simple and frequently occurring intransitive sentences types (*he went*), transitive types (*he loved her*), and basic three-place predicate types with a prepositional phrase (*she added the flowers to the bouquet*). At the higher levels (3 and 4) we see more complex and less frequent sentence types, for example, with different embeddings (*he explained how to do it*, *he asked whether he should come*, *he told the audience that he was leaving*) and gerundive verbs with –ing (*I caught him stealing*, *they worried about him drinking*). It turns out that there is a precise correlation between the order of acquisition and degree of frequency in the input, in accordance with MaF (see Williams 2007).

5.3. *Maximize structurally and semantically simple properties*

Frequency and structural complexity are often inversely correlated in language use, i.e., the more complex a structure is, the less frequently it is used (for relevant data see Hawkins 2004 and Wasow 2002). The progression involving frequency in native corpora also appears to correlate, in general, with the increasing complexity of these verb co-occurrence frames in the learner data (see Hawkins and Filipović 2012). Learners of English first learn the simpler co-occurrence frames of English before they learn more complex ones. But simplicity/complexity and frequency/infrequency are not always aligned (see Diessel 2004 for some illustrative dissociations between frequency and complexity in first language acquisition), and definitions of complexity are not always in agreement with one another (Section 3.3). Frequency effects can be more readily observed and quantified, so we do need separate principles here, for theoretical and methodological reasons. Our third more specific principle is accordingly (3) (see Hawkins and Filipović 2012):

(3) Maximize Structurally and Semantically Simple Properties (MaS)
Properties of the L2 are learned in proportion to their structural and semantic simplicity: simplicity means there are fewer properties to be learned and less learning effort is required.

We expect simpler properties to result, in general, in earlier L2 acquisition, in more of the relevant properties being learned, and in fewer errors. Learning is more effortful for complex structures and meanings as there are more properties to be learned, i.e., principle (3) derives from general principle (A) MiL. It also follows from general principle (B) MiP, since processing effort is minimized when structurally and semantically simple properties are used, and this preferred use of simple properties may also make communication more efficient, in accordance with general principle (D) MaC.

For instance, simpler consonants as well as simpler consonantal distinctions are often acquired earlier than more complex ones (Eckman 1984). On the syntactic level, Mellow (2008) showed in his case study of a 12-year old Spanish learner of English that complex syntax, including complex syntactic dependencies, was acquired relatively late. A further example is Tough Movement (*John is easy to please*), which is also acquired late in both first and second language acquisition (Anderson 2005; Callies 2008). This and other types of "raising" structures, Subject-to-Subject Raising (*John is likely to pass the exam*) and Subject-to-Object Raising (*I believe John to be sick*) are analyzed in more detail in Hawkins and Filipović (2012: 121–127). All these raisings involve complex displacements of arguments from a lower clause into a higher clause in which they do not engage in normal semantic relations with their matrix verbs and adjectives (it is not the case that *John is likely* in any sense, nor that I necessarily *believe John*). They also involve surface syntactic ambiguity between the raising interpretation and the more common "control" pattern in which surface arguments do contract normal semantic and syntactic relations with adjacent predicates (see Chomsky's 1957 *John is easy to please* vs. *John is eager to please*; cf. also Hawkins 1986).

Hawkins and Filipović (2012: Ch.7) give examples from their empirical corpus study, which show when raising structures of each of the different types are first attested for an illustrative set of raising triggers taken from Postal (1974). For example, Subject-to-Subject Raising (*Bill is likely to pass the exam*) first appears at

level 3, but solely with the verb *seem* and the adjective *supposed*. The majority of verbs (*appear, cease, fail, happen, prove, turn out*) and adjectives (*certain, likely, sure, unlikely*) that trigger this construction occur at level 4. This confirms a point that has been stressed in Hawkins and Filipović (2012) and also by others (e.g., Ellis and Larsen-Freeman 2009): we need to consider both grammatical constructions and their lexical triggers when making claims about stages of acquisition (and about criterial differences between levels in SLA). Hawkins and Filipović (2012) report that Subject-to-Object Raising (*I want him to do his homework*) is attested with three lexical triggers at level 3 (*expect, want and like*), but the remaining nine in the set of 12 we looked for are level 4 (*imagine, prefer*) or later (level 5 *believe, find, suppose, take* and level 6 *declare, presume, remember*). Thus the classification of Subject-to-Object Raising as a level 3 feature would be correct as suggested in the study by Williams (2007), but this is valid only for its first appearance: most instances occur later, sometimes significantly so at levels 5 and 6.

By the same token, Subject-to-Object Raising plus Passive (e.g., *He is expected to arrive early*) is first attested at level 4, while most of the triggering predicates actually occur at level 5 (*assumed, discovered, felt, found, proved*), and most (three) of the five attested Tough Movement triggers (as in *This game is easy to play*) in our sample are level 4 (*difficult, good, hard*).

The fundamental point to make in this context is that it is *vital to search for lexically specific syntactic patterns* when trying to determine whether a particular structure has been acquired, and to what extent. A construction may be present with only one or very few verbs at first, but this is not sufficient to conclude that the construction has become a productive part of the learner's grammar.

We see a similar pattern of development in lexical semantics: simpler and more basic meanings for verbs are acquired earlier than their more complex extensions. For instance, as illustrated in Capel (2012) and Hawkins and Filipović (2012) we can map the learner lexical progression using learner corpora. For instance, Capel (2012) shows that the acquisition of the verb *break* is registered at level 2 in its basic physical sense (e.g., *break a glass*). *Break* in the extended sense of 'interrupt' is level 3 (e.g., *break the routine*) and in its extended sense of 'end' at level 4 (e.g., *break an agreement*). The idiomatic meaning of *break* as in *break the bank* is first found at level 5 while an original figurative/metaphorical use (*break the wall his fears have forced him to* create) is level 6 (see Capel 2010, 2012 for further details and analysis; see also Hawkins and Filipović (2012: Ch.8) for many further examples of such lexical progressions).

It is important to combine lexis and grammar when describing the latter because a given construction type may appear at earlier proficiency levels with only a limited set of triggering verbs, generally the most frequently occurring ones, in accordance with principle (2) MaF. We therefore need to document the expansion in usage for constructions, just as we do for words and their meanings, by showing which additional verbs are found in particular constructions as learning progresses.

An important parallel can be drawn between our results on L2 acquisition and those of Tomasello (2000, 2003) on L1 acquisition. He argues that children's early language does not revolve around abstract categories and rules but rather around specific lexical items and expressions. Child learners first acquire some constructions with specific verbs, others with other kinds of verbs, and they are conservative in their usage in the early stages but start extending the use of constructions with more verbs as they progress

In Section 3.2 we emphasized the need to define both a theory of learning and a theory of processing, since certain linguistic phenomena are easier to learn but harder to process, and vice versa. There appears nonetheless to be a large overlap between the predictions for learning and for processing since the striving for simplicity impacts on-line processing just like it impacts learning. Infrequent linguistic items and their meanings are harder to learn and harder to access and process, in general. More complex structures are harder to learn and also harder to process, and they are rarer in corpora as a result. This all makes it difficult to tease learning and processing apart. We are not going to resolve this major issue in psycholinguistics in the present context. For a processing counterpart to specific principle (3) the reader is referred to Hawkins and Filipović (2012).

5.4. *Negative transfer and its constraints*

When grammatical and lexical properties are shared, transfers from L1 into L2 result in positive, i.e., correct, properties of L2 being learned in a simple and efficient way. However, when properties are not shared, and the transfer still takes place, this results in negative or incorrect properties in the L2. We have argued that positive transfers are maximized on account of general principle (A) Minimize Learning Effort, i.e., they are always advantageous for learning, and they are further motivated by principle (B) Minimize Processing Effort, principle (C) Maximize Expressive Power, and also principle (D) Maximize Communicative Efficiency.

Negative transfers appear to be a different, however. From a purely empirical perspective we cannot say that they are maximized, as positive transfers are, since we have seen (Section 3.1) that sometimes they occur and sometimes they don't. This is one of the central unresolved issues in SLA. Clearly our model must be set up so that it allows negative transfers on some occasions, but not others. The general and more specific principles defined in this article can help us understand this interplay of forces and make some predictions.

We see negative transfers as being motivated by the desire to maximize expressive power (principle (C) MaE) and also to maximize communicative efficiency (principle (D) MaC) in an L2 system that has been incompletely learned, while at the same time minimizing learning effort (principle (A) MiL) and processing effort (principle (B) MiP). In other words, the same general forces that structure positive transfers do structure negative transfers as well. The major difference between them is that there are real limitations on expressive power and on communicative efficiency that can be conveyed by linguistic properties that are not actually part of the L2 and not used by its native speakers. When native speakers communicate with L2 learners they can tolerate and compensate for departures from the native language conventions to a certain extent. On some occasions, however, when learners' outputs depart too radically from the native speaker's conventions, learners are not understood. Learners accordingly acquire a sensitivity to the native speaker's ability to compensate for these violations in conventions of grammar and use, and it is this, we would argue, that ultimately determines whether and when negative transfer can occur. Our principle (4) for negative transfers captures this interaction among general principles of the CASP model that can result in errors:

(4) Permit Negative Transfer (PNT)
Properties of the L1 which are not present in the L2 can be transferred, resulting in errors, as learners strive to achieve an expressive power and

communicative efficiency in L2 comparable to that in their L1 (see principles C and D), while minimizing learning effort (see principle A) and/or processing effort (see principle B).

Consider phonology. Substitutions of L1 consonants like [t] or [s] or [f] for L2 [θ] in words such as English *thin* minimize learning and processing effort for learners whose L1s do not have this consonant, and they succeed in expressing the word in question, generally with communicative success (see Lado's [1957] discussion of relevant languages and interlanguages). In syntax Spanish pro-drop (e.g., **is a beautiful country* for *it is a beautiful country*) is often transferred into early L2 English to express the proposition in question and the removal of the subject does not impede communicative success. This structure is simpler than its English counterpart with an overt subject, and transfer is not blocked, as predicted by our principle (3) MaS. Similarly, many article omission errors do not diminish expressive power and communicative success, and at the same time they minimize learning and processing effort through the transfer of L1 structures, e.g., *piece of X* for ***a*** *piece of X* by Serbian learners (see also Tables 1 and 2). By contrast, Chinese prenominal relative clauses do not result in errors whereby the English *man whom the woman loves* is changed into its Chinese prenominal counterpart **the woman loves whom man*, and this is in part because this is a complex and typologically marked structure in Chinese (Hawkins 1999, 2004). Complex lexical or constructional meanings in an L1 without an L2 equivalent will not generally transfer negatively, on account of principle (3) MaS or on account of their infrequency (2) MaF. Similarly complex structures and meanings may not immediately transfer from L1 to L2 even when they are shared, in opposition to principle (1) MaPT. What this all means is that principles (2) and (3) can block both positive and negative transfers into L2.

We can also account for errors like the overuse of topicalization by Chinese learners and by French learners (see Trévise 1986), and the underuse of the passive by Hebrew learners (see Seliger 1989). None of these transfers in usage affect the ability of learners to make themselves understood by the hearer. Grammatical errors or usage differences that do not affect communication may either be eliminated at later stages or they may persist throughout acquisition (as in the case of many uses and misuses of the articles).

We propose specific principle (5), which derives from the need for communicative efficiency (principle D), and which ultimately reflects the sensitivity of learners to their native speaking interlocutors and their tolerance for errors:

(5) Communicative Blocking of Negative Transfer (CBN)
The transfer of negative properties from L1 to L2 is filtered in proportion to communicative efficiency (principle D): the more an L1 property impedes efficient communication in L2, the less negative transfer there is.

Consider the basic word orders of English and Japanese. These languages have mirror-image word order patterns, head-initial versus head-final, that are equally simple and productive: [went [to [the cinema]]] versus [[[the cinema] to] went], see Greenberg 1966; Dryer 1992; Hawkins 1983, 1994, 2004). Head-final orders are not transferred into L2 English by Japanese learners because, we argue, that would result in extreme communicative inefficiency: speakers using Japanese word orders in English L2 would simply not be understood! By contrast, head-initial word order variants of Spanish that lack precise counterparts in English (e.g., *I read yesterday the*

book) are negatively transferred into L2 English, since they do not impact efficient communication.

We predict that because Japanese is a head-final language, the contrast with the mirror-image word order patterns of English is considerable and transferring head-final patterns into a head-initial language like English, and vice versa, would significantly impair communication. This is why it is imperative for Japanese learners of English, and English learners of Japanese, to acquire correct basic word order in their L2s early. On the other hand, speakers of L1 languages with flexible SVO like Spanish do not have the same incentive, because even when they transfer incorrect orders from their L1s into a fundamentally similar head-initial English L2, communication is not significantly impaired.

Tomasello (2003) makes an important general point of relevance here involving child language acquisition. The driving force behind child language acquisition, he argues, is the need to communicate and make sense of patterns, not an innate language acquisition device. Communication is driven by intention-reading skills whereby interaction is established with the outside world and the essential aim is to convey intentions and understand those of others. The sociocognitive urge to interpret the intentional and mental states of others drives the learning of language. But just as children need to interact, express their thoughts and understand those of others in the L1, so too do speakers acquiring an L2. Interacting socially is central to the acquisition of language. It is this, we believe, that ultimately underlies the communicative blocking of negative transfer (5).

5.5. *The order of second language acquisition*

The cumulative effect of the principles proposed in Sections 4 and 5 can be summarized in principle (6):

(6) Order of Second Language Acquisition (OSLA)
The order of acquisition for properties of the L2 is in accordance with general principles (A)-(D), and with the more specific principles and patterns that are supported empirically. These principles can be incorporated within a multi-factor model of SLA, the CASP model, and used to define possible versus impossible, and likely versus unlikely, interlanguage stages proceeding from a given L1 to a given L2.

These principles operate collectively to make constrained predictions for the acquisition of properties of L2 English and of other languages, and for their relative sequencing. Their interaction is complex, because there are several such principles, which sometimes compete and sometimes cooperate, because they are gradient, and because they have different relative strengths.

Issues of interaction and relative strength have recently been addressed in Hawkins (forthcoming: Ch. 9). He argues that the rules of cooperation and competition between principles in multi-factor models for all areas of language are driven by the same efficiency principles that ultimately explain the principles themselves. In the present context we have seen that specific principle 5 (CBN) will block 4 (PNT) when efficient communication is put at risk, in accordance with general principle D (MaC), as in the conflicting word order data of English and Japanese. Similarly when frequency or infrequency in the input for certain structures and meanings overrides the positive transfer of shared L1-L2 properties in early L2

learning (i.e., when MaF (2) trumps MaPT (1)) this will be because of the greater efficiency of attention to frequency as a learning strategy. The set of L1 properties that is shared with a given L2 is much smaller than the set of L2 properties to which the learner is exposed and which constitute the primary data for the task at hand, namely learning the L2. Hence it is generally more efficient for L2 learning to follow MaF (2) and to be sensitive to L2 input frequencies than to transfer properties, even positively, from the L1. The benefits of following MaF will outweigh the benefits of MaPT in cases of conflict, therefore, especially in the initial stages of learning when many shared properties are infrequently occurring and when there is insufficient evidence for their existence in the L2.

These examples illustrate the important point that the relative strength of competing principles is not arbitrary but reflects the same kind of ease of learning and processing that, we argue, ultimately underlies the principles themselves.

6. Conclusion

In this article we have outlined a general model, CASP, that contains a number of interacting principles that appear to underlie the kinds of evolving interlanguage data that we find in second language learning. We acknowledge that there are undoubtedly more factors that play a role in this process and that contribute to internal variation within L2 learners of different ages, motivation level, teaching background, etc. Our purpose here, however, has been to outline some of the more general linguistic, psycholinguistic and communicative factors that appear to structure our data, and these have been formulated as general and specific principles of the CASP model.

The CASP model has a number of advantages. It incorporates principles of both learning and processing and helps us predict variation patterns within the numerous transfer-related phenomena. Crucially, it shows how, and explains why, different mechanisms can relate to one other and co-operate or compete in the dynamic push-pull of SLA, which numerous previous attempts, focused on fewer principles, have not managed to account for, even though they have collectively drawn attention to a number of the principles that feature in our model. By changing our theoretical perspective from one of opposition and assessment of alternatives (Long 1993) to integration and interaction, we can potentially provide a more realistic model and solve problems that have proved elusive hitherto. The relative weight of competing principles in L2 acquisition can now be observed and measured, using learner corpora, and we can often account for which principle wins. For example, Japanese speakers have to make the effort to master English SVO word order early because if they stick to their native SOV they risk a serious breakdown of communication, which relates to the ultimate goal in language learning: *to understand and make yourself understood*. Further, while some general principles are found across the board regardless of L1 (e.g., all learners tend to minimize learning and processing effort), the CASP model makes many predictions for possible versus impossible, and more versus less likely interlanguages, for different L1-L2 pairs as exemplified in the definite and indefinite article omission data of Tables 1 and 2.

In summary, the model we propose builds on invaluable research that has preceded ours but provides a novel multi-factor perspective with redefined principles and a new methodology for formulating and testing predictions. We hope it will inspire further interdisciplinary research in SLA and bring grammarians, learning and processing theorists, and corpus linguists closer together, as well as opening channels

of communication between theoreticians and practitioners who work on teaching and assessment.

The CASP model presented in Sections 4 and 5 will need to be supplemented in the future by more principles, therefore, including social and pedagogical ones. Their strength and the manner of their interaction with other principles will have to be specified and motivated theoretically. More empirical testing is now needed for different L1-L2 pairs and for the predicted versus unpredicted interlanguage data. Many other grammatical and lexical features will have to be examined beyond those we have discussed here, using more learner data from English as well as learner corpora for other L2s exemplifying various typological mixes of L1 and L2. Other methodologies can also be employed in testing our model, both production and comprehension experiments from psycholinguistics, as well as traditional grammaticality judgment tests of the kind used in generative linguistics. These tests are useful for subtle syntactic phenomena that occur very rarely in actual usage and corpora. Further data testing the predictions of CASP, of the kind we have illustrated here, can then yield deeper insights into the rules of interaction and the relative strength of our principles, which can in turn generate further predictions for empirical studies.

It would also be desirable to set up a computer simulation to examine the possible/impossible and likely/unlikely routes from a given L1 to L2 in the manner of Kirby's (1999) simulation for evolving word order variants in different grammars, which used Hawkins' (1994) efficiency metric for phrase structure processing. By modeling multiple factors rather than just a single factor as in Kirby's study, further insight could be gained into their interaction and relative strength.

Received 16 November 2010
Revised version received
22 March 2012

University of East Anglia
University of Cambridge and
University of California Davis

Note

* We thank Cambridge University Press for permission to include Tables 1 and 2, which we adapted from Hawkins and Buttery (2010). Work on this article was supported by the Leverhulme Trust and Newton Trust postdoctoral research fellowship granted to the first author. The second author was supported by research funds from the University of California Davis. Cambridge ESOL provided generous financial support for the research reported in Hawkins and Filipović (2012), which is cited in this article. All of this support is hereby gratefully acknowledged. We owe our gratitude to two anonymous reviewers and to Ann Kelly, the Assistant Editor of *Linguistics*, whose comments and suggestions helped us improve our article. Any remaining errors are exclusively our own. Correspondence address: Luna Filipović, Language and Communication Studies, Faculty of Arts and Humanities, University of East Anglia, Norwich NR4 7TJ, United Kingdom. E-mail: L.Filipovic@uea.ac.uk

References

Andersen, Elaine S. 1992. Complexity and language acquisition: Influences on the development of morphological systems in children. In John A. Hawkins & Murray Gell-Mann (eds.), *The evolution of human languages*, 241–271. Redwood City, CA: Addison-Wesley.

Andersen, Roger W. 1984. The one-to-one principle of interlanguage construction. *Language Learning* 34. 77–95.

Anderson, Deborah L. 2005. *The acquisition of tough movement in English*. Cambridge: University of Cambridge dissertation.

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.

Atkinson, Dwight. 2002. Toward a sociocognitive approach to second language acquisition. *The Modern Language Journal* 86. 525–545.

Bates, Elizabeth & Brian MacWhinney. 1981.Second language acquisition from a functionalist perspective: Pragmatic, semantic and perceptual strategies. In Harris Winitz (ed.), *Native language and foreign language acquisition*, 190–214. New York: Annals of the New York Academy of Sciences.

Bates, Elizabeth & Brian MacWhinney. 1982. Functionalist approaches to grammar. In Eric Wanner & Lila R. Gleitman (eds.), *Language acquisition: The state of the art*, 173–218. Cambridge: Cambridge University Press.

Bates, Elizabeth & Brian MacWhinney. 1987. Competition, variation and language learning. In Brian MacWhinney (ed.), *Mechanisms of language acquisition*, 157–193. Hillsdale, NJ: Lawrence Erlbaum.

Bowerman, Melissa. 1988. The "no negative evidence" problem: How do children avoid constructing an overly general grammar? In John A. Hawkins (ed.), *Explaining language universals*, 73–101. Oxford: Blackwell.

Bybee, Joan L. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.

Callies, Marcus. 2008. Argument realization and information packaging in tough-movement constructions: A learner-corpus-based investigation. In Danuta Gabrys-Barker (ed.), *Morphosyntactic issues in second language acquisition studies*, 29–46. Clevedon: Multilingual Matters.

Capel, Annette. 2010. A1–B2 vocabulary: Insights and issues arising from the English Profile Wordlists project. *English Profile Journal* 1(1). http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=7908266 (accessed 3 February 2012).

Capel, Annette. 2012. Getting to B2: Language skills for Cambridge *First*. ETAS conference presentation. http://www.e-tas.ch/downloads/events/28_D4.pdf (accessed 19 March 2012).

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Clahsen, Harald. 1984. The acquisition of German word order: A test case for cognitive approaches to L2 development. In Roger W. Andersen (ed.), *Second languages: A cross-linguistic perspective*, 219–242. Rowley, MA: Newbury House.

Cross, John & Szilvia Papp. 2008. Creativity in the use of verb+noun combinations by Chinese learners of English. In Gaëtanelle Gilquin, Szilvia Papp & Maria Belen Diez Bedmar (eds.), *Linking up contrastive and learner corpus research*, 57–81. Amsterdam: Rodopi.

Culicover, Peter W. 1999. *Syntactic nuts: Hard cases, syntactic theory and language acquisition*. Oxford: Oxford University Press.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam & Philadelphia: John Benjamins.
Davies, Alan. 1991. *The native speaker in applied linguistics*. Edinburgh: Edinburgh University Press.
Diessel, Holger. 2004. *The Acquisition of Complex Sentences*. Cambridge: Cambridge University Press.
Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68. 81–138.
Eckman, Fred. 1984. Universals, typologies, and interlanguage. In William Rutherford (ed.), *Language universals and second language acquisition*, 79–105. Amsterdam & Philadelphia: John Benjamins.
Ellis, Nick C. 1998. Emergentism, connectionism and language learning. *Language Learning* 48(4). 631–664.
Ellis, Nick C. & Diane Larsen-Freeman. 2009. Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage. In Nick C. Ellis & Diane Larsen Freeman (eds.), *Language as a Complex Adaptive System*, 90–125. Chichester: Wiley-Blackwell.
Ellis, Rod. 2003[1994]. *The study of second language acquisition.* Oxford: Oxford University Press.
Francis, Elaine J. 2010. Grammatical weight and relative clause extraposition in English. *Cognitive Linguistics* 21. 35–74.
Gass, Susan. 1987. The resolution of conflicts among competing systems: A bidirectional perspective. *Applied Psycholinguistics* 8. 329–350.
Gass, Susan. 1989. How do learners resolve linguistic conflicts? In Susan Gass & Jacquelyn Schachter (eds.), *Linguistic perspectives on second language acquisition*, 183–199. Cambridge: Cambridge University Press.
Gass, Susan & Larry Selinker, L. 2008. *Second language acquisition: An introductory course.* London: Routledge.
Gell-Mann, Murray. 1992. Complexity and complex adaptive systems. In John A. Hawkins & Murray Gell-Mann (eds.), *The evolution of human languages*, 3–18. Redwood City, CA: Addison-Wesley.
Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68. 1–76.
Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. In Yasushi Miyashita, Alec Marantz & Wayne O'Neil (eds.), *Image, language, brain*, 95–112. Cambridge, MA: MIT Press.
Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–113. Cambridge, MA: MIT Press.
Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the second meeting of the North American chapter of the Association for Computational Linguistics (NAACL), v*ol. 2. 159–166.
Hauser, Mark, Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298. 1569–1579.
Hawkins, John A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. Atlantic Highlands, NJ: Humanities Press & London: Croom Helm.
Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.
Hawkins, John A. 1986. *A comparative typology of English and German: Unifying the contrasts*. Austin, TX: University of Texas Press & London: Routledge.

Hawkins, John A. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21. 223–261.
Hawkins, John A. 1991. On (in)definite articles: Implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27. 405–442.
Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
Hawkins, John A. 1999. Processing complexity and filler-gap dependencies. *Language* 75. 244–285.
Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
Hawkins, John A. 2007. Acquisition of relative clauses in relation to language universals. *Studies in Second Language Acquisition* 29. 337–344.
Hawkins, John A. 2009. An efficiency theory of complexity and related phenomena. In David Gil, Geoffrey Sampson & Peter Trudgill (eds.), *Complexity as an evolving variable*, 252–268. Oxford: Oxford University Press.
Hawkins, John A. 2011. Discontinuous dependencies in corpus selections: Particle verbs and their relevance for current issues in language processing. In Emily M. Bender & Jennifer E. Arnold (eds.), *Language from a cognitive perspective: grammar, usage and processing*, 269–290. Stanford, CA: CSLI Publications.
Hawkins, John A. forthcoming. *Cross-linguistic variation and efficiency*. Oxford: Oxford University Press.
Hawkins, John A. (ed.). 1988. Explaining language universals. Oxford: Blackwell.
Hawkins, John A. & Paula Buttery. 2009. Using learner language from corpora to profile levels of proficiency: Insights from the English profile programme. In Linda Taylor & Cyril J. Weir (eds.), *Language testing matters: Investigating the wider social and educational impact of assessment*, 158–175. Cambridge: Cambridge University Press.
Hawkins, John A. & Paula Buttery. 2010. Criterial features in learner corpora: Theory and illustrations. *English Profile Journal* 1(1). http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=7908278&fulltextType=RA&fileId=S2041536210000103 (accessed 16 February 2012).
Hawkins, John A. & Luna Filipović. 2012. *Criterial features in L2 English*: *Specifying the reference levels of the common European framework*. Cambridge: Cambridge University Press.
Hawkins, John A. & Murray Gell-Mann (eds.). 1992. *The evolution of human languages* (Santa Fe Institute Studies in the Sciences of Complexity: Proceedings Volume XI.). Reading, MA: Addison-Wesley.
Hawkins, Roger. 2001. *Second language syntax: A generative introduction*. Oxford: Blackwell.
Hoekstra, Teun & Jan G. Kooij. 1988. The innateness hypothesis. In John A. Hawkins (ed.), *Explaining language universals*, 31–55. Oxford: Blackwell.
Kawaguchi, Satomi. 1999. The acquisition of syntax and nominal ellipsis in JSL discourse. In Peter Robinson (ed.), *Representation and process* (Proceedings of the Third Pacific Second Language Research Forum 1), 85–93. Tokyo: Pacific Second Language Research Forum.
Kawaguchi, Satomi. 2002. Grammatical development in learners of Japanese as a second language. In Bruno di Biase (ed.), *Developing a second language: Acquisition, processing and pedagogy issues in Arabic, Chinese, English, Italian, Japanese, and Swedish*, 17–28. Melbourne: Language Australia.

Kellerman, Eric. 1983. Now you see it, now you don't. In Susan Gass & Larry Selinker (eds.), *Language transfer in language learning*, 112–134. Rowley, MA: Newbury House.
Kirby, Simon. 1999. *Function, selection and innateness*. Oxford: Oxford University Press.
Konieczny, Lars. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29(6). 627–645.
Krashen, Stephen D. 1982. *Principles and practice in second language acquisition*. Oxford: Pergamon.
Lado, Robert 1957. *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor, MI: University of Michigan.
Larsen-Freeman, Diane. 1997. Chaos/complexity science and second language acquisition. *Applied Linguistics* 18(2). 141–165.
Larsen-Freeman, Diane & Lynne Cameron. 2008. *Complex systems in applied linguistics*. Oxford: Oxford University Press.
Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106. 1126–1177.
Long, Michael H. 1993. Assessment strategies for second language acquisition theories. *Applied Linguistics* 14(3). 225–249.
MacWhinney, Brian. 1992. Transfer and competition in second language learning. In Roger Harris (ed.), *Cognitive processing in bilinguals*, 371–390. Amsterdam: Elsevier.
MacWhinney, Brian. 2005. A unified model of language acquisition. In Judith F. Kroll & Annette M. B. de Groot (eds.), *Handbook of Bilingualism: Psycholinguistic Approaches*, 49–67. Oxford: Oxford University Press.
McDonald, Janet & Kathy Heilenman. 1991. Determinants of cue strength in adult first and second language speakers of French. *Applied Psycholinguistics* 12. 313–348.
Mellow, J. Dean. 2008. The emergence of complex syntax: A longitudinal case study of the ESL development of dependency resolution. *Lingua* 118. 499–521.
Müller, Ralph-Axel. 2009. Language universals in the brain: How linguistic are they? In Morten H. Christiansen, Chris Collins & Shimon Edelman (eds.), *Language universals*, 224–252. Oxford: Oxford University Press.
Müller-Gotama, Frans. 1994. Grammatical relations: A cross-linguistic perspective on their syntax and semantics. Berlin & New York: Mouton de Gruyter.
O'Grady, William. 2005. *Syntactic carpentry: An emergentist approach to syntax*. Mahwah, NJ: Lawrence Erlbaum.
O'Grady, William. 2008. The emergentist program. *Lingua* 118. 447–464.
Pienemann, Manfred. 1998. *Language processing and second language development: Processability theory*. Amsterdam & Philadelphia: John Benjamins.
Pienemann, Manfred. 2003. Language processing capacity. In Catherine Doughty & Michael H. Long (eds.), *Handbook of second language acquisition theory and research*, 679–714. Oxford: Blackwell.
Pienemann, Manfred. 2005. Processing constraints on L1 transfer. In Judith F. Kroll & Annette M. B. de Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 128–153. Oxford: Oxford University Press.
Postal, Paul. 1974. *On raising*. Cambridge, MA: MIT Press.
Preston, D. 1989. *Sociolinguistics and second language acquisition*. Oxford: Blackwell.

Radford, Andrew. 1990. *Syntactic theory and the acquisition of English syntax.* Oxford: Blackwell.

Reali, Florencia & Morten Christiansen. 2009. On the necessity of an interdisciplinary approach to language universals. In Morten H. Christiansen, Chris Collins & Shimon Edelman (eds.), *Language universals*, 266–277. Oxford: University Press.

Rutherford, William. 1983. Language typology and language transfer. In Susan Gass & Larry Selinker (eds.), *Language transfer in language learning*, 358–370. Rowley, MA: Newbury House.

Seliger, H. 1989. Semantic transfer constraints on the production of English passives by Hebrew-English bilinguals. In Hans Dechert & Manfred Raupach (eds.), *Transfer in language production*, 21–33. Norwood, NJ: Ablex.

Slobin, Dan I. 1973. Cognitive prerequisites for the development of grammar. In Charles A. Ferguson & Dan I. Slobin (eds.), *Studies of child language development*, 175–208. New York: Holt, Rinehart & Winston.

Slobin, Dan I. 1977. Language in childhood and in history. In John Macnamara (ed.), *Language learning and thought* (Perspectives in neurolinguistics and psycholinguistics), 185–214. New York: Academic Press.

Slobin, Dan I. 1985. Crosslinguistic evidence for language-making capacity. In Dan I. Slobin (ed.), *The crosslinguistic study of language acquisition*, vol. 2, 1157–1256. Hillsdale, NJ: Lawrence Erlbaum

Stabler, Edward P. 2009. Computational models of language universals: Expressiveness, learnability, and consequences. In Morten H. Christiansen, Chris Collins & Shimon Edelman (eds.), *Language universals*, 200–223. Oxford: Oxford University Press.

Strömqvist, Sven, Åke Hellstrand & Åsa Nordqvist. 1998. From child speech to literary art: On language development and descriptions of motion. In Hrafnhildur Ragnarsdóttir & Sven Strömqvist (eds.), *Learning to talk about time and space*, 203–234. Reykjavík & Göteborg: Kompendiet.

Tomasello, Michael. 2000. Do young children have adult syntactic competence? *Cognition* 74. 209–253.

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Trévise, Anne.1986. Is it transferrable, topicalization? In Eric Kellerman & Michael Sharwood Smith (eds.), *Crosslinguistic influence in second language acquisition*, 186–206. New York: Pergamon.

VanPatten, Bill. 1996. *Input processing and grammar instruction*. Norwood, NJ: Ablex

Wasow, Thomas. 2002. *Postverbal behavior*. Stanford: CSLI Publications.

White, Lydia. 2003. *Second language acquisition and universal grammar*, Cambridge: Cambridge University Press.

Williams, Caroline. 2007. A preliminary study into verbal subcategorisation frame usage in the CLC. Unpublished ms., Research Centre for English and Applied Linguistics (RCEAL), University of Cambridge.