Statistical Learning in Second Language Acquisition

Duygu Bayram

Abstract

Language is a uniquely human ability (Anderson, 2008). This bears some questions about the nature and the mechanisms of language and language acquisition. Traditionally, Universal Grammar (UG) has been the mainly accepted framework to explain the process of language acquisition; however, statistical learning has started to regain some popularity. In light of this, the role of statistical learning in second language acquisition is an existing question. This paper gives a brief overview of some existing literature and presents a speculative discussion on the issue.

1 Introduction

The roots and the mechanisms of human language is an interesting but difficult question to tackle, especially considering it seems to be one uniquely human quality that animals seem to be unable to perform so far (Anderson, 2008).

To start with the very basics, in an evolutionary sense, language is a surprising formation. When evolution is mentioned, most of us tend to think of it as a process of improvement on existing skills. This is the most commonly taught type of evolutionary change at schools, where there is a gradual, incremental, and slow change based on natural selection. However, there is another type of evolutionary change called punctuated equilibrium, which describes a rapid process of change in a species that is usually preceded and followed by long periods of little to no change. There has been some ideas that language, or something that allows for the emergence of language in humans, could potentially be an example of this type of evolutionary change (Bowern, 2006).

So what could be this change, or what could be the separate elements of the change, and how does it tie into our understanding of language? For the purposes of this paper, I will investigate first and second language acquisition, consider these in terms of universal grammar and statistical learning, and speculate where compositional generalization may come into the picture.

1.1 Universal Grammar

Historically, the idea of statistical learning has been proposed to explain the process of learning a language but somewhat went into the background when Chomsky put forward his ideas about the Universal Grammar.

Universal Grammar (UG) can be summarized as a set of innate constraints that all languages shape around. In other words, there are certain principles (innate constraints) and parameters (on-off switches for certain grammatical properties such as pro-drop which either allows or does not allow a language to drop nouns from sentences).

One of the main arguments of UG is the poverty of stimulus, referring to the idea that children learn considerable linguistic information when the input is limited and not explained very clearly. This idea went against the popular assumption that children learn language through experience at the time, similar to a simple and pure statistical learning view. The fact that children do not receive enough data to learn solely through experience is still a strong argument.

UG has received enthusiastic support since then, much of research being built on it and confirming its proposals. Two commonly known examples of this are the cases of feral children relating

to the later developed Critical Period Hypothesis in support of UG, and the cases of hard of hearing communities inventing sign language among themselves (Senghas et al., 2001).

1.2 The Critical Period Hypothesis

The Critical Period Hypothesis (Lenneberg, 1967) (CPH) has also been an important finding which suggests strong support for Universal Grammar and serves as an important concept for our main topic, Second Language Acquisition. There has been several findings that children who have been isolated from linguistic input fail to acquire full linguistic competence when they are exposed to it later in life (Curtiss & Whittaker, 2014). Perhaps the most famous example is Genie, who has been found locked up in a dark room at 13 years old. Up until that point she had been severely neglected and abused, and was deprived of any linguistic input. While she showed some improvements and succeeded in learning words, she seemed unable to use grammar. This supported the notion that there is an innate grammar required for human language capacity, and that this was somewhat biological and dependent on age and maturation (Lenneberg, 1967).

However, it is important to note that the findings on such language impairment could be due to the general mistreatment and neglect these isolated children grow up with. It is well-know that children in these conditions suffer from cognitive impairments (Bücker, 2012), the demonstrated inability to reach linguistic competence could be dependent on this fact. Similarly, isolated deaf children are often unable to access education, resulting in the same unclarity around the findings.

Regardless, traditionally, language acquisition has mostly been thought of in terms of UG and the critical period. First language acquisition is characterized as a blank initial state, in other words, there is no existing grammar specific to a language, but an innate set of UG constraints (White 2003). From this starting point, children then go through a series of developmental stages where they acquire the right parameter settings based on limited input from their environment. During this developmental phase, all grammatical states children go through fit within UG constraints, and children then reach a "steady state grammar" where they become native speakers of the language.

2 Theories of Second Language Acquisition

When it comes to second language acquisition (SLA), the biggest difference is that the learners already have certain parameter settings in place because they already speak a language. The question is then, are they only able to learn within the constraints of these settings? Do they still have access to UG to be able to reset the parameters? Where does the critical period come in? How much does the first language aid and disrupt the SLA process?

According to White (2003) SLA can be investigated under three subcategories: the initial state, stages of development, and the final state. She also explains that this process can be thought of in terms of the involvement of the first language (L1) and UG. The involvement of L1 is coined as transfer, describing the transfer of the grammatical properties of L1 to the second language (L2) grammar during the learning process.

Based on these, the main theories of second language acquisition are as follows (White 2003):

1. **Full Transfer/Partial Access:** L1 has full involvement, and there is only partial access to UG in which the only available constraints of UG are the ones used in the acquisition of L1.

2. **No Transfer/Full Access:** L1 has no involvement, the learner acquires L2 much the same way as they acquire their L1, through full access to UG only.

3. **Full Transfer/Full Access:** L1 has full involvement and can serve as a filter, it is used as the initial referral point during the acquisition process, but the learner is able to access UG to reset parameters if the input does not make sense within UG constraints with L1's parameter settings.

4. **Partial Transfer/Full Access:** Learner interprets input through both L1 and UG, instead

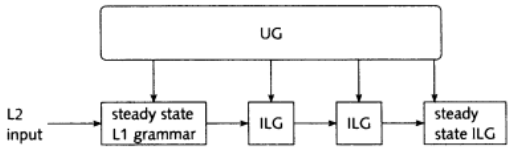

Figure 1: Figure 4.4, Full Transfer/Full Access, White (2003).

of using L1 as a filter first, and not all parameters of L1 are used.

5. **Partial Transfer/Partial Access:** Similar to the Partial Transfer/Full Access view, the input is interpreted through both L1 and UG, but not all parts of L1 are transferred, and not all constraints of UG are accessible.

For the purposes of this paper, theories 3-5 will be detailed further.

2.1 Full Transfer/Full Access

The initial state is characterized as the final state of L1, pointing to full transfer of L1 grammatical properties (White 2003). All UG constraints are available for full access, not limited to the ones that were used in L1. To summarize this framework, when given input, the learner first resorts to making sense of it through the available L1 grammar. When this leads to error, the learner then resets parameters for the L2 grammar based on UG constraints. In this sense, L1 acts as a filter during the acquisition process. In cases where the L2 input does not exactly follow the same parameters of L1, but fails to raise enough alarm to reset parameters, an interlingual grammar forms, and fossilization is possible. As the final state, the learner may be able to reach a native-like level, but may also fall short of native competency, consistent with conflicting findings on this issue (White 2003).

2.2 Partial Transfer/Full Access

The main difference of this framework from Full Transfer/Full Access is that it proposes not all properties of L1 is used for the initial state. Vainikka Young-Scholten (1994, 1996a, 1996b) find that only lexical categories are transferred, but functional categories are not, and UG is accessed

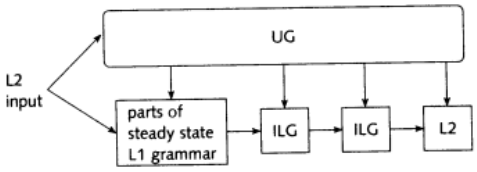

Figure 2: Figure 4.5, Partial Transfer/Full Access, White (2003).

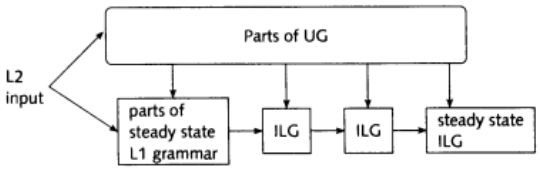

Figure 3: Figure 4.5, Partial Transfer/Partial Access, White (2003).

to project these categories later on. An issue with this view is that according to the model, the learner should converge on the L2 grammar (White 2003). However, as mentioned, many (if not most) learners fail to achieve this.

White (1996) also argues against the view that functional categories are not accessible, and instead asserts that certain functional projections are acquired through UG when they are not available in L1, such as French clitics for L1 English speakers (White 2003).

2.3 Partial Transfer/Partial Access

White (2003) summarizes this framework by the views of Beck (1997) and Eubank et al. (1997) who propose that some functional categories are never specified, resulting in an impairment in the local domain. This would mean that L2 speakers never reach native competence. However, results are conflicting on this, for example, Montrul Slabakova (2001) found no significant difference between the competencies of natives and L2 learners in the domain of aspects for L1 English - L2 Spanish speakers.

So far, this is the traditional framework around SLA in terms of UG. Recently, with the developments in machine learning, linguists have started to take the potential role of statistical learning into consideration again, sparking a debate around the

validity of the Universal Grammar. Recent work has shown that transitional probabilities play an important role in learning word segmentation for children (Saffran et al., 1996).

3 Criticism of the Critical Period Hypothesis in Second Language Acquisition

Vanhove (2013) puts forward a developed critical view on the topic. Taking a closer look at the role of CPH in second language acquisition (SLA), he highlights the common problems that show up in the field which create ambiguous and unsystematic research questions and methods.

To start with, according to Vanhove (2013), the cutoff point seems to be unclear, different researchers choosing different age periods. Moreover, he points out that the conditions in which CPH holds are undefined in SLA, resulting in questions around whether this hypothesis is accepted for only immersion settings, or whether it includes conditions of explicit acquisition in schools. Another issue is the lack of clarification around the specific areas of language that it encapsulates, whether it is phonology, syntax, morphology, or other areas. Finally, he criticizes the lack of consideration of learning rate in regards to CPH and the over-focus on the final state, stating that adult learners show faster progress in the earlier stages of learning compared to children.

Further, he claims the expected outcomes of CPH in SLA are undefined, resulting again in poor research methodology. As SLA research relies heavily on the final state, accepting the CPH would mean a learner past the CPH age would never obtain native-like linguistic competence. Vanhove (2013) points out that one case of an individual who has succeeded in this would then disprove this idea; however, due to the unclarity around what native-like competence looks like, the discussion continues. He argues that motivated by a desire to support CPH, the competence level that is expected from learners keeps increasing in the eyes of the researchers, and that these bilingual subjects are tested against monolingual speakers, a difference that could explain some performance discrepancies.

Finally, he investigates the proposed non-linear CPH function that takes the final state and the age of acquisition as variables, by reanalyzing some papers making the proposition. He argues against the popular notion that the negative relationship between age of acquisition and the final competence level supports CPH. Instead, he claims CPH would be supported by not looking at the ultimate attainment, but by looking at changes in learning and success rate in regards to age. This way, research would not have to look for native-like competence, but learning rates instead, and L2 learners could be tested among themselves rather than against monolingual native speakers.

Throughout the paper, he re-performs statistical analysis on published papers, looking at them in light of the proposed "flattened slope" and demonstrates that the papers are not using the correct statistical methods for assessing this prediction. Further, he argues that the results of his analysis do not necessarily support CPH. He finds that there is a breakpoint for a cutoff age as predicted, but also shows that a completely linear alternative is well-within the confidence interval, arguing that it is negligible.

4 Statistical Learning and Universal Grammar

4.1 The Variational Model

In response to this, Yang (2004) experiments with the idea of combining both ideas. In his paper, he argues and demonstrates that UG serves as a set of constraints on a probabilistic learning space, which allows for statistical learning to be successful in face of limited input. He also proposes that the "switch" framework should be thought of in terms of probabilities rather than a binary triggering mechanism.

Yang (2004) presents a computational model developed on Saffran et al. (1996)'s SLM model, trained on real world child directed speech data, finding that the model performs poorly on real world data. This supports his argument that even with statistical learning, children need to have certain constraints in place to know what they are learning, limiting the learning space.

In light of this, he proposes the idea that children try to learn through the constraints first (such as syllable having only one stress) and then resorts to statistical learning when the available constraints are not enough to make sense of the input, and suggests that several findings support this as models have shown children prefer prosody over statistical information (Johnson & Jusczyk, 2001).

He also tackles the pure-UG view, especially in context of parameter setting, which is where the paper becomes more relevant for SLA.

Yang (2004) proposes a variational learning model as opposed to the parameter setting framework that is traditionally accepted under the umbrella of UG. First, he points out that convergence cannot be guaranteed, as explained above in the SLA section as well. Second, similar again to the SLA frameworks, he explains that the triggering framework requires that at any given time, the grammar needs to fall under possible manifestations of UG. Third, he also points out that if this triggering property were true, there would be an immediate switch in children's use of grammar, rather than a gradual change.

In their variational model, they form a probabilistic learning space, but with UG constraints, where certain grammars are rewarded and punished to be selected. In this model, the target model is selected for gradually as the input eliminates other potential grammars. They verify this model with a computational model on null-subjects.

5 Statistical Learning in Second Language Acquisition

The concept of statistical learning in linguistics is not necessarily new; however, it has been somewhat disregarded and forgotten until recently due to the strong support found for Universal Grammar.

Prior to recent developments in computation, most statistical learning studies have been done on artificial languages but there was some resistance around their relevance to natural language acquisition (Onnis, 2012).

Onnis (2012) also explains that there have been findings that show that linguistic processing and statistical learning of sequences are potentially performed by the same mechanisms (Misyak Christiansen, 2010).

Going off of these, Onnis (2012) identifies four learning principles that may be relevant for SLA:

1. Integrate probabilistic sources of information
2. Seek invariance
3. Reuse learning mechanisms
4. Learn to predict

5.1 Integrate probabilistic sources of information

As also exemplified by Yang (2004), this principle proposes that statistical information aids language acquisition, and language input carries significant statistical cues. This is especially studied in terms of speech (Saffran et al., 1996; Dell et al., 2000) and phonetics phonotactics (Thiessen, 2007; Chambers et al., 2003).

5.2 Seek invariance

Onnis (2012) summarizes that certain features of inflectional morphology show to be difficult to acquire for L2 learners (Montrul et al., 2008; Slabakova, 2008). He argues that while this has been discussed in terms of L1 access (as described in the SLA frameworks above), there has not been enough studies focusing on potential statistical learning explanations. He reminds the artificial languages example, where learners are able to recognize certain hidden rules by repeated exposure, arguing that this kind of statistical sensitivity may be important in the difficulty and ease of acquisition of certain features. He also gives an example from Gomez (2002) who showed that variability of the intervening words used in her study predicts sensitivity.

5.3 Reuse learning mechanisms

Onnis (2012) proposes that learners reuse the same mechanisms to achieve performance on different tasks. To support this, he gives an example from Onnis et al. (2011) where adults are presented

with picture-word pairings where the meaning mappings are ambiguous. They find that when a certain pair is repeated across trials the disambiguation is more successful. In terms of meaning mapping, this information may hold some value for aiding SLA.

5.4 Learn to predict

In this principle, Onnis (2012) points out that corpus analyses have shown that words hold probabilistic information for the following word. This concept is frequently exploited in the field of Natural Language Processing, there is value in considering that a similar process may play a role for human language acquisition.

6 Discussion

6.1 Statistical Learning in Second Language Acquisition

For my speculative discussion I would like to return to White (2003)'s summary of SLA theories. The conflicting findings on acquisition and the initial state at times point to Partial Transfer and Partial Access theories. However, as White (2003) also mentions, it is important to note that studying the initial state empirically is quite difficult as the starting point of experiments may not be the actual initial state. The mentioned conflicting findings could be caused by the difficulty of measuring this state of language to begin with.

Furthermore, when we look at their predictions about the final state, they require that the learner does not converge and reach native-like competence. However, there are findings against this where some people, albeit possibly not common, achieve this level. As detailed in Section 3, Vanhove (2013) puts forward stronger arguments against this strong reliance on the final state, suggesting that the research might benefit more from focusing on learning rate instead, which could potentially point to more involvement of statistical learning in the L2 acquisition process. As mentioned, if the learner rate beings high in adult learners in the earlier stages and later slows down, rather than being an evidence for the critical period, this could point to some learning strategy switches. As an example, adults could be initially approaching the task more from an analytical view, possibly relying more on compositional generalization, and tending towards explicit learning. This could potentially show a stronger reliance on statistical learning; however, it also does not discredit access to UG, as the potential use of compositional generalization may require some syntactic constraints, and statistical learning itself benefits from these constraints, as supported by Yang (2004).

Furthermore, the Full Transfer/Full Access view puts forward a filtering idea, where L1 serves as an initial barrier before UG is accessed. In light of statistical learning and the variational model, there might be value in revisiting this approach.

It could be investigated whether the L1 filter effect is related to probabilistic learning, as the parameters set from L1 will have high probabilities in the learning space, acting as a potential barrier for the incoming input. However, if the input cannot be reconciled, then different settings may be considered under the constraints of UG. This would both allow for fossilization, as the input may not conflict enough for reconstruction, as supported by the initial framework, but it could also allow for reaching native competency depending on the language pairs, and perhaps the variance in people's openness to reconsideration of probabilities. It is also valuable to consider that this variance in speakers could also be dependent on above discussed learning strategies, whether they prefer implicit or explicit approaches, and to what extent they are allowed to indulge in these preferences considering that modern SLA instruction often tends towards encouraging implicit acquisition, which may not be suited for adult learners.

Another point is, as reviewed, most of the studies on statistical learning with respect to language acquisition seem to be related to speech and auditory input. It is also important to note here that while children acquire language largely through auditory input, L2 learners often have access to more mediums than this, and arguably they make use of these more than auditory input, which could further explain some of these age of acquisition differences. To this end, experiments similar to Yang (2004), inspecting multimodal statistical learning approaches may be beneficial in making compar-

isons. Such studies could also take into consideration the learning principles proposed by Onnis (2012). It is also an interesting open question what similar studies would show on grammatical aspects of language rather than phonological, as that might give us more insight on UG and CPH.

As a final note, it is important to take the critical period into consideration here and not disregard it completely. It could be that there is a biological limit to our access to UG, this could surely explain the partial access accounts of SLA theories. However, it could also be that the UG constraints are not the cognitive loss related to age in these cases, but rather the losses in the overall learning domain that allows for statistical learning are to blame for the observed impairments. After all, these children are often abused and neglected, conditions which cause some decrease cognitive abilities, as is aging. The idea that this loss of ability to learn language due to impairments in statistical learning rather than a loss of access to UG could be an interesting research question.

6.2 Compositional Generalization

In the context of this seminar, I would also like to briefly discuss where compositional generalization may fit into this topic. Simply defined, compositional generalization is the idea that meaning is derived from form. We understand and construct meaning from the syntax of the language, making semantics and syntax inherently interconnected.

In terms of language evolution, this is an interesting concept, especially with the questions of whether meaning or form began first, whether they can exist independently, or which aspect relies on the other. Additionally, another question is whether compositional generalization is a global cognitive skill that is employed in other areas of learning, or whether it is one of the aspects of language that render it unique. Assuming it is a skill used in other areas of learning, and knowing its reliance on syntax in language, it is worth examining if it could potentially be an overlapping point for UG-based learning and statistical learning.

Moreover, it is an interesting question how this interaction between meaning and syntax affects language acquisition. An idea is that in consideration of compositional generalization, I would almost assume L2 would be easier to acquire as we already have access to a semantic representation and we simply have to learn a new way to formalize it. This could be an explanation for why adults tend to rely on analytical approaches to new languages rather than implicit acquistion. Though, evidently, the process of language acquisition for adults is not this simple, potentially suggesting that there is still some need to access UG to construct the new syntax.

7 Conclusion

To conclude, the involvement of statistical learning in the acquisition of second language seems to be an open question. Traditionally, UG was the main point of explanation for theories of language acquisition, with L1 playing some role in SLA theories. The frameworks for L2 acquisition were mainly shaped around how much transfer from L1 the learner performed, and how much they could access UG. Furhtermore, the partial access theories mainly assumed a critical period, explaining why there might be a loss of access to UG as the learner ages, and why some learners do not converge and achieve native-like competence.

However, the Critical Period Hypothesis would benefit from a revisit in light of the reconsideration of statistical learning in language acquisition. It is possible that certain affects previously attributed to CPH could be explained by statistical learning models, which still allow for frameworks that suggest full access to UG. New frameworks that may be developed from these understandings could potentially explain the conflicting findings in the outcomes of L2 acquisition in terms of convergence.

It may also be valuable to investigate multimodal statistical learning models for L2 acquisition, considering that these experiments are so far largely performed with children and mainly auditory input in mind. However, L2 learners often have access to many more mediums and do not learn from primarily audio input.

Moreover, experiements that take compositional generalization into consideration may be important. This could potentially explain the analytical learning strategies adults tend to employ. It

could further give insight into overlapping areas of UG and statistical learning as it involves analytical components while relying heavily on syntactic information when it comes to language input.

The question of where statistical learning may fit into L2 acquisition remains open. It has many directions to take, and certain traditional theories may benefit from revisits. Current unclarities and conflicting findings in the area could be explained by the involvement of these approaches.

8 References

Anderson, S. R. (2008). The Logical Structure of Linguistic Theory. Language, 84(4), 795–814.

Bowern, Claire. (2006). Punctuated Equilibrium and Language Change. Encyclopedia of Language Linguistics. 10.1016/B0-08-044854-2/01870-8.

Bücker, J., Kapczinski, F., Post, R., Ceresér, K. M., Szobot, C., Yatham, L. N., ... Kauer-Sant'Anna, M. (2012). Cognitive impairment in school-aged children with early trauma. Comprehensive Psychiatry, 53(6), 758-764.

Chambers, K.E., Onishi, K.H., Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. Cognition, 87, B69-B77.

Curtiss, S., Whittaker, H. (2014). Genie: A Psycholinguistic Study of a Modern-Day Wild Child.

Dell, G.S., Reed, K.D., Adams, D.R., Meyer, A.S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. Journal of Experimental Psychology: Learning, Memory, Cognition, 26, 1355-1367.

Gomez, R.L. (2002). Variability and Detection of Invariant Structure. Psychological Science, 13, 5, 431-436.

Johnson, E.K. and Jusczyk, P.W. (2001) Word segmentation by 8-month-olds: when speech cues count more than statistics. J. Mem. Lang. 44, 1–20.

Lenneberg E 1967 Biological foundations of language. John Wiley New York.

Misyak, J. B., Christiansen, M. H. (2011). Statistical Learning and Language: An Individual Differences Study. Language Learning, 62(1), 302–331.

Montrul, S., Foote, R., Perpiñán, S. (2008). Gender agreement in adult second language Learners and Spanish heritage speakers: The effects of age and context of acquisition. Language Learning, 58, 3, 503-553.

Montrul, S.A., Slabakova, R. (2001). Is Native-like Competence Possible in L2 Acquisition?

Onnis, L., Edelman, S., Waterfall, H. (2011). Local statistical learning under crosssituational uncertainty. In L. Carlson, C. Hölscher and T. Shipley (Eds.). Proceedings of the 33rd Annual Conference of the Cognitive Science Society.

Onnis, L. (2012). The potential contribution of statistical learning to second language acquisition. Statistical Learning and Language Acquisition, 203–236.

Saffran, J.R. et al. (1996) Statistical learning by 8-month old infants. Science 274, 1926–1928.

Slabakova, R. (2008). Meaning in the second language. Berlin: Mouton de Gruyter.

Senghas, A., Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. Psychological science, 12(4), 323-328.

Thiessen, E.D. (2007). The effect of distributional information on children's use of phonemic contrasts. Journal of Memory and Language, 56, 16-34.

Vainnikka, A. and Young-Scholten, M. (1996a) The early stages of adult L2 syntax: additional evidence from Romance speakers. Second Language Research 12: 140-76.

Vainnikka, A. and Young-Scholten, M. (1996b) Gradual development of L2 phrase structure. Second Language Research 12: 7-39.

Vainnikka, A. and Young-Scholten, M. (1994) Direct access to X'-theory: evidence from Korean and Turkish adults learning German. In T. Hoekstra and B. Schwartz (eds.), Language Acquisi-

tion Studies in Generative Grammar, pp. 265-316. Amsterdam: John Benjamins.

Vanhove, J. (2013). The critical period hypothesis in second language acquisition: A statistical critique and a reanalysis. PloS one, 8(7), e69172.

White, Lydia. Second language acquisition: From initial to final state. Second language acquisition and linguistic theory (2003): 130-155.

White, Lydia. Clitics in French. In H. Clahsen (ed.), Generative Perspectives on Language Acquisition: Empirical Findings, Theoretical Considerations, Crosslinguistic Comparisons, pp. 335-68. Amsterdam: John Benjamins.

Yang, Charles D. Universal Grammar, statistics or both? Trends in cognitive sciences 8, no. 10 (2004): 451-456.