second
language
research

Second Language Research
27(3) 289–311

DOI: 10.1177/0267658310384567
slr.sagepub.com
SAGE

Straight on through to Universal Grammar: Spatial modifiers in second language acquisition

David Stringer, Beatrix Burghardt, Hyun-Kyoung Seo, Yi-Ting Wang
Indiana University, IN, USA

Abstract

There has been considerable progress in second language (L2) research at the syntax–semantics interface addressing how syntax can inform phrasal semantics, in terms of interpretive correlates of word order (Slabakova, 2008). This article provides evidence of a flow of information ostensibly in the opposite direction, from meaning to grammar, at the interface between lexical semantics and syntax. It is argued that there is a functional hierarchy of modifiers in the domain of adpositions, which enables the linguistic elaboration of trajectories, but that not all languages lexicalize all types. This study examines whether L2 learners of English are able to overcome the poverty of the stimulus and recruit the relevant functional categories despite their absence in the first language (L1). Modifiers were taught to learners individually, but never in combination. A computer-animated narrative was designed in order to create felicitous contexts for combinations of modifiers, and preference and grammaticality judgment tasks were administered to 121 students from various L1 backgrounds, as well as 20 native speakers. Accuracy scores were remarkably targetlike on binary combinations of modifiers (1) across proficiency levels, (2) across L1s, and (3) across the two tasks, revealing that with the semantics of modifiers in place, the syntactic hierarchy is naturally manifested.

Keywords

prepositions, lexical semantics, modifiers, motion events, Universal Grammar

I Introduction

Recent work on the syntax–semantics interface in second language (L2) acquisition has made considerable progress in identifying the influence of syntax on phrasal semantics,

Corresponding author:
David Stringer, Department of Second Language Studies, Indiana University, Memorial Hall 310, Bloomington, IN 47405, USA
Email: ds6@indiana.edu

in terms of semantic correlates of word order (for an overview, see Slabakova 2008). Once the syntax has been acquired, certain aspects of semantic interpretation appear to be already in place. The question remains as to whether there may be a comparable flow of information in the opposite direction, such that knowledge of semantics may carry syntactic implications. The current study addresses the question of whether lexical semantics can inform L2 syntax in the domain of prepositional modifiers. Just as there are robust universal orderings of adjectival modifiers in the nominal domain (Cinque, 1994; Crisma, 1996; Shlonsky, 2004), and of adverbial modifiers in the verbal domain (Alexiadou, 1997; Cinque, 1999), it is here argued that modifiers of the category P (prepositions, postpositions and particles) appear to stack in a fixed order to the left of the head. However, not all languages lexicalize all types of modifiers, which raises the question of whether L2 learners, on gaining awareness of the semantics of modifiers, may reveal knowledge of the syntactic hierarchy despite its absence in the first language (L1), a lack of instruction in this domain,[1] and a paucity of evidence in natural input. We examine whether L2 learners of English are able to overcome the poverty of the stimulus and employ functional categories not instantiated in the L1 (White, 2003: 139–141). In Section II, we present an account of the syntax of P-modifiers in English, in relation to their manifestation in other languages. In Sections III and IV, descriptions are given of two experiments with novel methodology involving computer animation incorporated into slideshows. In both cases, accuracy rates were resolutely above chance and at extremely similar levels across the proficiency range and across L1s. In Section V, the implications of these findings are drawn out: knowledge of the hierarchy of spatial modifiers is in evidence at all stages of development, revealing that lexical semantics can trigger functional projections in L2 syntax. This is commensurate with a view of the language faculty with unidirectional information flow between modules, from lexical semantics to syntax at the lexical interface prior to syntactic derivation, and from syntax to phrasal semantics at the interpretive interface following syntactic derivation. The fact that learners show knowledge of semantics–syntax mappings from the outset, despite the absence of particular mappings in the L1, provides further evidence of the role of Universal Grammar (UG) in second language acquisition.

II The syntax of spatial modifiers in L1 and L2

An emerging consensus in research on the syntax of adpositions is that there is a universal layered structure inside PP, with a higher directional P (PathP), a lower locational P (PlaceP), a locative nominal projection (LocN), and a semantically vacuous PP that assigns case to DP (e.g. van Riemsdijk, 1990; Koopman, 2000; den Dikken, 2006; Stringer, 2007; Svenonius, 2008), as exemplified in English:

(1) $[_{PathP}$ from $[_{PlaceP}$ on $[_{LocN}$ top $[_{P}$ of $[_{DP}$ the table]]]]]

This hierarchy is attested in many languages despite great variability in manifestations of the category P, which may be expressed by means of prepositions, postpositions, particles, or affixes of various types. For example, German circumpositions are best captured by means of the layered PP structure, as shown by van Riemsdijk (1990), and nominal case suffixes in languages such as Lezgian and Inuit are found in precisely the reverse order of prepositions, in line with the Mirror Principle (Baker, 1985), as shown by van Riemsdijk and Huybregts (2007). Languages with less regular paradigms, such as Northern Sami, and languages with mixed systems of adpositions and case suffixes, such as Hungarian, also appear to respect the syntactic hierarchy (Stringer, 2008).

In an extension of this work, we examine how the expression of trajectories may be elaborated by means of modifying elements within spatial PP. Following observations made by Stringer (2005), at least three types of P-modifier may co-occur in a fixed structural hierarchy, as exemplified in (2). That these are indeed P-modifiers with a fixed word order to the left of the head, and not verb particles or 'satellites', is shown by tests of displacement (3a–d).

(2) The fish swam [$_{\text{DEG}}$ {right/straight} [$_{\text{FLOW}}$ {on/back} [$_{\text{TRAJECT}}$ {through/down} [$_{\text{PP}}$ into the cave]]]].

(3) a. The fish swam {* straight through on/* on straight through/* on through straight, etc.} into the cave.
b. It was [straight on through into the cave] that she swam.
c. * It was [through into the cave] that she swam straight on.
d. * It was [into the cave] that she swam straight on through.

The highest modifiers are those of Degree (or intensity), which are well recognized, and standardly used as a test of prepositional status (Emonds, 1976). Degree modifiers are hard to define precisely, but usually have a sense of 'directly', 'exactly', or 'completely'. *Right* may be used with either directional or locational P, whilst *straight* may only be used with directional P, as shown below.

(4) a. The bird flew {right/straight} into the hole.
b. The bird lived {right/* straight} in the hole.

The motivation for an independent class of Flow modifiers springs from the fact that they have their own particular position, always following Degree modifiers, and always preceding Trajectory modifiers (Stringer, 2005). *On* expresses the continuation of the directional flow, and *back* expresses the reversal of the directional flow. The third class consists of elements normally appearing as lexical P, but functioning in this case as P-modifiers, thus elaborating on simple trajectories. This class of elements includes *up*, *down*, *through*, *over*, and *across*. Their status as P-modifiers can be distinguished from their status as prepositions by means of tests of syntactic distribution as in (3), and by *right*-modification (Ayano, 2001: 79, footnote 1).

It must be stressed that the observations offered here concerning P-modifiers only apply to these lexical items on the relevant interpretations. *Straight* is often ambiguous between a Degree modifier reading and a directional adverb reading (in contrast with *diagonally*, *round*, etc.). The following example could mean that Pat ran right to the post office, without getting sidetracked or taking any detours (P-modifier), or that he ran in a straight line (adverbial).

(5) Pat ran straight to the post office.

Similarly, *back* can sometimes mean the opposite of *front*, in which case it cannot be construed as a reverse Flow modifier, as shown below.

(6) Harry moved back, because he hated sitting at the front.

In the context of discussion of directional adverbs, Svenonius (2008) considers the co-occurrence of *straight* with a fourth type of modifying element, namely Measure Phrases, observing that the order of measure phrases and directional adverbs is rigid, as in *{twenty centimeters straight / * straight twenty centimeters} below the window*. We find Svenonius' observation to be accurate on the intended interpretation (i.e. the directional adverbial reading: *straight* as against *diagonally*), but Measure Phrases appear to be in strict complementary distribution with *straight* as a Degree modifier – indeed with any of the P-modifiers in the hierarchy identified above – and will not be considered here.

Prosody plays a pivotal role in the parsing of phrases with multiple modifiers. Just as in the adverbial hierarchy (Cinque, 1999), the insertion of pauses, shifting of stress or other variance in the intonational contour results in the assignment of a different syntactic structure with a different semantic interpretation. Consider the sentence in (7a), with variants on the postverbal elements given in (7b–c), where φ indicates the relevant prosodic boundaries.

(7) a. Although she was tired after the great migration, the bird flew [$_{\varphi}$ right on down] to the lake.
b. {*[$_{\varphi}$ right down on] to the lake / [$_{\varphi}$ right down], onto the lake}
c. {*[$_{\varphi}$ down right on] to the lake / [$_{\varphi}$ down], right onto the lake}

In a context of 'continuation' that renders the Flow modifier meaningful, and with the P-modifiers in a single intonational phrase, the meaning of (7a) is clear, and the word order corresponding to this interpretation is fixed. However, if a pause is inserted after *right down*, and *on* forms a prosodic unit with the prepositional head *to*, as in (7b), the syntax changes such that *right* modifies *down*, which is in this case not a modifier but an intransitive head P (following Emonds, 1985), and *on* forms a complex P *onto*. In this case, the meaning changes such that the bird lands on the surface of the lake, which is not entailed by (7a). In the variation in (7c), the degree modifier is not applied to the intransitive P *down* but to the transitive P *onto*, such that emphasis is given not to the swooping downward but to the precision of the landing. Clearly, if such subtleties of interpretation are to be investigated in L2 acquisition, then not only must there be manipulation of syntax, but assiduous control of both context and prosody.

Not all languages lexicalize all types of P-modifier, but when two or more are found, they conform to syntactic predictions. In a pattern suggestive of an implicational hierarchy,[2] a language may lexicalize all three (e.g. German, English), the higher two (Estonian, Hungarian), only the highest (French, Spanish), or none at all (Japanese, Korean), as shown below.

(8) direct zurück hoch auf den Berg DEG–FLOW–TRAJECT (German)
straight back up on the Mountain
'straight back up on the mountain' (**zurück hoch direkt*, etc.)

(9) otse tagassi teatri-sse DEG–FLOW (Estonian)
right back theatre-into
'right back into the theatre' (* *tagassi otse*)

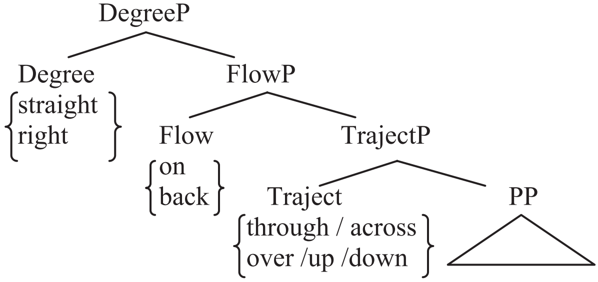
(10) juste par en dessous le pont DEG (French)
right via LocP underneath the bridge
'right under the bridge'

(11) tēburu no (ma-) ue (* ma-) de Ø (Japanese)
table GEN (right-) top (* right-) LocP
'right on top of the table' (literally: 'on right top of the table')

In languages such as Japanese and Korean, the inventory of adpositions is typically very small, as trajectories are generally expressed by means of verbs or complex PPs with spatial nouns. Such intensity modifiers as exist (e.g. Japanese *ma(n)* 'right', *sugu* 'immediately', *chōdo* 'exactly') attach not to P but to spatial nouns.

There may be several explanations of why Degree, Flow and Trajectory modifiers are aligned in this order: perhaps there is a syntactic principle at work, or perhaps this is a syntactic manifestation of a semantic hierarchy. Just as with adverbs and adjectives, an argument could be made that they do not project distinct phrasal levels, and are left-adjoined to VP. The logical problem for L2 acquisition that we address in this article remains irrespective of the account adopted. For current purposes, we assume a parallelism between the verbal, nominal and adpositional domains: just as there are functional projections in a fixed hierarchy supporting adjectives (Cinque, 1994; Crisma, 1996; Shlonsky, 2004) and adverbs (Alexiadou, 1997; Cinque, 1999), a similar system appears to obtain in the realm of adpositions, although with a more limited inventory. Cinque (1999: 106) posits approximately thirty functional projections hosting adverbial modifiers. The system of spatial P-modifiers outlined above contains three categories, and can be represented as follows.[3]

(12)



Such hierarchies have not been the subject of second language research to date, although their implications have sometimes been drawn out with respect to other phenomena. In the adjectival domain, for example, Androutsopoulou et al. (2008) examined noun–adjective placement in L2 Spanish by L1 French learners with specific reference to the adjectival hierarchy, although the hierarchy itself was not the target of the investigation.

In the adverbial domain, Ionin and Wexler (2002) found that thematic verbs in the L2 English of L1 Russian speakers were more likely to raise past Manner adverbs than Frequency or Epistemic adverbs, a fact which they attributed to the relative position of the adverbs on the functional hierarchy, but the focus of the investigation was verb-raising rather than how the hierarchy itself was revealed in the course of L2 acquisition.

Predictions of patterns of L2 development of spatial modifiers vary by theoretical framework. On the assumption that modifiers are hosted in distinct functional projections (Cinque, 1999), those accounts that posit impaired access to Universal Grammar in general (Clashen and Muysken, 1989; Meisel, 2008) or the domain of functional projections in particular (Eubank, 1993/94; Smith and Tsimpli, 1995; Hawkins and Chan, 1997) predict interminable confusion with regard to the hierarchy for types of modifier that are not instantiated in the L1. Accounts that argue for impairment only of uninterpretable features (Tsimpli, 2007; Tsimpli and Mastropavlou, 2007) have no bearing on this issue as the modifiers are presumably fully interpreted in the positions where they are base-generated. Accounts that rely on L2 input for the projection of new functional categories (Vainikka and Young-Scholten, 1994, 2006) predict initial difficulty but admit the possibility of acquisition. Strong continuity accounts such as the Full Transfer/Full Access hypothesis of Schwartz and Sprouse (1996) are neutral as regards when new functional categories might be in evidence, but allow for rapid convergence on the target grammar once the semantics of individual modifiers has been identified. In this latter vein, we consider that if knowledge of UG is available in L2 acquisition, the syntax of modification should reveal itself despite the absence of particular types of modifier in the L1.

The linguistics literature to date concerning the syntax of adpositions makes no mention of the syntax of spatial modifiers, and nothing is known of patterns of acquisition in either L1 or L2 acquisition; thus to what degree P-modifiers are learnable in a second language and what role the modificational system of the L1 might play were unknown factors prior to this study. The general research question we attempted to address was whether learners of English show knowledge of the universal hierarchy of P-modifiers over the course of L2 development. Three contrasting hypotheses were considered.

- Hypothesis 1: The hierarchy will be in evidence from the outset.
- Hypothesis 2: The hierarchy will emerge gradually.
- Hypothesis 3: The hierarchy is not acquirable for learners who lack L1 analogues.

III Experiment 1

1 Participants and location

The participants were drawn from six proficiency levels of an Intensive English Program at a large university in the Midwest of the USA. The proficiency levels were derived independently of this project by the battery of placement exams used by the program five weeks prior to experimentation. Initial placement criteria included composition, reading, vocabulary, grammar, listening comprehension and oral interviews, and promotion in the course involved integrating subsequent sets of test performance scores with previous course grades and current TOEFL scores. A total of 121 students from 6 different levels of proficiency successfully took part in Experiment 1, after 10 participants were eliminated according to pre-established criteria (persistent inattention, inaccurate responses on two or more of the four fillers, obviously artificial response patterns). The numbers of students by level were as follows: L2: 2, L3: 13, L4: 27, L5: 41, L6: 32, L7: 6. For purposes of analysis, these were conflated into three general proficiency groups: Lower-intermediate (L2–L4: 42), Intermediate (L5: 41) and Advanced (L6–L7: 38).

Learners came from 17 different L1 backgrounds, listed as follows with numbers of native speakers who completed the study: Arabic (15), Bambara (2), Chinese (14), French (2), Hungarian (2), Japanese (10), Korean (36), Mongolian (1), Portuguese (4), Russian (2), Spanish (5), Tajik (2), Thai (1), Tamil (1), Tartar (1), Turkish (25), and Vietnamese (2). Three participants were bilingual or multilingual from childhood (1 French/Bambara, 1 French/Bambara/Russian, 1 Tartar/Russian), and two languages were spoken only by bilinguals (Bambara and Tartar). While all language backgrounds were represented in the general analysis by proficiency level, a comparative analysis was also made of the performance of learners from the L1 groups with the most speakers: Korean, Turkish, Arabic, Chinese, and Japanese. None of these languages has more than one level of the hierarchy instantiated, so these learners must project functional categories that are absent in the L1.[4]

The main experiments were conducted in a language lab with learners seated in individual booths, using a communal main screen and surround speakers. Despite the range of affective factors and processing strategies among students, it was thus possible to synchronize aural and visual stimuli for all participants. The control experiments were conducted with 20 native speakers of English, aged 19–48, all of whom were Americans who had spent most of their lives in the Midwest of the USA, and none of whom were linguists.

2 Contextual materials

An original narrative was designed to contextualize PPs and their modifiers. The story was intended to be interpretable across cultures, and involved characters and scenes that were variations on the well-known Middle-Eastern folk-tale of Aladdin.[4] The narrative runs as follows. In a cave filled with treasure, Aladdin takes a magic lamp from under the nose of a wizard. He then jumps onto a magic carpet and flies up to an opening at the top of a flight of stairs at the side of the cave (*He flies right up out of the cave*). He passes through a tunnel to the outside (*He flies on through to the outside*), where he continues to fly through various spatial environments, each of which provides a plausible context for a targeted combination of prepositions and modifiers. In our version of the tale, he flies everywhere: reducing variability in manner of motion allowed for greater focus on trajectories. For example, in the course of his journey, he flies over some camels, up into the clouds, down to a lake, behind a waterfall, under a rock bridge, across a desert, through a city gate, etc. At one point he drops the lamp, but does not realize until later, and therefore has to fly back through the same environments, enabling further test materials to be utilized. Aladdin retrieves the lamp, and in the renewed pursuit, the wizard falls into a lake. Aladdin finally manages to make it back into the desert, where he releases the genie. The sentences making up the narrative (which were subject to experimental manipulation) are reproduced in Appendix 1.

The narrative was presented visually. An advantage of using visual stimuli together with linguistic encodings of motion events is that they can force an intended interpretation of a given sentence, despite the polysemy that is rife in adpositional systems. Every P-modifier in the hierarchy that we examine is potentially ambiguous, but whether *right* and *straight* are interpreted as Degree modifiers or as a directional adverbials, whether *on* is interpreted as a Flow modifier or a preposition, and whether *back* signifies return or movement to a posterior location may be largely controlled with appropriate visual stimuli. To embed the visual stimuli within a narrative was necessary in order to provide appropriate context for Flow modifiers, which necessarily express continuation or return with specific reference to prior events.

Previous experimental work on motion events has tended to rely either on two-dimensional images (Berman and Slobin, 1994; Strömqvist and Verhoeven, 2004) or on video (Pourcel, 2002; Hohenstein et al., 2004) for elicitation of utterances or judgments. For the current project, novel experimental stimuli were developed, by making use of basic computer animation. The advantages of animation include the incorporation of actual rather than inferred motion in the stimuli (as with video), and the expression of a full range of motion events (as with pictures), without placing actual actors in peril. The animation was achieved in the following way. First, the various characters, objects and background scenes were hand-drawn and coloured. Second, the cut-outs and background scenes were scanned as digital images. Third, they were incorporated into Microsoft PowerPoint slides, arranged in layers depending on the desired visibility of objects, and animated to create motion events. So, for example, when the protagonist flies through a gate in a city wall, he disappears as he passes behind the gate as the latter is the top layer among the objects in the slide. The complete animation is available for download from the first author's professional webpage (www.indiana.edu/~dsls/faculty/stringer.shtml).

3 Linguistic materials

As discussed previously, prosodic cues are also indispensable to establish the intended interpretation of a string of modifiers. Following a pilot study, two problems in particular were identified in the formulation of stimuli for preference and grammaticality judgment (GJ) tasks. The first was that the interactions between prosody and parsing were much more complex than we had at first assumed. The main issue was our attempt to ensure that the lexical elements of the modificational hierarchy were in fact parsed as part of a prosodic unit with P, and interpreted as P-modifiers, rather being prosodically aligned with V, and interpreted as verb particles. This might have been simple if there were only two contrastive patterns of prosody. However, prosodic units are somewhat underdefined in the phonology literature. There may be a pause between prosodic units; there may be a change in pitch such that the first sequence is high and the second low, or vice versa; or both can have internal pitch and stress variation, the shift to the extended PP being marked by sudden rise and gradual fall. The problem was that we could not systematically predict which prosodic pattern would be the most unambiguous delivery of the materials. Sometime a pause was felicitous; sometimes not. Within a given combination

Table 1 Targeted combinations in Experiments 1 and 2

(A) DEG–FLOW (× 6)	(a1) straight on, (a2) straight on, (a3) right on, (a4) straight back, (a5) right back, (a6) right back
(B) DEG–TRAJECT (× 6)	(b1) right up, (b2) right down, (b3) straight through, (b4) straight down, (b5) right out, (b6) straight out
(C) FLOW–TRAJECT (× 3)	(c1) on through, (c2) on down, (c3) back over
(D) DEG–FLOW–TRAJECT (× 3)	(d1) right on up, (d2) right back down, (d3) straight back across

type (i.e. DEGREE-FLOW, FLOW–TRAJECTORY, etc.), the felicity of particular prosodic patterns varied from item to item, apparently depending at least in part on the choice of lexical items. For this reason, the most appropriate prosody for stimuli was selected an item-by-item basis, based on native-speaker judgments. The second problem was that any written cues on the response sheets allowed participants to read as well as listen. This made prosodic rephrasing possible after the oral stimulus, so all written cues were removed from the response sheets, and linguistic stimuli were restricted to oral delivery. With these changes, once all stimuli had been digitally recorded and incorporated into the animated PowerPoint slides as sound files, we were able to ensure the most felicitous prosodic phrasing for the stimuli based on native speaker intuitions, and secure the uniform delivery of stimuli in both experiments across many sessions.

There were 26 slides in total: 3 initial example slides, 2 fillers for narrative coherence, 3 slides targeting onomatopoeia (outside the scope of the current discussion), and 18 test slides targeting the hierarchy of spatial modifiers. The stimuli were formulated as shown in Table 1, and the actual sentences subjected to manipulation are given in Appendix 1. The stimuli were not as balanced as they could have been if created outside the context of a narrative; however, it was decided that the felicitous embedding of multiple modifiers in a coherent narrative was of greater importance, so we strived to balance stimuli within narrative constraints.

4 Protocol

In advance of the experimentation, the vocabulary to be used was presented to the students for the purpose of making clear the meaning of each of the P-modifiers on the intended interpretations in English. Acquisition of the lexical items themselves was not the subject of investigation, but rather their interaction with one another, so pains were taken to ensure that individual lexical meanings were understood and accessible. The items on which they received instruction were the Degree modifiers *right* and *straight*, the Flow modifiers *on* and *back*, the Trajectory modifiers *up*, *down*, *through*, *over*, and *across* in prepositional contexts, and the locative nouns *front* and *top*.

As with the test materials, the instructional materials were presented in the form of animated PowerPoint slides. A handout was also created, which was left on learners' desks throughout the experiments, so that they could quickly recall, by means of written and visual aids, the meanings of individual items. The most important aspect of the logic of this part of the experimentation was that students were taught modifiers in isolation (i.e. 1 modifier + PP), but they were tested on modifiers in combination (i.e. 2 or 3 modifiers + PP).

During testing, one experimenter was responsible for oral delivery of all instructions, and another for the manipulation of visual images and pre-recorded test stimuli. The third and fourth experimenters administered all documents to participants, and performed the crucial function of monitoring the behaviour of participants during the session. Notes were taken if participants did not look at the screen, or changed any of their initial answers, or otherwise ignored instructions. As each new slide was presented, we checked that all participants were looking at the screen again, and only following an agreed signal did the experimenter in control of the visual materials start the animation, to make sure that learners processed the oral and visual stimuli in sync. All linguistic stimuli were repeated once after a 4-second pause.

Experiment 1 was a preference task: following oral delivery of two variants of a sentence, learners circled (a) or (b) on their answer sheets, according to which sounded better. To clarify: we use the term preference task, despite the fact that such tasks are often termed grammaticality judgments, because one cannot be sure that in choosing one answer, the learner believes the other to be ungrammatical. The results of this preference task will later be compared to the results of Experiment 2, in which straightforward judgments of grammaticality were elicited. An example stimulus from Experiment 1 is given in (13).

(13) Experiment 1 Sample:
‘He flies ______________ over the camels.’ (* *on straight/straight on*)
A B

The rationale behind the forced choice was to avoid a preponderance of ‘don’t know’ answers given the processing difficulties inherent in many test items and the indeterminate nature of this type of second language judgment. If participants felt 75% sure of their preference, we wanted to ensure that a positive value was recorded, whilst if the opposite were true, we wanted a negative value for the response. We also wanted the initial reactions of students to the stimuli, without the reflective time spent on more elaborate systems of judgment such as Likert scales, which increase the likelihood of prosodic rephrasing during sentence recall. Both learners and controls were told to listen to how the sentences sounded, and to judge them immediately on the way they sounded, without considering other pronunciations. The order of presentation of targetlike and non-targetlike variants was systematically varied across stimuli.

5 Results

A mixed design ANOVA was performed with stimulus type as the within-participant factor and proficiency group and L1 as between-participant factors. The stimulus types were as described earlier. The four proficiency levels included the three learner groups and the native

Table 2 Experiment 1: Preference task accuracy scores by proficiency level: Group means showing percentage accuracy, with significance above chance

	Group1 (*n* = 42)	Group 2 (*n* = 41)	Group 3 (*n* = 38)	Controls (*n* = 20)
(A) DEG–FLOW	76***	74***	84***	95***
(B) DEG–TRAJECT	71***	78***	81***	98***
(C) FLOW–TRAJECT	41	38	44	83***
(D) DEG–FLOW–TRAJECT	64**	68***	76***	100***

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 3 Experiment 1: Preference task accuracy scores by L1: Group means showing percentage accuracy, with significance above chance

	Korean (*n* = 36)	Turkish (*n* = 25)	Arabic (*n* = 15)	Chinese (*n* = 14)	Japanese (*n* = 10)
(A) DEG–FLOW	78***	73***	79**	77**	80**
(B) DEG–TRAJECT	78***	80***	82***	73**	60
(C) FLOW–TRAJECT	37	40	49	48	47
(D) DEG–FLOW–TRAJECT	72***	67**	69*	71*	70*

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

controls. L1 was assessed in terms of the five largest L1 populations: Korean (36), Turkish (25), Arabic (15), Chinese (14), and Japanese (12), as well as the native English controls.

The results for all four proficiency groups and the five main L1 populations are given in Tables 2 and 3. The *p*-values displayed indicate significance above chance, and are unadjusted from *t*-statistics using estimated means and standard errors from the repeated measures ANOVA. The possibility of false positive increases due to multiple comparisons was controlled by using Benjamini and Hochberg's (1995) method for False Discovery Rate. The results of the native English controls, who served as both a proficiency group and a language group, are reported only once, in Table 2.

A main effect of stimulus type was found, $F(2.539, 309.722) = 57.216$, $p < .001$. The results for Types A, B and D did not reveal any significant differences, but Type C was significantly different from the other 3 types, $p < .001$. A main effect of proficiency was also found, $F(2, 122) = 4.652$, $p = .011$. Group 4 (the native controls) was significantly different from the other groups, $p < .001$. A significant difference was also found between Proficiency Groups 1 and 3: $t(78) = 2.878$, $p = .031$, although the effect size was quite small ($\eta^2 = .071$). With respect to L1 background, no significant differences were found between learner groups. *T*-tests reveal a significant difference between the native English controls and the learners, as follows: English vs. Arabic: $t(33) = 5.1$, $p < .001$; English vs. Chinese: $t(32) = 5.812$, $p < .001$; English vs. Japanese: $t(28) = 6.038$, $p < .001$; English vs. Korean: $t(54) = 7.07$, $p < .001$; English vs. Turkish: $t(43) = 7.275$, $p < .001$.

A glance at the descriptive statistics immediately reveals a difference between the relatively accurate performance on Types A and B (DEG–FLOW and DEG–TRAJECT) at all proficiency levels, the particularly non-targetlike performance on Type C (FLOW–TRAJECT) at all proficiency levels, and performance on Type D (DEG–FLOW–TRAJECT), which showed significant accuracy at all levels, but improvement with general proficiency. The high accuracy rates on Types A and B in evidence at all proficiency levels are particularly striking. In comparison, the generally weak performance on Type C (FLOW–TRAJECT) calls out for further scrutiny. Although at first pass it might appear that the lower reaches of the hierarchy pose a higher degree of difficulty, analysis by individual stimuli reveals that poor performance on Type C might be alternatively explained in terms of a lexical effect. In Experiment 1, participants treated items (c1) *on through* and (c2) *on down* very differently from (c3) *back over*. Accuracy rates for Proficiency Groups 1, 2 and 3 were, respectively, for (c1): 43%, 20% and 42%; for (c2): 12%, 15% and 8%; and for (c3): 67% (p = .023), 80% (p < .001) and 82% (p < .001). A similar effect was found when results by L1 were analysed. The scores from the five L1 groups were: (c1) 28%, 44%, 40%, 29% and 30%; (c2) 8%, 8%, 20%, 14% and 0%; (c3) 75% (p = .001), 68% (p = .06), 87% (p < .001), 100% (p < .001) and 80% (p = .03). One possible reason for this discrepancy might be that the PPs modified by these combinations were headed by *to*: *on through [to the outside]*; *on down [to the ground]*; *back over [to the waterfall]*. If participants rephrased the first two utterances prosodically as they considered their responses, the resultant forms could be interpretable with *through* or *down* either as verb particles or as P-modifiers, with *on* analysed not as a modifier at all but as part of the complex preposition *onto*. We shall return to this problem later, as the results of Experiment 2 did not reveal the same discrepancy.

Performance on the ternary combinations of Type D was significantly above chance, though showing an increase in accuracy with proficiency. This was to be expected given the increase in processing load. These examples were included to stretch learners, as native responses were so robust: the controls attained 100% accuracy for this type. While the proficiency group scores attained significance, again there were item effects. Scores for (d1) *right on up* neither reached significance nor showed any improvement across the proficiency levels, standing at 64%, 63% and 63%. The other two types, however, improved consistently with proficiency level: accuracy scores for (d2) *right back down* were 60%, 68% (p = 0.017) and 79% (p < 0.001) and those for (d3) *straight back across* were 67% (p = 0.023), 73% (p = 0.001) and 87% (p < 0.001).

Performance on individual stimuli was checked for item effects with regard to particular modifiers. In this regard, differences emerge on close examination of the data from the lower-intermediate group. A comparison between accuracy rates of Group 1 and Group 3 responses with *right* reveal approximately the same scores; similar comparisons of responses with *straight* usually, although not always, reveal jumps in accuracy rates between Groups 1 and 3. Such differences on the preference task are as follows: (a1) 62% to 79%; (a2) 62% to 79%; (a4) 79% to 92%; (b3) 76% to 89%; (b4) 52% to 74%; (b6) 55% to 82%. It is plausible that some individuals may have persistent misunderstanding of *straight* as a P-modifier, although Proficiency Groups 2 and 3 seem to have acquired the relevant meaning and syntax. On the analysis by L1 group, proficiency levels were conflated, so this item effect disappeared.

There was no interaction between L1 background and proficiency level; performance was remarkably uniform across the levels within each language. To take the largest L1 group as an example, accuracy scores for Korean learners in Proficiency Groups 1 (11 learners), 2 (16 learners), and 3 (9 learners), respectively, were as follows: Type A: 81%, 81%, 82%; Type B: 70%, 89%, 84%; Type C: 36%, 34%, 45%; and Type D: 67%, 69%, 85%.

To summarize the results: The learners were significantly outperformed by the controls in all cases, but they nevertheless showed rates of accuracy that were well above chance for the binary combinations of Types A and B, consistently underperformed on Type C (which contained a design flaw), and showed improvement and eventual accuracy on the ternary combinations of modifiers of Type D. There was no L1 effect.

IV Experiment 2

1 Participants and location

Experiment 2 (the GJ task) targeted the same linguistic knowledge as Experiment 1. One purpose was to obtain binary judgments of grammaticality, rather than preference judgments.[5] Another purpose was to control for task effects by enabling the triangulation of results across experiments. It was conducted with the same participants immediately after the completion of the first experiment. Following the same criteria for exclusion, 13 participants were eliminated (compared to 10 in Experiment 1) leaving a total of 118 students. The numbers of students by level were as follows: L2: 2, L3: 12, L4: 27, L5: 40, L6: 31, L7: 6; again, these were collapsed into three general proficiency groups: Lower-intermediate (L2–4: 41), Intermediate (L5: 40) and Advanced (L6–7: 37). As before, learners came from 17 different L1 backgrounds, five of which had sufficient numbers to test for L1 effects: Korean, Turkish, Arabic, Chinese, and Japanese.

2 Contextual materials, linguistic materials and protocol

The Aladdin animation was run again, but this time with different embedded sound files. For each slide, immediately following the animation, a male voice asked a question about the narrative, and a female voice answered by means of a sentence fragment. Positive or negative responses were elicited from participants following delivery of these pre-recorded sentence fragments. Once more, these were manipulations of the sentences provided in Appendix 1. The order of presentation of targetlike and non-targetlike variants was systematically varied across stimuli. An example stimulus from Experiment 2 is given below.

(14) Experiment 2 Sample:
‘Now where does he go?’ ‘Straight back across the desert.’ (*ok*)
A: good B: bad

Table 4 Experiment 2: GJ task accuracy scores by proficiency level: Group means showing percentage accuracy, with significance above chance

	Group 1 (*n* = 41)	Group 2 (*n* = 40)	Group 3 (*n* = 37)	Controls (*n* = 20)
(A) DEG–FLOW	76***	81***	80***	93***
(B) DEG–TRAJECT	69***	77***	79***	90***
(C) FLOW–TRAJECT	34	39	32	82***
(D) DEG–FLOW–TRAJECT	58	63**	68***	98***

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 5 Experiment 2: GJ task accuracy scores by L1: Group means showing percentage accuracy, with significance above chance

	Korean (*n* = 36)	Turkish (*n* = 25)	Arabic (*n* = 13)	Chinese (*n* = 14)	Japanese (*n* = 10)
(A) DEG–FLOW	79***	78***	78***	81***	73***
(B) DEG–TRAJECT	82***	69***	71***	76***	67**
(C) FLOW–TRAJECT	28	29	49	45	43
(D) DEG–FLOW–TRAJECT	65**	59	56	72**	43

Notes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

The rationale behind the use of sentence fragments was to further control for prosodic reanalysis by participants. For example, prosody can disambiguate between *he flies back* [$_{\varphi}$ *right into the desert*], which is grammatical, and * *he flies* [$_{\varphi}$ *back right into the desert*], which is not. An ungrammatical sentence fragment answer such as * [$_{\varphi}$ *back right into the desert*], provides a clear contrast to the grammatical [$_{\varphi}$ *right back into the desert*], and reduces the chance of the P-modifier being reanalysed as a verbal particle. The rationale behind the forced choice with time limitation was the same as described for Experiment 1, and the general protocol in terms of running the animation was also unchanged.

3 Results

Again, a mixed design ANOVA was performed with stimulus type as the within-participant factor and proficiency group and L1 as between-participant factors. The stimulus types were as described earlier: given that narrative context was crucial for interpretation, combinations of modifiers were tied to particular scenes. As before, the analysis by proficiency included the three learner groups and the native controls, and L1 was assessed in terms of the five largest L1 populations: Korean (36), Turkish (25), Arabic (13, as compared to 15 in Experiment 1), Chinese (14), and Japanese (12).

The results for all four proficiency levels and the five main L1 populations are given in Tables 4 and 5. As before, *p*-values indicate significance above chance, and are unadjusted from *t*-statistics using estimated means and standard errors from the repeated measures ANOVA. The possibility of false positive increases was controlled by using Benjamini and Hochberg's (1995) method for False Discovery Rate. Again, the results of the native English controls are reported only once, in Table 4.

As in Experiment 1, a main effect of stimulus type was found, $F(2.457, 292.409) = 53.374$, $p < .001$. Again, there were no differences between Types A and B, and Type C was significantly different from the other 3 types, $p < .001$. In Experiment 2, significant differences were also found for Type D vs. Type A and Type D vs. Type B, both $p < .001$. There was no effect of proficiency level. With respect to language background, no significant differences were found between learner groups, although *t*-tests again revealed significant differences between the native English controls and the learners, as follows: English vs. Arabic: $t(31) = 5.192$, $p < .001$; English vs. Chinese: $t(32) = 4.547$, $p < .001$; English vs. Japanese: $t(28) = 5.672$, $p < .001$; English vs. Korean: $t(54) = 6.523$, $p < .001$; English vs. Turkish: $t(43) = 7.113$, $p < .001$. Unlike in Experiment 1, an interaction of stimulus type and language group was found, $F(12.286, 292.409) = 2.092$, $p = .017$, due to the poor performance of the smallest groups (Arabic and Japanese) on Types C and D.

The descriptive statistics again clearly indicate the difference between the relatively accurate performance on Types A and B (DEG–FLOW and DEG–TRAJECT) in comparison with the other two types. Type C (FLOW–TRAJECT) stimuli produced a notably non-targetlike performance at all proficiency levels, and Type D (DEG–FLOW–TRAJECT) again showed improvement with general proficiency, although the accuracy levels were lower at each proficiency level than in Experiment 1.

In the previous experiment, the analysis of Type C results revealed considerably lower rates of accuracy for items (c1) *on through* and (c2) *on down* as compared to (c3) *back over*, and it was hypothesized that the first two might have been prosodically recast by participants, so that the displaced *on* could merge with the following preposition *to*, resulting in *onto*. However, in Experiment 2, the results did not reveal the same discrepancy. The accuracy rates by stimulus were as follows: (c1) 32%, 25%, 16%; (c2) 54%, 62%, 63%; (c3)15%, 28%, 19%, with poor performance on all stimuli. We have no account for this difference.

It is notable that the control participants also had difficulty with (c1) in particular, with scores of 50% in Experiment 1 and 65% in Experiment 2, bringing down the average accuracy for this type. The possibility of prosodic reorganization is somewhat complex here: as an anonymous reviewer points out, *He flew through onto the outside* is semantically incongruous, so such an interpretation might reveal problems with prepositional semantics on the part of learners. However, the native speakers presumably do not have such problems. It remains a possibility that some controls inserted a pause between the modifiers, which resulted in a legitimate structure: *He flew through, on to the outside*. Some evidence that this might have been the case is given below. Given the design flaw in juxtaposing *on* and *to* (albeit a legitimate combination in the target language), we are forced to leave more detailed examination of L2 knowledge of FLOW–TRAJECT for future work, in which the *to*-PP might be replaced with, for example, an *into*-PP.

The ternary combinations of Type D again proved more difficult for lower-level learners, although accuracy generally improved with proficiency. Group 1 results were not significantly above chance, while Groups 2 and 3 showed increasingly significant rates of accuracy. This pattern conforms to our understanding of these combinations as involving a higher processing load. Native-speaker responses were unequivocal at 98% accuracy. As before, there were differences in the learner responses to individual stimuli. Scores for (d1) *right on up* showed no improvement across the proficiency levels, with the learner groups performing at 49%, 43% and 41%. In contrast, the other two types showed a marked improvement across the proficiency range: accuracy scores for (d2) *right back down* were 56%, 68% (p = 0.019) and 73% (p = 0.003) and those for (d3) *straight back across* were 68% (p = 0.017), 78% ($p < 0.001$) and 89% ($p < 0.001$).

Performance on individual stimuli was checked for item effects with regard to particular modifiers. In Experiment 1, responses with *right* were comparable across proficiency levels, while responses with *straight* generally revealed jumps in accuracy rates between Groups 1 and 3. This suggestive pattern was weakly replicated in Experiment 2. Binary combinations with straight were judged by Groups 1 and 3 as follows: (a1) 68% to 76%; (a2) 78% to 76%; (a4) 76% to 84%; (b3) 83% to 92%; (b4) 46% to 54%; (b6) 59% to 81%. As in the first experiment, this difference in accuracy with *right* and *straight* is visible only in the lower proficiency groups; further investigation of this difference therefore requires further experimentation specifically with lower-intermediate learners.

As in Experiment 1, there was no interaction between L1 group and proficiency, and performance was generally consistent across the levels. Using the largest L1 group as an example, accuracy scores for Korean learners in Experiment 2 for Groups 1 (11 learners), 2 (16 learners), and 3 (9 learners), respectively, were as follows: Type A: 73%, 80%, 84%; Type B: 78%, 88%, 87%; Type C: 21%, 29%, 33%; Type D: 45%, 84%, 78%. Once more, similarity across proficiency levels held within as well as across language groups.

To summarize the results: just as in the previous experiment, the learners revealed impressive rates of accuracy for the binary combinations of Types A and B, consistently underperformed on the flawed Type C, and showed increased accuracy with proficiency on the ternary combinations of Type D. Again, there was no L1 effect.

4 Triangulating results across tasks

In order to uncover any task effects, a third ANOVA was performed with task and stimulus type as within-participant factors and proficiency group and L1 as between-participant factors. A main effect of task was found, $F(1, 119) = 8.632$, $p = .004$, which we will seek to understand in terms of the performance of learners grouped by proficiency level and by language background. The proficiency groups displayed slightly different patterns of responses by task. Group 1 and the control group showed no task effect. Group 2 performed differently on Type A in task 1 and task 2: $t(41) = 2.142$, $p = .035$; however the effect size was small ($\eta^2 = .037$), and there were no differences for Types B, C and D. Group 3 performed differently on Type D ($t(36) = 2.774$, $p = .006$); again the effect size was small ($\eta^2 = .061$), and there were no differences for Types A, B, and C. Despite the general task effect, this more detailed analysis in terms of proficiency levels reveals that the similarities in performance are more striking than the differences.

An analysis by L1 group furnishes a similar understanding. The Korean, Chinese and control groups showed no task effects. The Arabic speakers showed differences on Type D, $t(12) = 2.316$, $p = .022$, but not on Types A, B and C. The Japanese speakers also showed differences on Type D, $t(9) = 2.637$, $p = .009$, but not on Types A, B and C. The

Turkish speakers showed differences on Type B, $t(24) = 2.116$, $p = .036$, but not on types A, C and D. Therefore, of the 24 possible pairwise comparisons (6 language groups × 4 stimulus types), only three produced a significant difference. Our conclusion is that while the ANOVA did reveal a main effect of task, the results of the two experiments remain highly comparable.

On considering the results of the two experiments more generally, one might ask why, despite performing so well above chance for Types A and B, the learners did not perform as well as native speakers. We believe the reason for this can be found in the examination of individual results. This is most clearly seen with the control participants, who all remained for some time after the experiment to give feedback on the experience. While the majority either performed exactly as expected or gave at most one or two unexpected answers, some produced negative results with a degree of consistency. Notably, one participant produced 5 out of 6 non-standard preference judgments for Type A, and in Experiment 2 accepted all variants of Types A and B, resulting in a total of six accurate responses to the grammatical variants and 6 inaccurate responses to the ungrammatical variants. Fortunately, following the experiment (and prior to these results coming to light), she provided an explanation of this by saying, 'Most of the sentences were fine if you just changed the pronunciation a bit.' Similarly, two other participants said that once or twice they thought it was 'OK if you just said it a different way.' These responses show clearly that at least some native speakers disregarded explicit instructions not to change the way sentences sounded before judging them. Perhaps this was to be expected given the fact that people are naturally bemused by ungrammatical sentences and have a natural tendency to repair deviant utterances. However, if some native speakers performed in this way, it seems very likely that non-native speakers found it difficult to adhere to the instructions, and prosodically rephrased some of the stimuli despite our efforts to prevent this by means of restricted visual context, pre-recorded stimuli, and explicit instructions. Although this may explain certain responses, the behaviour of some individuals is likely to remain mysterious in a cross-sectional study of this type. While one Turkish learner had an accuracy score of 24/24 (100%) on Types A and B over both experiments, another Turkish learner in the same proficiency level scored 8/24 (33%). It was considered that numbers of participants were sufficient to overcome the behaviour of outliers.

V Discussion

The results reveal knowledge of the functional hierarchy across general proficiency levels, and irrespective of L1 background, despite the fact that the relevant syntax is not taught in the classroom, and despite the scarcity of examples of multiple modifiers in natural input. Of the three hypotheses posited earlier, Hypothesis 3 can be discounted. Hypothesis 2 may be true to a degree, in that clearly the lexical semantics and prosody have to be in place, but the gradual development predicted by this hypothesis was not attested: even very low-level learners performed significantly above chance. It appears that there is direct mapping from lexical semantics to a pre-existing syntax right from the beginning, and that universals of prosody are likely to be in play. More gradual development was observed in performance on ternary combinations of modifiers, presumably due to improvement in processing skills. With regard to binary combinations of modifiers, Hypothesis 1 is the hypothesis that is best supported by these results: the hierarchy is robustly in evidence from the outset.

This phenomenon permits an interesting perspective on investigations of syntax–semantics correspondences in L2 acquisition. Previously, several lines of research have shown that subtle differences in syntax can lead learners to semantic interpretations that could not possibly be gleaned from instruction or negative evidence, with phenomena such as adjectival restrictions on *wh*-quantifiers (Dekydtspotter and Sprouse, 2001) and *combien* extraction in French (Dekydtspotter et al., 2001). Slabakova (2008) provides an insightful overview of such work, which sheds light on the flow of information at the interface from syntax to phrasal semantics. In an important sense, the current investigation has revealed the reverse of this process, with semantics informing syntax. However, the view adopted here is that there is not a single interface between syntax and semantics with a bidirectional flow of information. Rather, the correspondence is unidirectional, in different directions at two interfaces: from lexical semantics to syntax at the lexical interface, and from syntax to phrasal semantics at the interpretive interface. Slabakova (2006, 2008) is rightly cautious about distinguishing between lexical and phrasal semantics, and the interface with which she is concerned is that of interpretive semantics, following processes of syntactic derivation. If the lexical semantics of modifiers informs the syntax of directional predication, this flow of information from semantics to syntax is at a different point of contact between meaning and grammar, namely, the lexical interface. It is here that the semantics of lexical items helps mould initial syntactic representations (Pinker, 1989; Jackendoff, 1990; Hale and Keyser, 2002). While L2 research in this area has been productive with regard to the selectional properties of verbs (locatives: Bley-Vroman and Joo, 2001, Schwartz et al., 2003; datives: Bley-Vroman and Yoshinaga, 1992, Whong-Barr and Schwartz, 2002; psych verbs: White et al., 1999), the current study has charted new ground in its focus on lexical semantics informing the projection of hierarchies of modifiers. The logic was as follows: if the modificational hierarchy is indeed part of the toolkit of UG, and if adult learners have continued access to UG beyond any purported critical period, then once knowledge of the lexical semantics of modifiers is acquired, the syntax of modification should be naturally manifested. This is indeed what was found.

VI Conclusions

This study explored the question of whether L2 learners have prior knowledge of the functional hierarchy of adpositional modifiers, despite the lack of instantiation of the hierarchy in the L1, the lack of formal instruction, and the paucity of combinations of multiple modifiers in natural input. After instruction on the semantics of individual modifiers, a computer-animated narrative was used to contextualize combinations of modifiers, and a preference task and a grammaticality judgment task were administered to L2 learners of English from a variety of L1 backgrounds. The results show a conspicuous awareness of the functional hierarchy from lower-intermediate to advanced learners, irrespective of the L1, with responses triangulated across the two tasks. Learners were able to project the hierarchy on the basis of the lexical semantics of individual modifiers, revealing a flow of information from lexical semantics to syntax. We suggest that there are two distinct places of intersection between syntax and semantics, each of which exhibits a unidirectional flow of information: the lexical interface, at which lexical semantics informs syntax, and the interpretive interface, at which syntax informs phrasal semantics. That L2 learners so clearly apply semantics–syntax mappings at the lexical interface from individual items to functional categories not instantiated in the L1 constitutes sound evidence for full access to Universal Grammar in second language acquisition.

Acknowledgements

This research was conducted with the collaboration of the Intensive English Program (IEP) and the Center for Language Technology and Instructional Enrichment (CeLTIE), at Indiana University, IN, USA. Many thanks especially to Kathleen Bardovi-Harlig, Roger Crandall, Doreen Ewert, Lisa McCune, Heidi Vellenga, and Tara Zahler. We would also like to acknowledge the statistical support offered by Stephanie Dickinson of Indiana Statistical Consulting Center (ISCC), whose patience apparently knows no bounds. Insight into the nature of the L1 grammars was provided by Ahmad Alqassas (Arabic), Yi-Ting Wang (Chinese), Seiki Ayano (Japanese), Hyun Kyoung Seo (Korean), Ercan Balci (Turkish), and Gozde Bahadir (Turkish), among others. Thanks are also due to Kamil Ud Dean, Laurent Dekydtspotter, Bonnie Schwartz and William Snyder for insightful discussion as we framed the research questions and designed the experiments. We are also very grateful to three anonymous *Second Language Research* reviewers, whose comments were extremely useful.

Notes

1 All current textbooks of participants in this study were examined, as well as the textbooks from the previous three decades in the university's TESOL library. No examples were found of multiple P-modifiers. Seventeen professional ESL teachers lent their support to this study, none of whom had ever covered this topic in class, and all of whom had instinctive but not conscious knowledge of the hierarchy.
2 Thanks to Kamil Ud Deen for this observation.
3 It is possible that spatial P-modifiers are in fact hosted in Specifier positions rather than in the heads of functional projections, in line with Cinque's (1994, 1999) analyses of adjectives and adverbs. In the absence of typological evidence of bound morphemes ever fulfilling the same function in the PP domain, we remain agnostic on this issue.
4 An anonymous *Second Language Research* reviewer asks for clarification on this point. Note that the issue is whether these L1s instantiate the hierarchy of P-modifiers, for which they must have at least two such types of modifiers. Korean functions much the same way as Japanese (see example 11): it has LocN-modifiers, but no P-modifiers at all. Turkish has an element *tam* 'right' that might be analysed as a LocN-modifer, as in *Masa-nın tam üst-ü-nde* / table.GEN right top.Poss.Loc / 'right on top of the table'. Other translational equivalents in Turkish such as *direkt* ' right/directly' and *geri* 'back' pattern like adverbs. For example, *direkt* may be used with a verb – *Direkt geldi* / directly came / 'He came immediately' (compare: * *He came right*) – but not with a bare PP. Arabic arguably has one P-modifier, *tamaam* 'right', as in *wada3 al-kiaab tamaam-an ala al-taawilat* / put.he the.book exactly.ACC on the.table / 'He put the book right on the table'; in contrast to the Turkish example, Arabic *tamaam* may not be used with intransitive verbs, and may be used in PP sentence fragments. Chinese has one element, *jiu* 'right', that may appear in PP sentence fragments, as in *Jiu zai jiaoluo* / right in corner / 'right in the corner'. However, other translation equivalents function as adverbs. The sentence *Zheng hui-qu shangdian* might be loosely glossed as 'right back-go shop', but on closer analysis, *hui* and *qu* are both verbs and, unlike *jiu*, *zheng* cannot be found without a verb, so that a more

accurate gloss might be ‘immediately return-go shop’. In short, Korean, Japanese and Turkish have no P-modifiers, Arabic and Chinese have at most one type, and none of these languages provides speakers with evidence that multiple P-modifiers must stack in accordance with a universal hierarchy.

5 Initial concerns about how learners from different L1 backgrounds might perceive the ‘exoticism’ of the setting were laid to rest by informal discussion with informants from a variety of L1 backgrounds, and the discovery that such exoticism exists in all of the various versions found internationally. In the oldest Arabic and Syrian folk-tale versions, as well as in early translations into English and French, Aladdin himself is from China (the farthest East imaginable), whilst the sorcerer comes from Morocco (the farthest West imaginable).

6 A further consideration is that grammaticality judgments may be divided into three types, although distinctions are occasionally blurred: (1) judgments of well-formedness; (2) judgments of interpretation; and (3) judgments of truth-value (for discussion, see Gordon, 1996; McDaniel and Cairns, 1996). With two exceptions that are ill-formed in all contexts irrespective of prosody (**down back right*/ **across back straight*), all the manipulations of spatial modifiers in the experiment might permit a grammatical interpretation with a different visual context and with different prosody. Some might thus consider these to be judgments of truth-value. However, we assume that prosody is an integral part of the grammar. As such, all the responses in this experiment were considered to be judgments of well-formedness.

References

Alexiadou A (1997) *Adverb placement: A case study in antisymmetric syntax*. Amsterdam: John Benjamins.

Androutsopoulou A, Español-Echevarría M, and Prévost P (2008) On the acquisition of the prenominal placement of evaluative adjectives in L2 Spanish. In: Bruhn de Garavito J and Valenzuela E (eds) *Selected proceedings of the 10th Hispanic linguistics symposium*. Somerville, MA: Cascadilla, 1–12.

Ayano S (2001) The layered internal structure and external syntax of PP. Unpublished doctoral dissertation, University of Durham.

Baker M (1985) The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* 16: 373–415.

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 57: 289–300.

Berman RA and Slobin DI (eds) (1994) *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum.

Bley-Vroman R and Joo H-R (2001) The acquisition and interpretation of English locative constructions by native speakers of Korean. *Studies in Second Language Acquisition* 23: 207–19.

Bley-Vroman R and Yoshinaga N (1992) Broad and narrow constraints on the English dative alternation: Some fundamental differences between native speakers and foreign language learners. *University of Hawai’i Working Papers in ESL* 11: 157–99.

Cinque G (1994) On the evidence for partial N movement in the Romance DP. In: Cinque G, Koster J, Pollock J-Y, Rizzi L, and Zanuttini R (eds) *Paths toward Universal Grammar*. Georgetown: Georgetown University Press, 85–110.

Cinque G (1999) *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.

Clashen H and Muysken P (1989) The UG paradox in L2 acquisition. *Second Language Research* 5: 1–29.

Crisma P (1996) On the configurational nature of adjectival modification. In: Zagona K (ed.) *Grammatical theory and romance language*. Amsterdam: John Benjamins, 59–71.

Dekydtspotter L and Sprouse RA (2001) Mental design and (second) language epistemology: Adjectival restrictions of *wh*-quantifiers and tense in English–French interlanguage. *Second Language Research* 17: 1–35.

Dekydtspotter L, Sprouse RA, and Swanson KAB (2001) Reflexes of the mental architecture in second language acquisition: The interpretation of discontinuous *combien* extractions in English–French interlanguage. *Language Acquisition* 9: 175–227.

den Dikken M (2006) On the functional structure of locative and directional PPs. Unpublished manuscript, CUNY Graduate Center, New York

Emonds JE (1976) *A transformational approach to English syntax: Root, structure-preserving and local transformations*. New York: Academic Press.

Emonds JE (1985) *A unified theory of syntactic categories*. Dordrecht: Foris.

Eubank L (1993/94) On the transfer of parametric values in L2 development. *Language Acquisition* 3: 183–208.

Gordon P (1996) The truth-value judgment task. In: McDaniel D, McKee C, and Cairns HS (eds) *Methods for assessing children's syntax*. Cambridge, MA: MIT Press, 211–31.

Hale K and Keyser SJ (2002) *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.

Hawkins R and Chan CY (1997) The partial availability of Universal Grammar in second language acquisition: The failed functional features hypothesis. *Second Language Research* 13: 187–226.

Hohenstein JM, Naigles LR, and Eisenberg AR (2004) Keeping verb acquisition in motion: A comparison of English and Spanish. In: Hall DG and Waxman SR (eds) *Weaving a Lexicon*. Cambridge, MA: MIT Press, 569–602.

Ionin T and Wexler K (2002) Optional verb raising in L2 acquisition: Evidence from L1-Russian learners of English. In: Costa J and Freitas MJ (eds) *Proceedings of the GALA 2001 Conference on Language Acquisition*. Lisbon: Associação Portuguese de Linguistica, 134–41.

Jackendoff R (1990) *Semantic structures*. Cambridge, MA: MIT Press.

Koopman H (2000) Prepositions, postpositions, circumpositions and particles: The structure of Dutch PPs. In: *The syntax of specifiers and heads: Collected essays of Hilda J. Koopman*. London: Routledge, 204–60.

McDaniel D and Cairns HS (1996) Eliciting judgments of grammaticality and reference. In: McDaniel D, McKee C, and Cairns HS (eds) *Methods for assessing children's syntax*. Cambridge, MA: MIT Press, 233–54.

Meisel JM (2008) Child second language acquisition or successive first language acquisition? In: Haznedar B and Gavruseva E (eds) *Current trends in child second language acquisition: A generative perspective*. Amsterdam: Benjamins, 55–80.

Pinker S (1989) *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Pourcel S (2002) Investigating linguistic relativity: A research methodology. *Durham Working Papers in Linguistics* 8: 125–38.

Schwartz BD and Sprouse RA (1996) L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research* 12: 40–72.

Schwartz BD, Dekydspotter L, and Sprouse RA (2003) Pouring the fire with gasoline: Questioning conclusions on L2 argument structure. In: Liceras JM, Zobl H, and Goodluck H (eds) *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)*. Somerville, MA: Cascadilla Proceedings Project, 248–59.

Shlonsky U (2004) The form of Semitic noun phrases. *Lingua* 114: 1465–1526.

Slabakova R (2006) Is there a critical period for semantics? *Second Language Research* 22: 1–37.

Slabakova R (2008) *Meaning in the second language*. Berlin/New York: Mouton de Gruyter.
Smith N and Tsimpli IM (1995) *The mind of a savant: Language learning and modularity*. Oxford: Blackwell.
Stringer D (2005) Paths in first language acquisition: Motion through space in English, French and Japanese. Unpublished doctoral dissertation, University of Durham.
Stringer D (2007) Extending the PP hierarchy: The role of bare nominals in spatial predication. In: Scheffler T, Tauberer J, Eliam A, and Mayol L (eds) *Penn Working Papers in Linguistics (PWPL): Volume 13.1: Proceedings of the 30th Annual Penn Linguistics Colloquium*. Philadelphia, PA: Penn Linguistics Club, 379–92.
Stringer D (2008) Spatial affixes in syntax. Paper presented at Berkeley Workshop on Affix Ordering, University of California, Berkeley, 4 October.
Strömqvist S and Verhoeven L (eds) (2004) *Relating events in narrative: Typological and contextual perspectives*. Mahwah, NJ: Lawrence Erlbaum.
Svenonius P (2008) Projections of P. In: Asbury A, Dotlačil J, Gehrke B, and Nouwen R (eds) *Syntax and semantics of spatial p*. Amsterdam: John Benjamins, 63–84.
Tsimpli IM (2007) The interpretability hypothesis: Evidence from *wh*-interrogatives in second language acquisition. *Second Language Research* 23, 215–42.
Tsimpli IM and Mastropavlou M (2007) Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. In: Goodluck H, Liceras J, and Zobl H (eds) *The role of formal features in second language acquisition*. Mahwah, NJ: Lawrence Erlbaum, 143–83.
van Riemsdijk H (1990) Functional prepositions. In: Pinkster H and Genée I (eds) *Unity in Diversity*. Dordrecht: Foris, 229–41.
van Riemsdijk H and Huybregts R (2007) Location and locality. In: Karimi S, Samiian V, and Wilkins W (eds) *Clausal and phrasal architecture: Syntactic derivation and interpretation: A festschrift for Joseph E. Emonds*. Amsterdam: John Benjamins, 339–64.
Vainikka A and Young-Scholten M (1994) Direct access to X'-theory: Evidence from Korean and Turkish adults learning German. In: Hoekstra T and Schwartz BD (eds) *Language acquisition studies in generative grammar*. Amsterdam: John Benjamins, 265–316.
Vainikka A and Young-Scholten M (2006) The roots of syntax and how they grow: Organic Grammar, the Basic Variety and Processability Theory. In: Unsworth S, Parodi T, Sorace A, and Young-Scholten M (eds) *Paths of development in L1 and L2 acquisition: In honor of Bonnie D. Schwartz*. Amsterdam: John Benjamins, 77–106.
White L (2003) *Second language acquisition and Universal Grammar*. Cambridge: Cambridge University Press.
White L, Brown C, Bruhn de Garavito J, Chen D, Hirakawa M, and Montrul S (1999) Psych verbs in second language acquisition. In: Klein E and Martohardjono G (eds) *The development of second language grammars: A generative approach*. Philadelphia, PA: John Benjamins, 171–96.
Whong-Barr M and Schwartz BD (2002) Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. *Studies in Second Language Acquisition* 24: 579–616.

Appendix 1 Core linguistic descriptions linked to the narrative

Note: The italicized elements were subject to experimental manipulation.

Example slides

1: Here is Aladdin. Here is the wizard. Here is a *very beautiful* lamp.
2: Aladdin and the wizard are *going to* the cave.
3: Aladdin takes the *magic lamp* from the wizard.

Stimulus slides

4: He flies *right up* out of the cave.	[DEG [TRAJECT]]
5: He flies *on through* to the outside.	[FLOW [TRAJECT]]
6: He flies *straight on* over the camels.	[DEG [FLOW]]
7: He flies *right on up* into the clouds.	[DEG[FLOW[TRAJECT]]]
8: He goes *crash into* the birds.	ONOMATOPOEIA
9: The lamp falls *right back down* onto a tree.	[DEG[FLOW[TRAJECT]]]
10: The lamp falls *on down* to the ground.	[FLOW [TRAJECT]]
11: Aladdin flies *right down* in front of a waterfall.	[DEG [TRAJECT]]
12: He flies *whoosh over* a lake.	ONOMATOPOEIA
13: Aladdin flies *straight on* under a rock.	[DEG [FLOW]]
14: Aladdin flies *right on* across the desert.	[DEG [FLOW]]
15: He flies *straight through* into the city.	[DEG [TRAJECT]]
16: Oh no! The lamp is not *in his* bag!	FILLER
17: Aladdin flies *straight back* across the desert.	[DEG [FLOW]]
18: He flies *right back* under the rock.	[DEG [FLOW]]
19: He flies *back over* to the waterfall.	[FLOW [TRAJECT]]
20: He flies *straight down* behind the tree.	[DEG [TRAJECT]]
21: Aladdin flies *right out* from behind the tree.	[DEG [TRAJECT]]
22: The wizard falls *splash into* the lake.	ONOMATOPOEIA
23: Aladdin comes *straight out* from behind the waterfall.	[DEG [TRAJECT]]
24: He flies *straight back* across to the rock.	[DEG [FLOW]]
25: He flies *right back* into the desert.	[DEG [FLOW]]
26: Aladdin touches *the lamp*. The genie appears!	FILLER