L2 acquisition of P-modifiers: Fine-tuning the linguistic expression of trajectories

David Stringer, Beatrix Burghardt, Hyun Kyoung Seo, Yi-Ting Wang

Indiana University

1. Introduction

The following is a work-in-progress report from a project investigating the knowledge second language (L2) learners have of a syntactic phenomenon hitherto unrecognized in the linguistics literature, and not explicitly taught in the classroom environment: the hierarchy of modifiers of the syntactic category P. Crosslinguistic analysis reveals that if particular types of P-modifiers are lexicalized in a language, they appear to stack in a fixed order to the left of the head. However, not all languages lexicalize all types, which raises the question of whether L2 learners may have knowledge of the syntax of elements in the hierarchy that are not present in the first language (L1), despite lack of instruction and a paucity of evidence in naturalistic input. This project seeks to clarify the nature of the hierarchy, and examine whether L2 learners of English, across a range of L1s, and at various levels of general proficiency, are able to overcome the poverty of the stimulus. Furthermore, whilst previous experimental work on motion events has utilized either pictures or video for contextual stimuli, this report introduces novel experimental methodology involving computer animation incorporated into slideshows. This is shown to have a distinct set of advantages in experimental work of this type. In Section 2, we present an account of the syntax of P-modifiers in English, in relation to their manifestation in other languages. Whilst generalizations are necessarily tentative, the findings are both intriguing and robust enough to serve as the basis for L2 research on this topic. In Section 3, an account is given of a pilot conducted with the aim of refining expectations and methodology in this novel area of inquiry. Section 4 summarizes the results and the lessons learned for research design, and Section 5 offers a brief view of initial findings from the main set of experiments, which are currently in progress. Despite the preliminary nature of this report, the results offer an original insight into the workings of Universal Grammar in second language acquisition.

2. The syntax of spatial modifiers

It has been argued that Universal Grammar makes available a layered PP, with a directional P (PathP), a 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; Stringer, 2005; den Dikken, 2006; Svenonius, 2008), as exemplified in English below.

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

This hierarchy is attested crosslinguistically despite great variability in the manifestations of the category P, such as circumpositions in German (van Riemsdijk, 1990), nominal case suffixes in Lezgian (van Riemsdijk & Huybregts, 2007, based on Haspelmath, 1993), and affixes to postpositions in Hungarian (Stringer, 2008). In an extension of this work on universals in PP structure, we examine five types of modifying elements within spatial PP. Several of these elements have received attention in isolation, but their relation to each other within an overall

system of modification remains in need of investigation. Following observations made by Stringer (2005), at least three types of P-modifiers may co-occur in a fixed structural hierarchy, as exemplified in (2).

(2) Sally ran [$_{DEG}$ {right/straight} [$_{FLOW}$ {on/back} [$_{TRAJECT}$ {through/over} [$_{PP}$ into the room]]]].

At the top of the hierarchy are Degree (or intensity) modifiers, which are well-recognized, and standardly used as a test of prepositional status.[1] *Right* may be used with either PathP or PlaceP, whilst *straight* may only be used with PathP, as shown below.

(3) a. The mouse ran {right/straight} into the corner.
b. The mouse lived {right/*straight} in the corner.

Flow modifiers express the continuity or reversal of the directional flow, the former being expressed by *on* and the latter by *back*. Trajectory modifiers are elements normally appearing as lexical P which have taken on a modifying function, and include *up*, *down*, *through*, *over*, and *across*. The motivation both for the hierarchy itself and more basically for the classification of these elements as P-modifiers, rather than verb particles or 'satellites' (Talmy, 1991, 2000), is based on tests of syntactic distribution and displacement. The fixed word order of these elements is demonstrated in (4a-b); and constraints on movement are shown in (4c-e).

(4) a. Sally ran straight on through into the room.
b. *She ran {straight through on / on straight through / on through straight / through straight on / through on straight} into the room.
c. It was [straight on through into the room] that she ran.
d. *It was [through into the room] that she ran straight on.
e. *It was [into the room] that she ran straight on through.

In addition to these three elements, Measure Phrases are also used to modify trajectories, but appear to be in complementary distribution with modifiers in the above hierarchy, in the absence of pauses which allow for different parses of the string.

(5) Sally rode 20 miles *{straight/on/over} to the farm.

A fifth type of P-modifier is attested, in the form of onomatopoeia inserted above the head P. Similar restrictions on syntactic movement also obtain in this case. Like Measure Phrases, onomatopoeia appears to be in complementary distribution with other forms of modification, although this remains to be more fully investigated. Examples of such forms are given below.

(6) The helicopter flew {crash into /*into crash} a tree / {splash into / *into splash} a lake / whoosh over / *over whoosh} their heads.

In determining the system of co-occurrence of these five types, polysemy can confuse the issue. For example, *straight* is often ambiguous between a Degree modifier reading and a

[1] This observation regarding *right* modification is originally due to Jespersen (1992 [1924]).

directional adverb reading, such that in the phrase *straight to the house* it could mean *right* (no detours), or *in a straight line* (see Svenonius, 2008, for discussion of the syntax of *straight* as a directional adverb). Similarly, *back* can sometimes mean the opposite of *front*, in which case it cannot be construed as a reverse Flow modifier (Stringer, 2005: 443-444). It must be stressed that the observations offered here concerning the ordering of P-modifiers apply to these lexical items only on the relevant interpretations.

As regards crosslinguistic manifestations of the hierarchy, a brief look at the range of possible expression reveals that not all languages lexicalize all types of modifiers, but when two or more are found, they conform to syntactic predictions. For example, the configuration of Degree, Flow and Trajectory is found in German, as in *direct zurück hoch* 'straight back up' (NB. **zurück hoch direct*, **hoch direct zurück*, etc.). In a language such as Estonian, there are no Trajectory modifiers, but Degree and Flow appear in the predicted order, as in *otse tagassi* 'right back' (NB. **tagassi otse*). In a language such as French, there are no Flow or Trajectory modifiers, but there can be at most one Degree modifier above P, such as *juste* 'right' in phrases such as *juste par en dessous* - right via LOC underneath - 'right underneath'. In strong exemplars of 'verb-framed' languages such as Japanese, there may be no P-modifiers at all. Directional predication is characteristically lexicalized in verbs, and the inventory of adpositions is very small, although Japanese does have intensity modifiers like *ma(n)* which differ from Degree modifiers in English in that they attach to spatial nouns (for further discussion, see Stringer, 2005, 2007). Thus whilst initial investigations support the plausible existence of a universal hierarchy of modifiers above P, there is a range of realization such that all three, two, one or none of the P-modifiers may be actually manifested in a given language. Such distribution raises the question of whether there might be an implicational hierarchy: perhaps the existence of Trajectory implies Flow and Degree, and the existence of Flow implies Degree, but our current investigations have not yet involved a broad enough survey of typologically distinct languages to make this claim.

3. A preliminary investigation of P-modifiers in L2 acquisition

Initial hypotheses and pilot rationale.

The plausibility of a universal hierarchy of P-modifiers, coupled with crosslinguistic variation regarding which elements are actually lexicalized, creates an intriguing situation from the perspective of second language research. It is unclear how learners from different L1 backgrounds will fare in the acquisition of a language with a full set of modifiers. The absence of instruction and the relative paucity of such concatenations in naturalistic input create a serious problem of poverty of the stimulus, yet if knowledge of Universal Grammar is available in L2 acquisition, as is argued by many researchers, once the requisite acquisition of lexical items has taken place the syntax of modification may manifest itself despite deficiencies in input and instruction. The linguistics literature to date makes no mention of the syntax of spatial modifiers, and nothing is known of patterns of acquisition in either L1 or L2 acquisition; it is thus unknown to what degree P-modifiers are learnable in a second language, and what role the modificational system of the L1 might play. The general research question we attempted to address was whether learners of English show knowledge of the hierarchy of P-modifiers over the course of L2 development. For the pilot experiment, three initial, contrasting hypotheses were considered.

(7) Either (i) the hierarchy will be in evidence from the outset; or (ii) it will emerge gradually as learners establish appropriate L2 lexical representations; or (iii) learners who lack L1 analogues will not demonstrate any knowledge of the hierarchy.

The purpose of the pilot was not to have a dry run of the finished version of the newly designed methodology, nor to furnish developmental data to comprehensively test all hypotheses, but to consciously experiment with our approach to the project before committing to the greater logistics of a large-scale study. This proved to be the right decision, as several crucial elements of the methodology were altered, and expectations for the main study were significantly revised.

Participants and location.
The population participating in this ongoing project is drawn from a large number of learners of English with a variety of L1 backgrounds, across 6 proficiency levels of an Intensive English Program at the large public university in the Midwest of the United States. The proficiency levels were derived independently of this project by the battery of placement exams used by the program. In advance of the main experimentation, two tasks were piloted on 20 lower-proficiency learners, with 7 different L1s, listed here by number of speakers: Arabic (8), Korean (5), Japanese (3), Chinese (1), Spanish (1), Tajik (1) and Thai (1). Note that all these languages apart from Chinese are strong exemplars of the 'verb-framed' type discussed above, with few or no P-modifiers. All learners were young adults in their twenties. Two classes of lower-proficiency learners were chosen in part to see if the experiments could be run across the spectrum of learners on the program, and in part for logistical reasons. The experiments were conducted in a language lab, using a main screen and surround speakers.

Materials.
As mentioned earlier, previous experimental work on motion events has tended to rely either on two-dimensional images (Berman & Slobin, 1994; Strömqvist & Verhoeven, 2004) or on video (Pourcel, 2002; Hohenstein, Naigles, & Eisenberg, 2004) for elicitation of utterances or judgments. One advantage of providing visual context is that it can force an intended interpretation of a given sentence, despite the lexical polysemy discussed above and the ambiguity inherent in combinations of motion verbs and adpositions. Such cues are even more robust when the scene itself is embedded in a narrative. For the current project, novel experimental methodology was developed which makes use of a very basic kind of 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 as they fall into lakes, tumble down mountains or fly through tunnels. The animation was created in the following way. First, the various characters, objects and background scenes were hand-drawn and colored. 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 desired visibility of objects, and animated to create motion events. Following the animation stage, sound files were also incorporated into the slides, providing effects for onomatopoeia, and eventually linguistic stimuli.

An original narrative was designed to contextualize PPs and their modifiers, involving characters and scenes which are variations on the well-known Middle-Eastern folk-tale of Aladdin. In a cave filled with treasure, Aladdin takes a magic lamp from a wizard. He then jumps onto a magic carpet and flies up to an opening at the top 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 pass through various spatial environments, each of which provides a plausible context for a targeted combination of prepositions and modifiers. 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., with the wizard in pursuit. Eventually, Aladdin manages to escape from the wizard, and he releases the genie. In our version of the tale, he flies everywhere: reducing variability in manner of motion allowed for greater focus on trajectories.

The linguistic materials were presented in the form of two exercises: a preference task and a forced grammaticality judgment (GJ) task. The five types of modifiers discussed above were manipulated, as well as variations on PathP-PlaceP, and PlaceP-LocN. The types and tokens of spatial elements targeted in the piloting of the preference task are given in Table 1, and those used in the GJ task are given in Table 2 (three tokens were replaced to produce extra fillers). With two exceptions in the GJ task, only binary combinations of P-modifiers were used in both tasks, in order to reduce the processing burden.

Table 1. Pilot Exp 1: Preference task, targeted combinations

(a) DEG-FLOW (x3)	(a1) right on, (a2) straight on, (a3) right back
(b) DEG-TRAJECT (x3)	(b1) right up, (b2) straight through, (b3) right down
(c) FLOW-TRAJECT (x3)	(c1) on through, (c2) back over, (c3) back across
(d) MEASURE-FLOW (x2)	(d1) 10 km on, (d2) 10 km back
(e) MEASURE-TRAJECT (x2)	(e1) 1000m up, (e2) 1000m down
(f) ONOM-PathP (x3)	(f1) crash into, (f2) whoosh over, (f3) splash into
(g) Path-Place (x3)	(g1) from on top of, (g2) from in front of, (g3) from behind
(h) Place-LocN (x3)	(h1) in front of, (h2) in front of, (h3) on top of

Table 2. Pilot Exp 2: GJ, targeted combinations.

(a) DEG-FLOW (x2)	(a2) *on straight, (a3) right back
(b) DEG-TRAJECT (x2)	(b2) *through straight, (b3) right down
(c) FLOW-TRAJECT (x2)	(c1) *through on, (c2) back over
(d) MEASURE-FLOW (x1)	(d2) *back 10 km
(e) MEASURE-TRAJECT (x1)	(e1) 1000m up
(f) ONOM-PathP (x3)	(f1) *into crash, (f2) whoosh over, (f3) *into splash
(g) Path-Place (x3)	(g1) *on from top of, (g2) from in front of, (g3) *behind from
(h) Place-LocN (x3)	(h1) in front of, (h2) *front in of, (h3) *top on of
(i) MEASURE+DEG+TRAJECT (x1)	(i1) *straight 1000m down
(j) MEASURE+FLOW+DEGREE (x1)	(j1) *10km right on

In the preference task, following oral delivery of two variants of a sentence, learners selected one of two written orderings on their answer sheets. The order of presentation of targetlike and nontargetlike variants (whether they appeared as A or B answers) was systematically varied across stimuli. In the GJ task, learner responses were elicited following oral delivery of answers to questions in the form of sentence fragments. Again, the ordering of targetlike and nontargetlike forms was systematically varied across stimuli. The rationale behind the use of sentence fragments was to further control for prosodic reanalysis by subjects. For example, in the context of Aladdin continuing on through a tunnel to the outside, prosody can disambiguate between [*he flies through*] [*on to the outside*], which is grammatical, and *[*he flies*

[*through on to the outside*]], which is not, in contrast to [*he flies on through to the outside*]. However, prosody can be notoriously difficult in second language acquisition, and a sentence fragment answer such as *[*through on to the outside*], with appropriate prosody, reduces the chance of the P-modifier being reanalyzed as a verbal particle. The rationale behind the forced grammaticality judgments 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 grammaticality judgments. If a participant felt 75% sure of the grammaticality of the sentence, 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 judgments such as Likert scales, which increase the likelihood of prosodic rephrasing during sentence recall.

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. A set of teaching materials was presented to participants immediately preceding the experiment proper, with the purpose of making clear the meaning of each of the P-modifiers on the intended interpretations in English. 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 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*, *across* in prepositional contexts, the locative nouns *front* and *top*, and the onomatopoeic terms *crash*, *whoosh* and *splash*. 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 +1 PP), but they were tested on modifiers *in combination* (i.e. 2 or 3 modifiers + PP).

Protocol.

During the session, one experimenter was responsible for oral delivery of all instructions and stimuli, and another for the manipulation of visual images. Across both experiments, all oral stimuli were repeated once after a short pause. Given the importance of the visual stimuli to interpretation, on the presentation of each new slide we checked that all participants had their gaze directed toward the screen, and only following an agreed signal did the experimenter in control of the visual materials start the animation, which was then in sync with the oral stimulus. The third and fourth experimenters took note of whether participants followed instructions. The experiments were conducted during a 50-minute period, which began with the teaching session discussed above, followed by the preference judgment task, followed immediately by the GJ task. The session was brought to a close with a further instructional session so that participants (as students of English) could gain something from the experience.

4. Pilot results and consequences

The principle aim of the pilot was the refinement of test methodology in a novel area of investigation, and in this the pilot was very successful; given the relatively low numbers of tokens to types and the low number of learners, it was not expected that the results be robust enough to test for statistical significance, and indeed they were not. That said, the results were messier than expected, with dramatic variation across particular stimuli within combination-

types in both experiments. For example, in Deg-Flow combinations, accuracy rates varied from 40% (8/20) to 75% (15/20) in the preference task, and between 30% (6/20) and 85% (17/20) in the GJ task. One important outcome of such variation, visible to a greater or lesser degree in all combination types, was to raise our awareness of the importance of lexical effects in testing syntactic combinations. In Deg-Flow and Deg-Traject combinations, across the two experiments, accuracy rates on combinations with *right* (60%, 75%, 85%, 80%, 75%, 75%) were reliably more targetlike than accuracy rates on combinations with *straight* (40%, 30%, 50%, 35%).

Table 3. Results of Pilot Exps. 1 and 2: (a) Deg-Flow, (b) Deg-Traject, (c) Flow-Traject, (d) Measure-Flow, (e) Measure-Traject, (f) ONOM–P; (g) Path-Place, (h) Place-LocN.

	Slide no.	*Exp1 test items*	*%accuracy*	*Exp2 test items*	*%accuracy*
(a) DEG-FLOW		TOTAL	58.3%(35/60)	TOTAL	70%
(x3)	4	(a1) right on	60%(12/20)	(a1) FillPP	95%(19/20)
	11	(a2) straight on	40%(8/20)	(a2) *	30%(6/20)
	16	(a3) right back	75% (15/20)	(a3) ok	85%(17/20)
(b) DEG-TRAJECT		TOTAL	68.3%(41/60)	TOTAL	65%
(x3)	2	(b1) right up	80%(16/20)	(b1) FillPP	85%(17/20)
	13	(b2) straight through	50%(10/20)	(b2) *	35%(7/20)
	18	(b3) right down	75%(15/20)	(b3) ok	75%(15/20)
(c) FLOW-TRAJECT		TOTAL	43.3%(26/60)	TOTAL	48%
(x3)	3	(c1) on through	30%(6/20)	(c1) *	35%(7/20)
	17	(c2) back over	60%(12/20)	(c2) ok	65%(13/20)
	23	(c3) back across	40%(8/20)	(c3) FillPP	45%(9/20)
(d) MEASURE-FLOW		TOTAL	47.5%(19/40)	TOTAL	65%
(x2)	12	(d1) 10 km on	60%(12/20)	(j1) *M-D-F	40%(8/20)
	15	(d2) 10 km back	35%(7/20)	(d2) ok	90%(18/20)
(e) MEASURE-TRAJECT		TOTAL	30%(12/40)	TOTAL	55%
(x2)	5	(e1) 1000m up	30%(6/20)	(e1) ok	65%(13/20)
	7	(e2) 1000m down	30%(6/20)	(i2) *M-D-T	55%(11/20)
(f) ONOM-PathP		TOTAL	71.7%(43/60)	TOTAL	58%
(x3)	6	(f1) crash into	85%(17/20)	(f1) *	70%(14/20)
	10	(f2) whoosh over	65%(13/20)	(f2) ok	60%(12/20)
	21	(f3) splash into	65%(13/20)	(f3) *	45%(9/20)
(g) Path-Place		TOTAL	70%(42/60)	TOTAL	63%
(x3)	8	(g1) from on top of	60%(12/20)	(g1) *	65%(13/20)
	20	(g2) from in front of	80%(16/20)	(g2) ok	75%(15/20)
	22	(g3) from behind	70%(14/20)	(g3) *	50%(10/20)
(h) Place-LocN		TOTAL	96.7%(58/60)	TOTAL	68%
(x3)	9	(h1) in front of	100%(20/20)	(h1) ok	95%(19/20)
	19	(h2) in front of	100%(20/20)	(h2) *	75%(15/20)
	24	(h3) on top of	90%(18/20)	(h3) *	35%(7/20)
Non PP fillers:		TOTAL	100%(40/40)	TOTAL	90%
(x2)	14	my in / in my	100%(20/20)	the lamp	90%(18/20)
	25	the lamp / lamp the	100%(20/20)	the genie	90% (18/20)

Flow-Trajectory stimuli, as well as all combinations with Measure Phrases, furnished results around the level of chance, whilst the results with Onomatopoeia were more encouraging, with accuracy rates of between 65% and 85% in the preference task, and a higher acceptance of the grammatical variant in the GJ task, although, again, the results fell short of expectations. Other combinations of spatial elements tested were of the categories PathP-PlaceP and PlaceP-LocN, and five fillers were included. Participants generally scored highly on both types of spatial combination in the preference task (with averages of 70% (42/60) for PathP-PlaceP, and 96.7% (58/60) for PlaceP-LocN), and on two out of three tokens for each type in the GJ task, although *from behind* (**behind from*) and *on top* (**top on*) produced unexpected GJ results. GJs of the fillers also contained some surprises. Two participants gave nontartgetlike results for 3 out of 5 fillers, and although 4 of the 5 were unproblematic for participants in general, accuracy on one in particular was surprisingly low. During a scene in which Aladdin flies across a lake to a rock, the question was posed: *Where does Aladdin go?* The answer was the perfectly acceptable *To the rock*, which was rejected by 55% (11/20) of participants.

Our initial global interpretations of the pilot results were as follows. Of the three hypotheses set out in Section 3.1, Hypothesis (i) appeared likely to be disconfirmed, while hypotheses (ii) and (iii) remained to be evaluated as testing continued across the proficiency levels. Given the strong lexical effects that were attested, Hypothesis (ii) seemed to be the most plausible: after refining the methodology for the main experimentation, we expected that knowledge of the hierarchy would be manifested only in the wake of fine-grained acquisition of the lexicon, and of principles of prosody. However, several things persuaded us to keep Hypothesis 1 alive. One consideration was that whilst acceptance rates of ungrammatical examples were relatively high across the board, for each stimulus type there was invariably a higher acceptance rate for targetlike utterances than for nontargetlike utterances. An increase in the numbers of tokens used and the number of participants tested might make such differences more apparent. Another consideration was that, true to the purpose of the pilot, important methodological flaws were indentified leading to revisions in the protocol, in the auditory and visual stimuli, and in the response sheets used by participants. The most significant of the changes made on the basis of lessons learned from the pilot are listed below.

First, some students reported that it took several stimuli for them to understand with confidence what was required on the response sheet. This concern led to two changes: (i) the creation of two extra example slides, so that each task could be preceded with three examples; and (ii) the conscious articulation of a simple response sequence (LOOK, LISTEN, then WRITE), which aided tremendously in the smooth running of the experiments.

Second, it was soon realized that the controls we had in place for variations in prosody were woefully insufficient. Three problems were identified. The first was that the interactions between prosody and parsing were much more complex that we had at first assumed. For the intended interpretation we had to ensure that P-modifiers were not marked prosodically as verb particles. 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. Such differences did not correspond in one-to-one fashion with any of the types of syntactic variation we investigated, so we could not systematically predict which prosodic pattern would be the most unambiguous delivery of the materials. The felicity of particular prosodic patterns varied from item to item, apparently depending at least in part on the choice of lexical items. The

second problem was that despite an annotated script, much practice, and the best of intentions, the oral delivery of the stimuli by an experimenter resulted in small inconsistencies in prosodic delivery across initial and repeated utterances of stimuli, across items, and undoubtedly across the two experimental sessions. The third problem was that the written forms on the response sheets allowed participants to read as well as listen, either before during or after the oral stimulus, so that the individual's prosodic understanding of the stimulus was an uncontrolled variable. These three problems were tackled in the main experimentation by (i) selecting the most appropriate prosody for stimuli on an item-by-item basis, (ii) digitally recording all stimuli to be incorporated into the Powerpoint slides as sound files; and (iii) removing all written cues from the response sheets for both experiments, leaving only the letters (a) and (b). In this way, despite continuing concern about the current state of knowledge of the prosody-syntax interface, we were able to ensure the most felicitous prosodic phrasing for the stimuli based on native speaker intuitions, ensure the uniform delivery of stimuli in both experiments across many sessions, and restrict the linguistic stimulus to the oral delivery.

Third, to further distinguish the questions and sentence fragment answers of Experiment 2, it was decided to use a female voice for the questions, and a male voice for the answers. Not only did this serve its intended purpose, but it caused participants to visibly experience a fresh boost of attention at the beginning of the second experiment.

Fourth, following the discovery of strong lexical effects in the pilot, a concerted effort was made to balance types of P-modifiers within and across combination types. For example, in the main experiments Deg-Flow contained three tokens each of *right*, *straight on*, and *back* in various combinations. It must be noted that complete mathematical precision was not possible throughout the design, as all stimuli had to fit within the constraints of the narrative: each combination of modifiers had to be natural and appropriate for a particular pictorial context within the narrative.

Fifth, the ratios of tokens to types in the pilot were very low. To an extent, this is a problem endemic to a methodology which ties different types of linguistic stimuli to a visual narrative, because the confines of the narrative limit the number of tokens that can be presented. However, it is also a problem of too many types, so we decided to eliminate several categories as follows. Measure Phrases are problematic in that when they are combined with other P-modifiers, native speaker judgments are not clear cut. As the target language syntax remained undefined in this respect, it was decided to reserve Measure Phrases for future study. Combinations of PathP-PlaceP and PlaceP-LocN were also eliminated. Although performance on these items was impressive, it was decided that such responses were problematic in at least two ways. First, it is plausible that phrases such as *in front of* and *on top of* are learned as chunks, and accuracy is reflective of how well the collocation has been learned, rather than demonstrating knowledge of the internal structure of PP. It would be more revealing to study such combinations in a language that exhibits more productivity in the spatial noun system (e.g. Japanese, Korean). A second problem with these phrases is to understand how ungrammatical variants with this particular word order might be processed, as they also involve a third element, *of*, standardly analyzed as a semantically vacuous case-assigning P. A simple reversal of PlaceP-LocN strands *of* in its position close the head noun (e.g. *top on of the table*), but without any good reason to be there, as the P *on*, unlike the N *top*, can assign case directly to N. However, if *of* is deleted in such ungrammatical examples (e.g. *top on the table*), it is not obvious that this will be seen as a variant of the target structure at all. In sum, for the main experimentation, we chose to focus exclusively on the hierarchy of P-modifiers (Degree>Flow>Trajectory) and onomatopoeia.

5. Progress report: The genie out of the bottle

Although a full report of the main experimentation is pending, the initial results are intriguing enough to merit inclusion in this overview of the pilot study, as the refinement in experimentation wrought by the changes discussed above furnished a much more revealing set of responses. A total of 131 students from 6 different levels of proficiency took part in the study, with number of students by level as follows: L2-4, L3-15, L4-29, L5-45, L6-32, L7-6. There were 17 different L1s, listed as follows by numbers of native speakers in the study: A-Arabic (16), B-Bambara (2-bi), C-Chinese (14), F-French (2bi), H-Hungarian (2), J-Japanese (12), K-Korean (40), M-Mongolian (1) , P-Portuguese (4), R-Russian (2-bi), S-Spanish (5), Ta-Tajik (2), Th-Thai (1), Tm-Tamil (1), Tr-Tartar (1bi), Tu-Turkish (27), V-Vietnamese (2). The changes to the materials discussed above were implemented, and the stimuli were reformulated as shown in Table 4. The particular combinations used in Experiments 1 and 2 were the same, to facilitate triangulation of the results. For experiment 2, when categories allowed 6 tokens, 3 were grammatical and 3 ungrammatical, and when categories allowed only 3 tokens, 1 was grammatical and two were ungrammatical.

Table 4. Main experimentation: targeted combinations in Experiments 1 & 2.

(a) DEG-FLOW (x6)	(a1) straight on, (a2) straight on, (a3) right on, (a4) straight back, (a5) right back, (a6) right back
(b) DEG-TRAJECT (x6)	(b1) right up, (b2) right down, (b3) straight through, (b4) straight down, (b5) right out, (b6) straight out
(c) FLOW-TRAJECT (x3)	(c1) on through, (c2) on down, (c3) back over
(d) DEG-FLOW-TRAJECT (x3)	(d1) right on up, (d2) right back down, (d3) straight back across
(e) ONOM-PathP (x3)	(e1) crash into, (e2) whoosh over, (e3) splash into

The results, contrary to the prediction made after the pilot study, do not show a strong developmental pattern. It had been expected that accuracy on the hierarchy and on combinations with onomatopoeia would gradually increase as the lexicon was consolidated and principles of prosody were acquired, but even after consideration of certain lexical effects, the results are robustly above chance and remarkably flat (i) across the two tasks, (ii) across L1s, (iii) across the global proficiency range in general, and (iv) across the proficiency range within individual languages.

In order to measure effects related to proficiency, the program levels were conflated to derive 3 groups: Group 1 (Levels 2, 3, 4), Group 2 (Level 5), and Group 3 (Levels 6, 7). 10 subjects were eliminated from Experiment 1 and 13 from Experiment 2 following criteria applied to all subjects: persistent inattention, inaccurate responses on fillers, etc.

Because the stimuli were matched much more precisely across tasks than they had been in the pilot, it was possible to triangulate the results. Group responses in terms of accuracy rates are shown in Tables 5 and 6. The results for combinations of Deg-Flow and Deg-Traject and for the integration of onomatopoeia into syntax are particularly notable both in demonstrating that learners' performance on these aspects of the syntax of P-modifiers is well above chance at all levels of proficiency, and in their consistency across tasks. The responses for Flow-Traject were less revealing of knowledge of the proposed hierarchy, but the results once more were very consistent across tasks. The responses for Deg-Flow-Traject show slight but steady increases

across proficiency levels on both tasks, although on closer examination responses were difficult to interpret in the GJ task, with Group 1 performing around chance for two of the three items tested, and with 1 item in particular generating nontargetlike responses by all three groups.

Given the findings in the pilot concerning the possible importance of lexical effects, it is worth making brief note of such phenomena insofar as they can be found in the initial results of the main experimentation. In Groups 2 and 3, there appears to be no difference in response patterns with the items *right* and *straight*. However, differences do emerge on close examination of the data from Group 1, that is to say, those students at the same proficiency level as the pilot participants. A comparison between Group 1 and Group 3 responses with *right* reveal approximately the same scores; similar comparisons of responses with *straight* reveal jumps in accuracy rate between Groups 1 and 3. Such differences by item on the preference task are as follows: (a1) 62% to 79%; (a2) 62% to 79%; (a4) 79% to 92%; (b3) 83% to 92%; (b4) 46% to 54%; (b6) 59% to 81%. It is plausible that the relevant meaning of *straight* is acquired later than *right* in general, and some individuals may have persistent misunderstanding of this item.

Table 5. Main experimentation: Exp 1: Preference task initial results.

	Group means: % accuracy		
	Group1 (N=42)	Group2 (N=41)	Group3 (N=38)
(a) DEG-FLOW	76%	74%	84%
(b)DEG-TRAJECT	71%	78%	81%
(c) FLOW-TRAJECT	41%	38%	44%
(d) DEG-FLOW-TRAJECT	64%	68%	76%
(e) ONOM-PathP	83%	85%	89%

Table 6. Main experimentation: Exp 2: Grammaticality task initial results.

	Group means: % accuracy		
	Group1 (N=41)	Group2 (N=40)	Group3 (N=37)
(a) DEG-FLOW	76%	81%	80%
(b)DEG-TRAJECT	69%	77%	79%
(c) FLOW-TRAJECT	34%	39%	32%
(d) DEG-FLOW-TRAJECT	58%	63%	68%
(e) ONOM-PathP	87%	93%	90%

Another lexical effect might explain the particularly poor performance on Flow-Traject in the preference task. Participants treated items (c1) *on through* and (c2) *on down* very differently from (c3) *back over*. Accuracy rates by group for item (c1) were 43%, 20% and 42%; for (c2)

they were 12%, 15% and 8%; whilst for (c3) they were 67%, 80% and 82%. 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, the resultant forms could be interpretable with *through* or *down* either as verb particles or as P-modifiers, with *on* analyzed not as a modifier at all but as part of the complex preposition *onto*.

An initial analysis was also made of the performance of learners from different L1 backgrounds. The L1s with the most speakers were: Korean (40), Arabic (16), Turkish (27), Japanese (12), and Chinese (14). Learners from these L1s were almost indistinguishable from one another. For example, on the first three combination types in the preference task, the performance of these groups in the order listed above was as follows (in terms of means of accuracy scores): Deg-Flow: 78%, 79%, 73%, 80%, 77%; Deg-Traject: 78%, 82%, 80%, 60%, 73%; Flow-Trajectory: 37%, 49%, 40%, 37%, 48%. All language groups had the same discrepancy inside Flow-Trajectory (with c1 and c2 distinct from c3) discussed above. Within language groups, differences between proficiency levels were also negligible.

The analysis of Experiments 1 and 2 is far from complete, and the testing of control subjects is still underway. However, these preliminary results are striking enough to stand as evidence that the refinements made to the pilot experiment were successful, furnishing a surprising and robust set of findings concerning the L2 acquisition of P-modifiers.

6. Conclusion

In an extension of research on universals in PP structure, five types of modifying elements within spatial PP were identified, three of which are found in a fixed structural hierarchy, but not all of which are lexicalized in all languages. A learnability problem was identified for second language learners: how could knowledge of the syntactic properties of P-modifiers be acquired in the absence of L1 transfer, lack of instruction, and a paucity of such combinations in naturalistic input? A preference task and a grammaticality judgment task were designed and piloted on a group of lower-intermediate learners of English, with the primary purpose of testing originally developed experimental materials in a hitherto unexplored area of the acquisition of syntax, so as to refine the methodology for large-scale testing. In this, the pilot was successful, laying the groundwork for future research in this domain. The pilot results, however, were inconclusive. The interim interpretation was that the hierarchy is not in evidence in the initial stages of L2 acquisition, and it seemed plausible that accuracy was dependent on gradual acquisition of the lexicon. Although analysis of the main experimentation is still in progress, preliminary results contradict the expectations created by the pilot study. It appears that evidence for knowledge of the syntactic hierarchy on the part of L2 learners does not appear piecemeal over time, but is in existence from the outset. Quite surprisingly, accuracy rates are resolutely above chance and at extremely similar levels across the two tasks, across L1s, and across the proficiency range. Knowledge of the hierarchy of spatial modifiers is indeed in evidence, despite the poverty of the stimulus, suggesting a path straight on through to Universal Grammar.

Acknowledgements

This research was conducted with the collaboration of the directors, instructors and students of the Intensive English Program (IEP) at Indiana University, as well as the staff of the Center for Language Technology and Instructional Enrichment (CeLTIE). Many thanks especially to Doreen Ewert, Heidi Vellenga, Roger Crandall, and Lisa McCune. We would also like to thank Tara Zahler, for her wonderful voicework, and Bonnie Schwartz for insightful discussion as we moved from pilot to main experimentation. Discussion with conference participants at BUCLD 2008 was extremely useful, and ideas generated in exchanges with Kamil Ud Dean, Bonnie Schwartz, William Snyder, Lydia White and others continue to inspire our research on this topic. A final thank you is due to the organizers of the conference, who not only helped to ensure the continued excellence of this venue, but also made such a concerted effort to turn it green.

References

Berman, R.A. & Slobin, D.I. (Eds.) (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Haspelmath, M. (1993). *A Grammar of Lezgian*. Berlin: Mouton de Gruyter.

Hohenstein, J.M., Naigles, L.R., and Eisenberg, A.R. (2004). Keeping verb acquisition in motion: A comparison of English and Spanish. In D.G. Hall and S.R. Waxman (Eds.), *Weaving a Lexicon*. Cambridge, MA: MIT Press.

Jespersen, O. (1992 [1924]). *The Philosophy of Grammar*. Chicago: The University of Chicago Press.

Koopman, H. (2000). Prepositions, postpositions, circumpositions and particles: The structure of Dutch PPs. In H. Koopman, *The Syntax of Specifiers and Heads*. London: Routledge

Pourcel, S. (2008). Investigating linguistic relativity: a research methodology. *Durham Working Papers in Linguistics*, 8: 125-138.

Riemsdijk, H. van & Huybregts, R. (2007). Location and Locality. In S. Karimi, V. Samiian & W. Wilkins (eds.). *Clausal and Phrasal Architecture: Syntactic Derivation and Interpretation: A Festschrift for Joseph E. Emonds*. Amsterdam: John Benjamins.

Stringer, D. (2005). *Paths in First Language Acquisition: Motion through Space in English, French and Japanese*. Doctoral dissertation, University of Durham.

Stringer, D. (2007). Extending the PP hierarchy: The role of bare nominals in spatial predication. In T. Scheffler, J. Tauberer, A. Eliam, and L. Mayol (eds.), *Penn Working Papers in Linguistics (PWPL) Vol.13.1: Proceedings of the 30th Annual Penn Linguistics Colloquium*: 379-392.

Stringer, D. (2008). Spatial affixes in syntax. Paper presented at Berkeley Workshop on Affix Ordering, University of California, Berkeley, October 4.

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 A. Asbury, J. Dotlačil, B. Gehrke and R. Nouwen (Eds.), *Syntax and Semantics of Spatial P*. Amsterdam: John Benjamins.

Talmy, L. (1991). Paths to realization: A typology of event conflation. *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*: 480-519.

Talmy, L. (2000). *Toward a Cognitive Semantics, Vol. II: Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press.