The Psycholinguistics of the Interaction Hypothesis:
Balancing Usage-Based and Form-Focused SLA

Nick Ellis
Research Scientist, English Language Institute
Professor of Psychology
University of Michigan

ncellis@umich.edu

Invited AAAL symposium
Multiple Perspectives on Interaction in SLA
Costa Mesa, CA April 21-24, 2007
We learn Constructions

- Constructions as basic symbolic units of language representation:
  - Form meaning mappings
  - Conventionalized in the speech community
  - Entrenched as language knowledge in the learner’s mind

- Usage-based acquisition
  - We learn constructions through using language, engaging in communication.

- Emergence
  - Creative linguistic competence emerges from the collaboration of the memories of all of the utterances in a learner’s entire history of language use and the frequency-biased abstraction of regularities within them

- Cognitive Linguistics
- Functional Linguistics
- Psycholinguistics
- Can’t separate:
  - Grammar from lexis
  - Form from meaning
  - Meaning from context

Constructions

Today he walks to town

morphological, syntactic, lexical form

semantic, pragmatic, discourse functions

Fig. 1. (a) An expression, or ‘construction’, that is a combination of the constructions shown in (b), color-coded to the appropriate parts of the expression (VP, Verb-Phrase; NP, Noun-Phrase). See text for discussion.

ΔP, the one-way dependency statistic (Allan, 1980), drives associative learning (Shanks)

\[ \Delta P = P(O|C) - P(O|-C) \]
\[ = \frac{a}{a+b} - \frac{c}{c+d} \]

\(a, b, c, d\) represent frequencies
\(a\) is the frequency of conjunctions of the cue and the outcome, etc.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cue</td>
<td>a</td>
</tr>
<tr>
<td>No cue</td>
<td>c</td>
</tr>
</tbody>
</table>

\[ \Delta P = \frac{P(O/C)}{P(O/-C)} = \frac{a}{a+b} - \frac{c}{c+d} \]
Connectionist Networks, Construction Contingency, Competition

Complex Constructions
Dynamic - in Context, in Time

- Cognitive Linguistics and usage-based models emphasize:
  - Language is learned from experience of processing input and producing language
  - During interaction in social contexts
  - Where personally relevant non-linguistic outcomes (e.g., “a cup of tea”) are goals to be achieved
  - By communicating intentions, concepts and meaning with others

Cognitive Grammar - Construal and Attention

- Constructions are conventionalized linguistic means for presenting different conceptualizations and construals of an event.

- Language structures concepts and windows attention to aspects of experience through the options specific language make available to speakers (Talmy, 2000)

- Cross-linguistic research into how different languages lead speakers to prioritize different aspects of events in narrative discourse (Berman and Slobin, 1994)

English is a satellite-framed language. English verbs use particles to show the path of motion and its verbs usually show manner of motion

"he ran in"

Spanish (along with all the Romance languages) is a verb-framed language.

"entró corriendo"

literally 'he entered running'

Embodied Participation in Dynamic Situated Action

- Necessity of participation in situated action to learners' full understanding of L2 constructions.
- Meaning of spatial language does not simply derive from the addition of fixed meanings prepositions have for 'where' an object is to the meanings of other elements in the sentence describing 'what' is being located which can be taught by L2 rule and learned by rote.
- Meaning flexibly constructed on-line.
- Multiple constraints involving object knowledge, dynamic-kinematic routines and functional geometric analyses which come together in contexts that embody meaning for language users.

Embodied Networks
Transfer & Learned Attention

- As Usage-based theory views linguistic knowledge as a set of automatized patterns which are schematic to varying degrees, the first language must be viewed as both a help and a hindrance to second language acquisition.
- To the extent that the constructions in the second language are similar to those of the first language, the L1 constructions can serve as the basis for the L2 constructions.
- However, since even similar constructions across languages are likely to differ in detail, the acquisition of the L2 pattern in all its detail is hindered by the L1 pattern.


Rescorla & Wagner (1972) expressing the capacity of a cue (CS) become associated with an outcome (US) on any learning trial

$$dV = ab(L - V)$$

- $V$ the associative strength of the US to the CS
- $dV$ the change in this strength which occurs on each learning trial
- $a$ the salience of the CS
- $b$ the salience of the US
- $L$ the amount of processing given to a completely unpredicted US.

- Learning depends on the salience of the cue ($s$)
- Learning depends on the importance of the outcome (interpretation - 3rd person, present tense))
  - [Simon Satisficing: learners make decisions not by "maximization" by "satisficing", i.e. setting an aspiration level which, if achieved, they will be happy enough with, and if they don't, try to change either their aspiration level or their decision.]

- The more predicted the US (from context and other cues), the less the additional association on this trial
The very things that make L1 easy, make L2 hard

- Natives
  - Produce grammatical functors (because of their high frequency) automatized & fluently, eroded and non-salient.
  - These are fluently perceived by expectation-driven means, supplementing the weak data-driven source and its low contingency with outcome

- Non-natives
  - Find them difficult to perceive (low salience), hard to learn (low contingency)
  - Have no top-down support in their perception
  - Have various cognitive biases, as a result of their L1 experience, which tune their attention away from these cues

- The very things that make an L1 easy for a fluent speaker (automaticity and top-down knowledge) make L2A hard (non-salient cues, no top-down support)

Attention-Gated Networks

Dealing with Transfer (1): Consciousness

- **Interface** of Explicit and Implicit Learning
  - Explicit learning to overcome L1 learned attention, automaticity, and transfer
  - Consciousness as the Interface

Consciousness and Learning

- “The more novelty we encounter, the more conscious involvement is needed for successful learning and problem-solving” (Baars)

- The NCC involve a coalition of forebrain neurons implicated in working memory, interconnected via widespread cortico-cortico and cortico-thalamic feedback loops with sets of neurons in sensory and motor regions coding for particular features.

Global Workspace theory
Global Workspace theory is a cognitive architecture with an explicit role for consciousness (Fig. 2). It makes minimal assumptions:

1. that the brain can be viewed as a collection of distributed specialized networks (processors);
2. that consciousness is associated with a global workspace in the brain – a fleeting memory capacity whose focal contents are widely distributed (‘broadcast’) to many unconscious specialized networks;
3. conversely, a global workspace can also serve to integrate many competing and cooperating input networks;

- ○ Conscious, accurately reportable with qualitative content
- ● Fringe conscious, accurately reportable without qualitative content
- ★ Unconscious, not accurately reportable
Consciousness & Learning

- Implicit learning from usage occurs largely within modality and involves the priming or chunking of representations or routines within a module; it is the means of tuning our zombie agents, the menagerie of specialized sensori-motor processors that carry out routine operations in the absence of direct conscious sensation or control. It is largely automatized. It operated in parallel.

- Conscious processing is spread wide over the brain and unifies otherwise disparate areas in a synchronized focus of activity. Conscious activity affords much more scope for focused long-range association and influence than does implicit learning. It brings about a whole new level of potential associations. It operates serially.
Consciousness is the interface (Ellis, SSLA, 2005)

- Consciousness is the publicity organ of the brain. It is a facility for accessing, disseminating, and exchanging information and for exercising global coordination and control.

- “Paying attention—becoming conscious of some material—seems to be the sovereign remedy for learning anything, applicable to many very different kinds of information. It is the universal solvent of the mind”. (Baars)
Dealing with Transfer (2): The Dialectics of Dialogue

- Interface of Explicit and Implicit Learning
  - Explicit learning to overcome L1 learned attention, automaticity, and transfer
  - Consciousness as the Interface

- CREED Dialectic Dialogic
  - Consciousness is socially shaped in interaction

---

Socially-Gated Networks

Language and Social Interaction

Contingent social interaction but not simple exposure changes phonetic discrimination after 9–10 months

a Nine-month-old American infants being exposed to Mandarin Chinese in twelve 25-min live or televised sessions.

b After exposure, infants in the Mandarin exposure groups and those in the English control groups were tested on a Mandarin phonetic contrast using the head-turn technique.

c Results show phonetic learning in the live-exposure group, but no learning in the TV- or audio-only groups.

Social Feedback

- Social feedback modulates the quantity and quality of utterances of young infants.
- Mothers’ responsiveness to their infants’ vocalizations was manipulated after a baseline period of normal interaction:
  - Half of the mothers were instructed to respond immediately to their infants’ vocalizations by smiling, moving closer to and touching their infants: these were the ‘contingent condition’ (CC) mothers.
  - The other half of the mothers were ‘yoked controls’ (YC) — their reactions were identical, but timed (by the experimenter’s instructions) to coincide with vocalizations of infants in the CC group.
- Infants in the CC group produced more vocalizations than infants in the YC group, and their vocalizations were more mature and adult-like.

Interaction and Speech Cues

Cue: \( S \)

Construction

- outcome
- 3rd person singular simple present
Speech to Learners: Motherese

- ‘Motherese’ or Caretaker speech
- a special speech ‘register’ used to talk to infants and children, we use that has a unique acoustic signature
- Caretakers in most cultures use it when addressing infants and children.
- When compared to adult-directed speech, infant-directed speech is
  - slower,
  - has a higher average pitch
  - contains exaggerated pitch contours

Exaggerated Input

- Infant-directed speech might assist infants in learning speech sounds.
- Women speaking English, Russian or Swedish were recorded while they spoke to another adult or to their young infants.
- Acoustic analyses showed that the vowel sounds (the /i/ in ‘see’, the /a/ in ‘saw’ and the /u/ in ‘Sue’) in infant-directed speech were more clearly articulate.
- Women from all three countries exaggerated the acoustic components of vowels (see the ‘stretching’ of the formant frequencies, creating a larger triangle for infant-directed, as opposed to adult-directed, speech).
- This acoustic stretching makes the vowels contained in motherese more distinct.

Speech clarity aids Learner Speech Discrimination

- Mothers who stretch vowels to a greater degree have infants who are better able to hear subtle distinctions in speech.

Foreigner Talk

- Infant-directed speech has three main roles
- attracts attention ~ pitch
- conveys emotional affect
- conveys language-specific phonological information ~ vowel hyperarticulation.

- Compared British English speech directed to first language English learners (infants), and second language English learners (adult foreigners), populations with similar linguistic but dissimilar affective needs.
- Vowels were equivalently hyperarticulated in infant- and foreigner-directed speech.
- Pitch higher in speech to infants than to foreigners or adult British controls.
- Positive affect affect was highest in infant-directed and lowest in foreigner-directed speech.
- Linguistic modifications found in both infant- and foreigner-directed speech are didactically oriented, and that linguistic modifications are independent of vocal pitch and affective valence.

Foreigner Talk

- Speakers adjusted conversational tempo according to the status of their listeners.
- They talked more slowly to foreigners than to native speakers and produced longer vowels.
- FDS is indeed an acoustically distinct speech style from standard native-directed speech.
- Adjustments consistent with those seen in other listener-directed speech styles:
  - Speakers produce a signal that is clearer and easier to process when speaking to listeners who may have had extra processing difficulties (in this case, due to limited language experience).

Speech not just an auditory signal
Embodied Speech easier to perceive

- Hardison (2002) found somewhat better learning of /r/ and /l/ by Japanese and Korean speakers when training involved a frontal view of the talker than simply auditory speech.

  - Baldi for teaching non-native phonetic contrasts, by comparing instruction illustrating the internal articulatory processes of the oral cavity versus instruction providing just the normal view of the tutor’s face.
  - Eleven Japanese speakers of English as a second language were bimodally trained under both instruction methods to identify and produce American English /r/ and /l/ in a within-subject design.
  - Speech identification and production improved under both training methods.
  - A generalization test showed that this learning transferred to the production of new words.


Interaction and Interpretation

Construction

outcome interpretation
Intentionality

Understanding and sharing intentions: The origins of cultural cognition

Michael Tomasello, Malinda Carpenter, Joseph Call, Tanya Belaev, and Henrik Moll

Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

tomas@eva.mpg.de, carpenter@eva.mpg.de, call@eva.mpg.de, belaev@eva.mpg.de, moll@eva.mpg.de
http://www.eva.mpg.de/predprenser/

Figure 1. Human intentional action. The goal is an open box; reality is a closed box. The actor chooses a means (plan), depicted as handle doing things, which forms an intention. The resulting action causes a result, which leads to a reaction from the actor.

Figure 2. Each partner’s conception of a collaborative activity in which a shared goal and joint intention (with complementary roles) are formed.
Intention Reading
Dyadic Situated Interaction

<table>
<thead>
<tr>
<th>6ms</th>
<th>12ms</th>
<th>18ms</th>
<th>2yrs</th>
<th>......</th>
<th>adulthood</th>
</tr>
</thead>
</table>

- **attention**
- **detection**
  - gaze
  - following

- **attention**
- **manipulation**
  - directive
  - pointing

- **intention**
- **understanding**
  - others are
  - goal-directed

**social**
**coordination**
**with shared**
**intentionality**
  - joint activities,
  - shared interest,
  - meaning negotiation

---

<table>
<thead>
<tr>
<th>Subject Group</th>
<th>Eye Presence</th>
<th>Simple Gaze</th>
<th>Gaze Following</th>
<th>Joint Attention</th>
<th>Mental Attribution</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish</td>
<td>✔</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
<td>[41]</td>
</tr>
<tr>
<td>Birds</td>
<td>✔</td>
<td>✔</td>
<td>?</td>
<td>✔ (?)</td>
<td>?</td>
<td>[59, 65, 111, 130, 133, 134]</td>
</tr>
<tr>
<td>Dogs (domestic)</td>
<td>?</td>
<td>✔</td>
<td>✔</td>
<td>?</td>
<td>?</td>
<td>[66, 100]</td>
</tr>
<tr>
<td>Prosimians</td>
<td>?</td>
<td>✔</td>
<td>X</td>
<td>?</td>
<td>?</td>
<td>[5, 40, 75]</td>
</tr>
<tr>
<td>Monkeys</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>X</td>
<td>[3, 4, 53, 75, 76, 83, 89, 96, 115, 132, 141, 143]</td>
</tr>
<tr>
<td>Great Apes</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>X (? )</td>
<td>[33, 61, 62, 75, 77, 91, 110, 120, 122–125, 143]</td>
</tr>
<tr>
<td>Human</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>—3 months</td>
<td>✔</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>[9, 27, 28, 30, 39, 45, 77, 106, 135, 139]</td>
</tr>
<tr>
<td>—9 months</td>
<td>✔</td>
<td>✔</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>[9, 27, 28, 30, 39, 45, 77, 106, 135, 139]</td>
</tr>
<tr>
<td>—12 months</td>
<td>✔</td>
<td>✔</td>
<td>X</td>
<td>✔</td>
<td>X</td>
<td>[9, 27, 28, 30, 39, 45, 77, 106, 135, 139]</td>
</tr>
<tr>
<td>—18 months</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>X</td>
<td>[9, 27, 28, 30, 39, 45, 77, 106, 135, 139]</td>
</tr>
<tr>
<td>—24 months</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>X</td>
<td>[9, 27, 28, 30, 39, 45, 77, 106, 135, 139]</td>
</tr>
<tr>
<td>—48 months</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td>[9, 27, 28, 30, 39, 45, 77, 106, 135, 139]</td>
</tr>
<tr>
<td>—Autism</td>
<td>✔ (?)</td>
<td>✔</td>
<td>?</td>
<td>X</td>
<td>X</td>
<td>[9, 11, 14, 15, 37, 93, 94]</td>
</tr>
<tr>
<td>—Down syndrome</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td>[9]</td>
</tr>
<tr>
<td>—Amygdala damage</td>
<td>✔ (?)</td>
<td>X</td>
<td>X (? )</td>
<td>?</td>
<td>X (?)</td>
<td>[138, 159]</td>
</tr>
</tbody>
</table>

Development of social gaze

A. Mutual versus Averted Gaze
B. Gaze Following
C. Joint Attention
D. Shared Attention
E. "Theory of Mind"

From Emery (2000)
Referential Indeterminacy

- Quine (1960): Single words cannot be paired with experiences, since they confront experience in clusters.
- Other things being equal, it is natural to translate the word as 'rabbit'. But why not translate it as, say, 'undetached rabbit-part'? For any experience which makes the use of 'rabbit' appropriate would also make that of 'undetached rabbit-part' appropriate.

“Gavagai”

Shared attention, noticing, and scaffolding, solve the problem of referential indeterminacy

The Role of Embodied Intention in Early Lexical Acquisition

- Influence of inferring interlocutors’ referential intentions from their body movements at the early stage of lexical acquisition.
- Embodied intentions facilitate both word discovery and word-meaning association.

Fig. 1. The snapshots when the speaker uttered “The cow is looking at the little boy” in Mandarin. Left: No non-speech information in audio-only condition. Center: A snapshot from the fixed camera. Right: A snapshot from a head-mounted camera with the current gaze position (the white arrow).

Fig. 2. The mean percentages of correct answers in tests.
The Role of Embodied Intention in Early Lexical Acquisition

Fig. 3. The level of synchrony between eye movement and speech production. Most spoken object names were produced after eye fixations, and some of them were uttered before eye fixations. Occasionally, the speaker did not look at the objects at all when he referred to them in speech. Thus, there is no perfect synchrony between eye movement and speech production.

Fig. 5. The problems in word learning. The raw speech is first converted into phoneme sequences. The goal of our method is to discover phoneme substrings that correspond to the sound patterns of words and then infer the meanings of these words from nonlinguistic modalities.
The Role of Embodied Intention in Early Lexical Acquisition

- Computational model that can identify the sound patterns of individual words from continuous speech, using nonlinguistic contextual information, and employ body movements as deictic references to discover word-meaning associations.

- First model of word learning that not only learns lexical items from raw multisensory signals to closely resemble infant language development from natural environments, but also explores the computational role of social cognitive

Fig. 9. Overview of the method for word learning. Spoken utterances are categorized into several bins that correspond to temporally co-occurring attentional objects. Then we compare any pair of spoken utterances in such bins to find the similar subsequences that are treated as wordlike units. Next, those wordlike units in such bins are clustered based on the similarity of their phoneme strings. The expectation-maximization (EM) algorithm is applied to find lexical items from hypothesized word-meaning pairs.
Embedded Networks

Constructions
Connectionist Nets
Embodied
Attentional
Consciousness
Social-Interactional
Instructional & Cultural Context