Computational Cognitive Morphosemantics

Modeling morphological compositionality in Hebrew verbs with Embodied Construction Grammar

Nathan Schneider ~ BLS 36 ~ 7 February 2010

http://www.cs.cmu.edu/~nschneid/bls36-slides.pdf

Overview

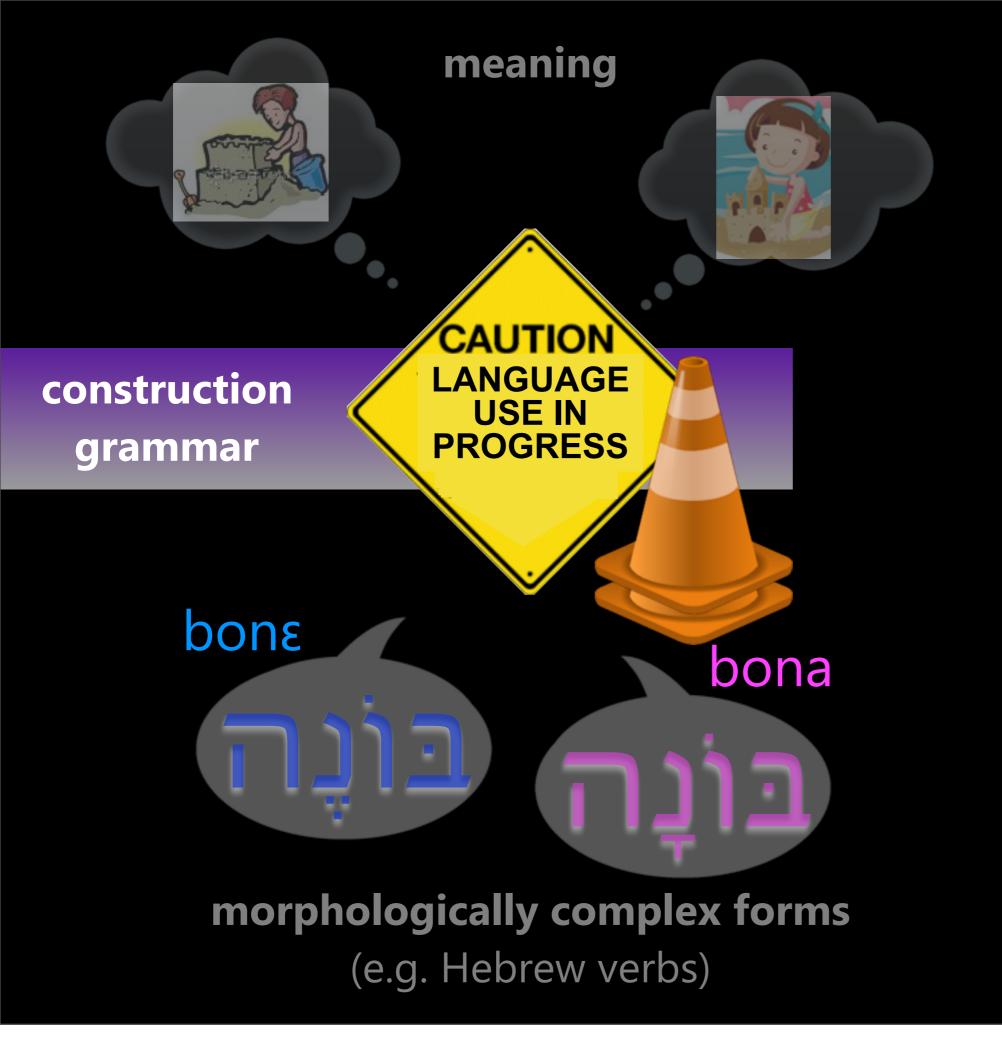
- An analysis of Hebrew verbs linking cognitive semantics to complex morphological constructions
 - Must account for compositionality as well as idiosyncrasy
- Cast within the ECG formalism to facilitate computational processing
 - Previously, ECG was only used for syntax

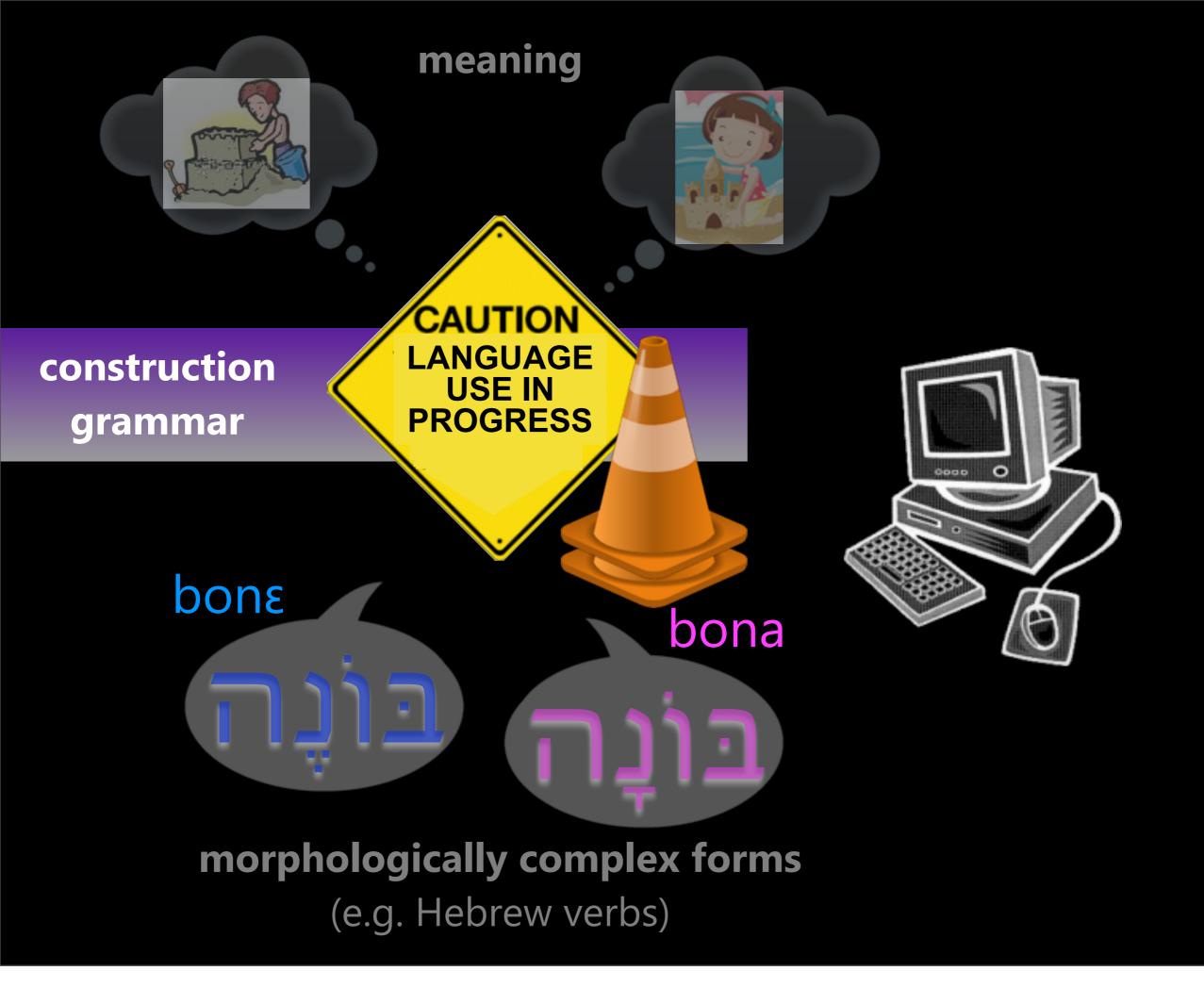


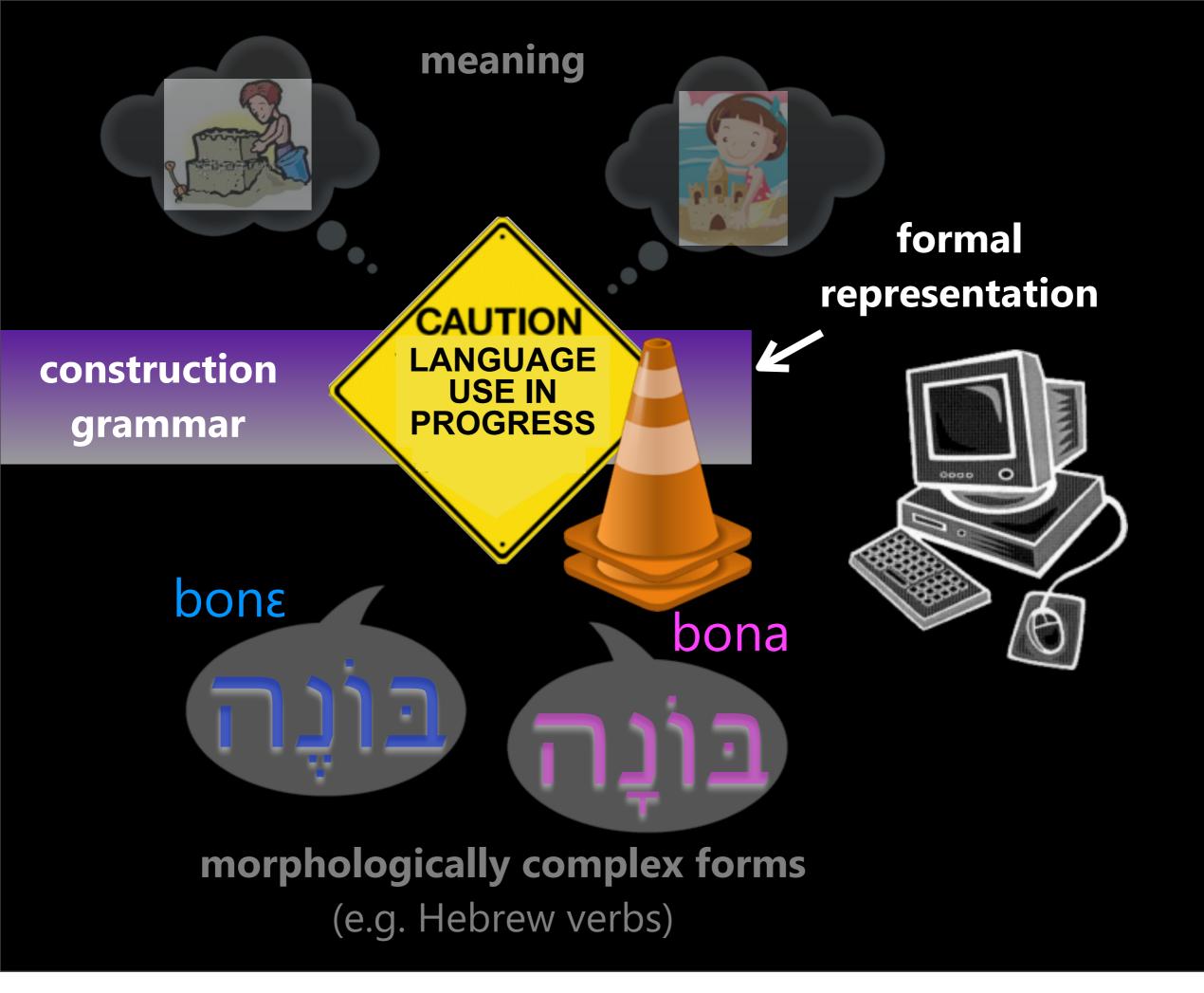


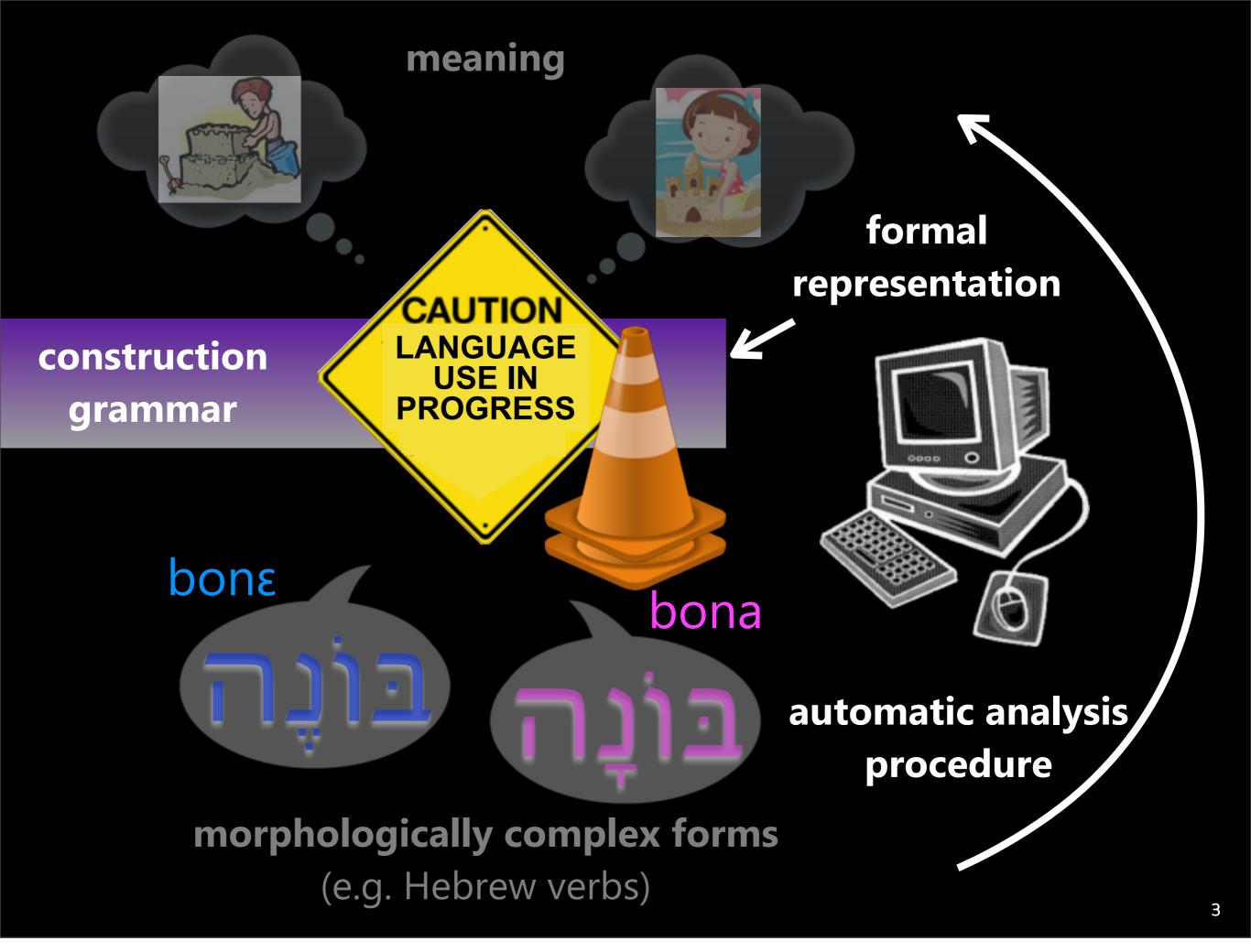


morphologically complex forms (e.g. Hebrew verbs)









Construction Grammar for Hebrew Verb Morphology

- Challenges:
 - Nonconcatenative morphology
 - Semantics of roots, paradigms, and verbs (whether compositional or idiosyncratic)
- I will use Embodied Construction Grammar, a formalism designed to support computational analysis and simulation of sentences

Construction Grammar

- In the family of cognitive theories known as Construction Grammar, there is no separation between lexicon and grammar
- Words, lexical categories, multiword expressions, syntactic phrases, idioms all form-meaning pairs: constructions, albeit with different levels of generality
- Usage-based theories of grammar: constructions may be stored redundantly in memory ("constructicon"); sensitive to factors such as frequency

e.g.: [Fillmore et al. 1988] [Kay & Fillmore 1999] [Goldberg 1995, 2006] [Langacker 1990] [Croft 2001] [Tomasello 2003]

Other Related Work

Formal Approaches to Semitic/Nonconcatenative Morphology

[McCarthy 1979] proposed an autosegmental analysis for the root-pattern morphology of Arabic. [Finkel & Stump 2002] used inheritance in the KATR formalism to describe Hebrew verb forms. For other approaches to nonconcatenative morphology, see [Orgun 1996] [Rubba 2001] [Roark & Sproat 2007].

Morphology in Construction Grammar

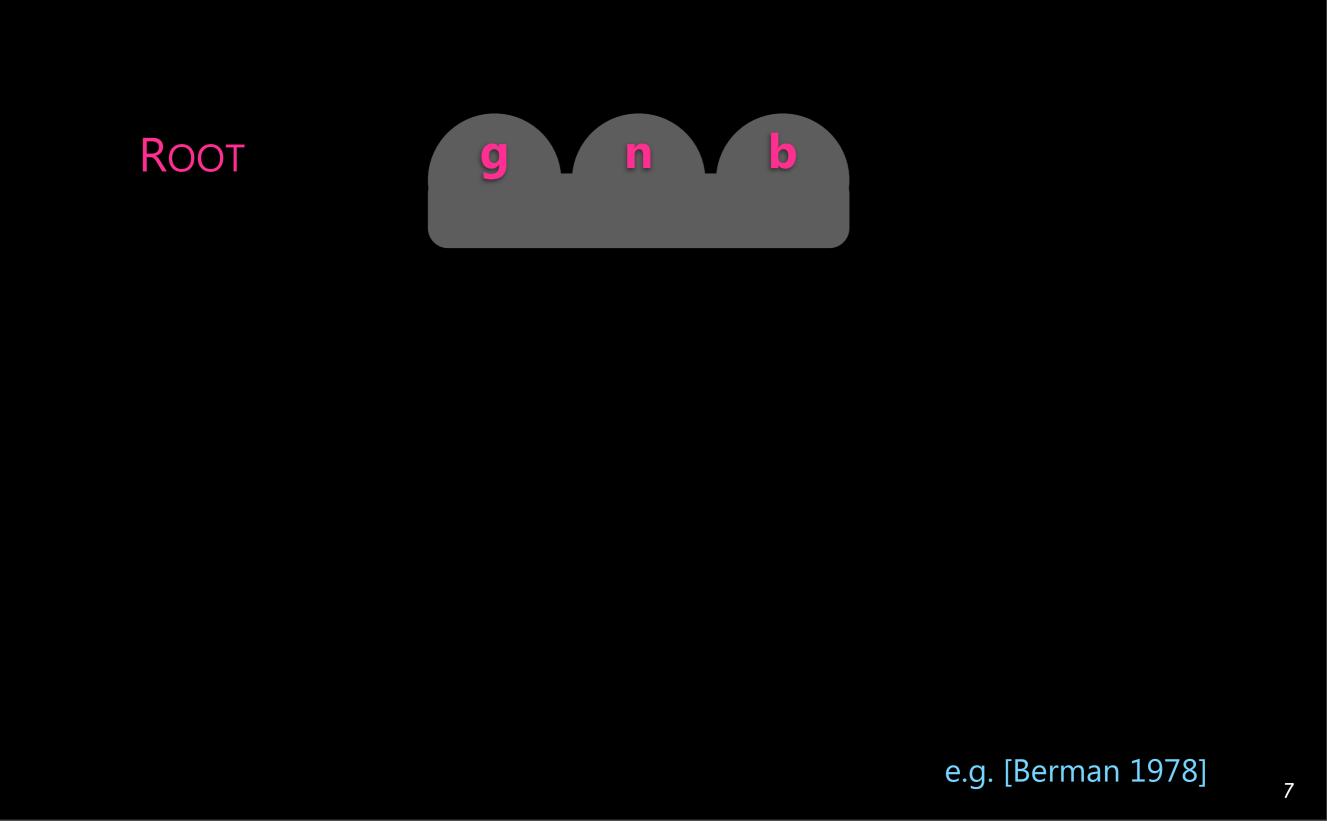
Previous work has described composition of morphological constructions [Riehemann 1998] [Booij 2005, 2007] [Gurevich 2006]. Several mechanisms for adding morphology to ECG were entertained in [Bergen 2003], but none were implemented. [Rubba 1993] (synopsis in [Rubba 2001]) takes a Cognitive Grammar approach to nonconcatenative morphology, situating words in a network (cf. [Bybee 1985, 2001]). Two other relevant approaches to phonology are found in [Inkelas 2008] and [Nathan 2007]. [Mandelblit 1997] offers an extensive semantic account of Hebrew verb paradigms.

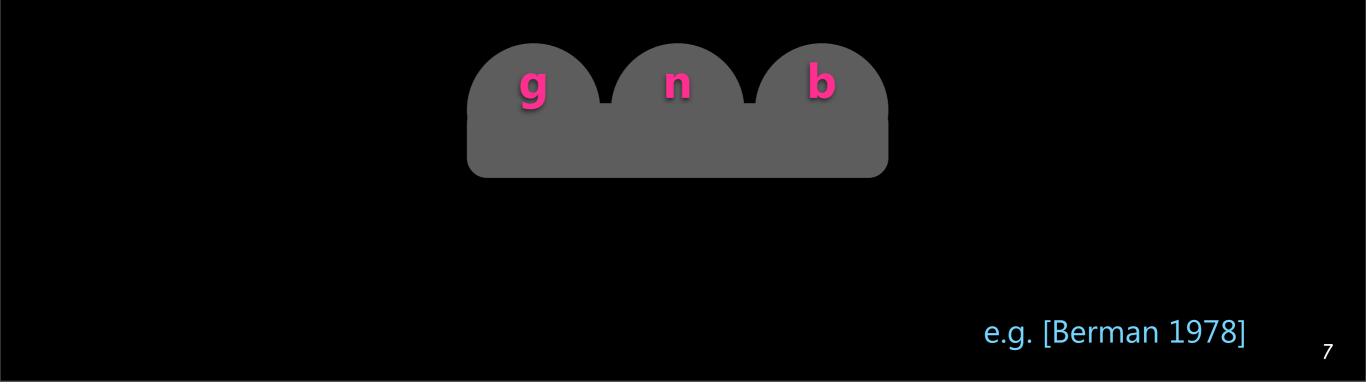
6

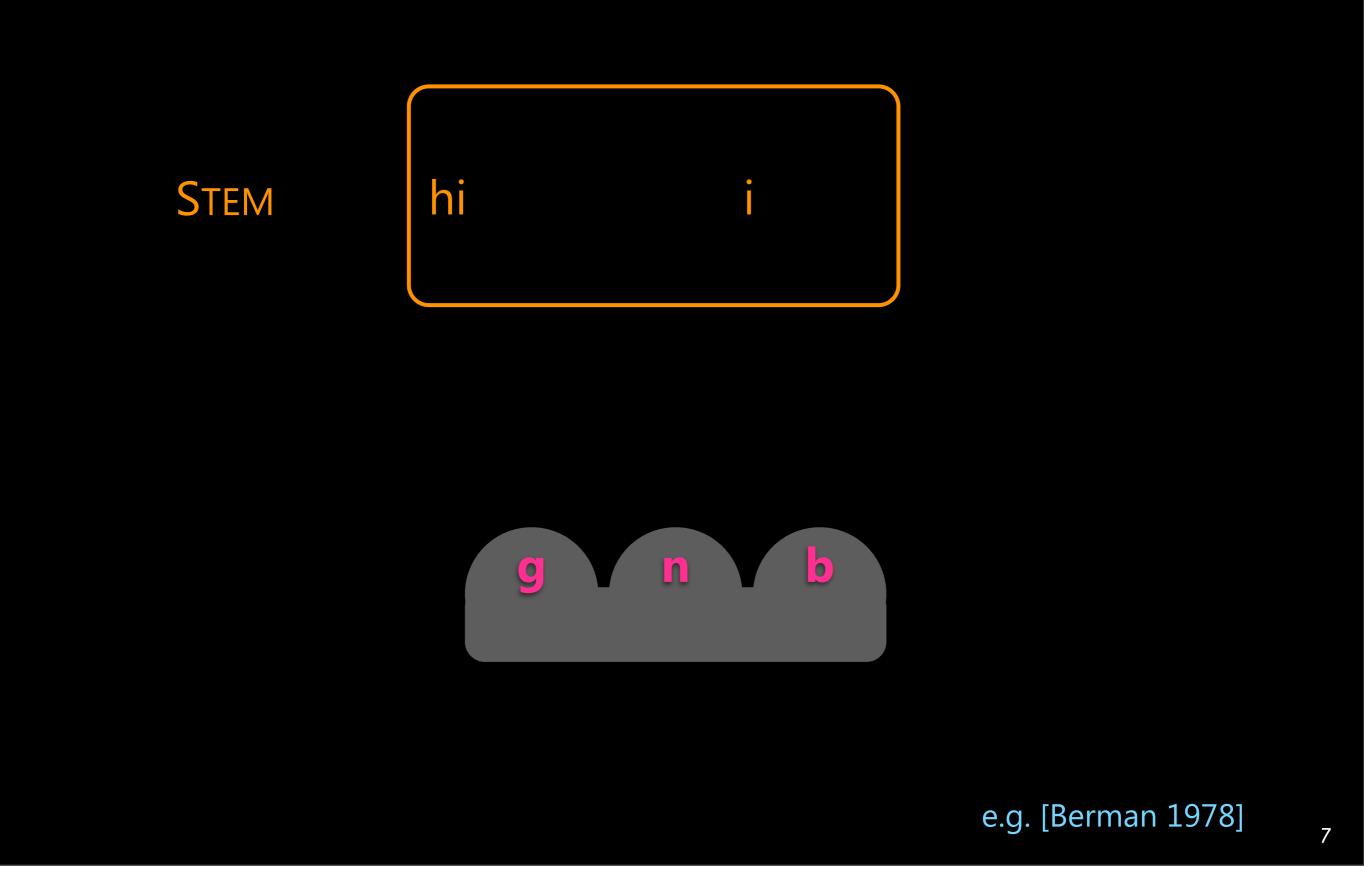
There has been a great deal of formal work on Semitic morphology using a variety of approaches, including rules, autosegmental phonology, and unification grammars. There has also been some work on morphology in Construction Grammar and related theories. To my knowledge, this is the first work to explicitly combine detailed semantic representations from cognitive linguistics in a formal description of morphological constructions.

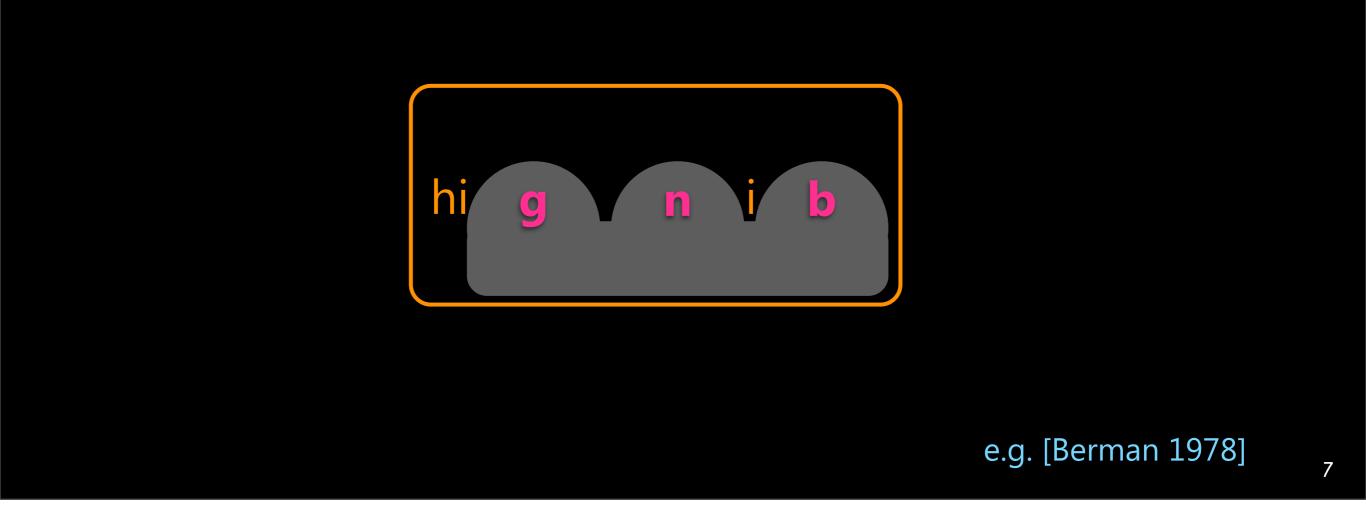
e.g. [Berman 1978]

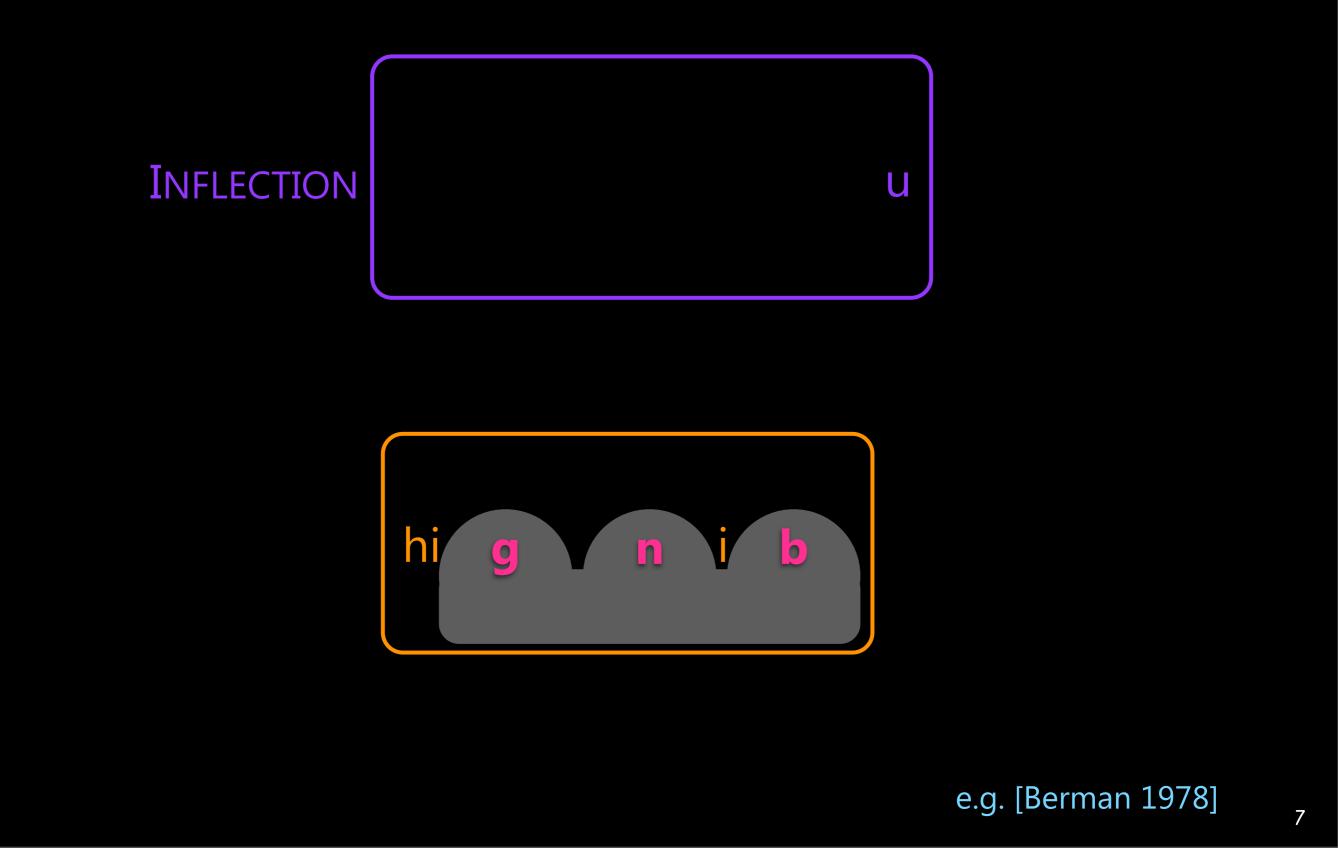
7

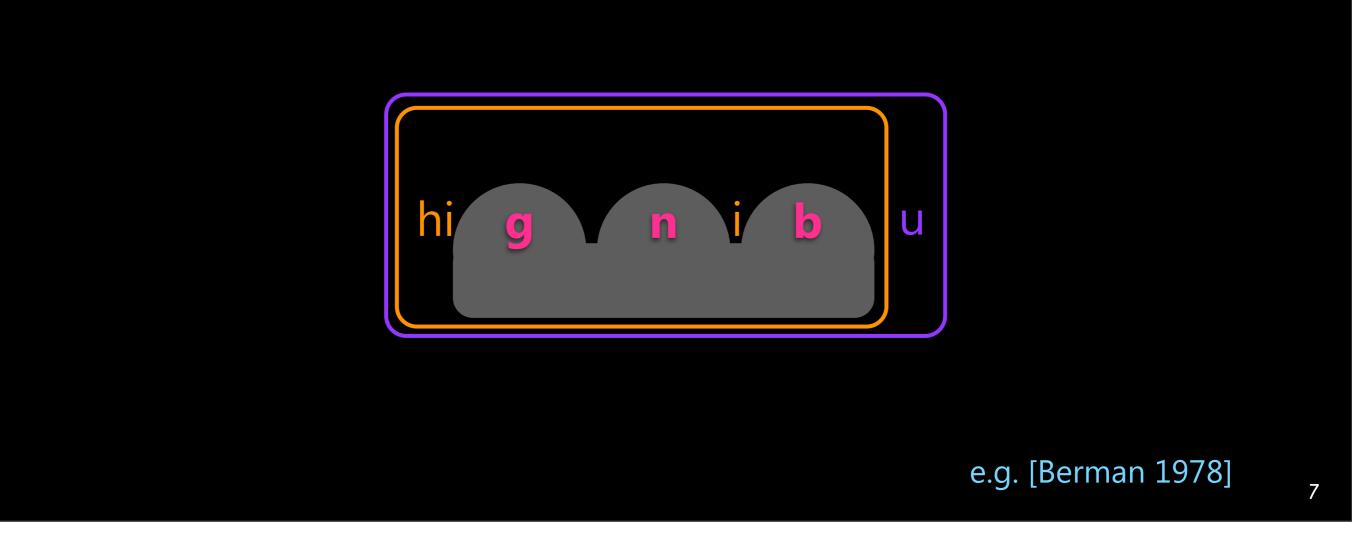












Root	/g/□/n/□/b,	/ ~ 'steal'	Paradigm	Hif'il (P5)	Meaning	'smuggle in'
Tense/Num	Past m	Past f	Present m	Present f	Future m	Future f
1.sg	hignav ti				? agniv	
2.sg	hignav ta	hignav t	magniv	magniv a	t agniv	tagnivi
3.sg	<u>higniv</u>	higniv a			y agniv	t agniv
1.pl	higna	av nu			n agniv	
2.pl	hignav tem	hignav ten	magniv im	magniv ot	ta	gniv u
3.pl	higniv u				y agniv u	

A single stem for each root/paradigm/tense: here past /hignib/*, present /magnib/, future /agnib/

The root fits into a pattern: /hi^oi^o/, /ma^oi^o/, /a^oi^o/ **Affixes specifying person, gender, and number—not sensitive to paradigm**: /-ti/, /-im/, /t- -u/, etc.

Root	/g/¤/n/¤/b/ ~ 'steal'		Paradigm	Hif'il (P5)	Meaning	'smuggle in'	
Tense/Num	Past m	Past f	Present m	Present f	Future m	Future f	
1.sg	hignav ti				?	? agniv	
2.sg	hignav ta	hignav t	magniv	magniv a	t agniv	t agniv i	
3.sg	<u>higniv</u>	higniv a			y agniv	t agniv	
1.pl	higna	a∨nu			n agniv		
2.pl	hignav tem	hignav ten	magniv im	magniv ot	ta	gniv u	
3.pl	higniv u					y agniv u	

A **single stem for each root/paradigm/tense**: here past /hignib/*, present /magnib/, future /agnib/

The root fits into a pattern: /hi^oi^o/, /ma^oi^o/, /a^oi^o/ **Affixes specifying person, gender, and number—not sensitive to paradigm**: /-ti/, /-im/, /t- -u/, etc.

Root	/g/¤/n/¤/b/ ~ 'steal'		Paradigm	Hif'il (P5)	Meaning	'smuggle in'	
Tense/Num	Past m	Past f	Present m	Present f	Future m	Future f	
1.sg	hignav ti				? agniv		
2.sg	hignav ta	hignav t	magniv	magniv a	t agniv	t agniv i	
3.sg	<u>higniv</u>	higniv a			y agniv	t agniv	
1.pl	hignav nu				n agniv		
2.pl	hignav tem	hignav ten	magniv im	magniv ot	ta	gniv u	
3.pl	higniv u					y agniv u	

For brevity, assume some phonological details are handled elsewhere:

- Consonant allophony: /b/ is sometimes realized as [v], /k/ as [x], and /p/ as [f]
- Certain root consonants (e.g. /?/, /w/, /h/) will affect the pattern in systematic ways
- Stress-sensitive vowel reduction and deletion
- * The last vowel in this paradigm's past tense stem undergoes the phonological change /i/→[a] in 1st & 2nd person

Morphological Generalizations: Stored or Inferred?

- I will present general morphological constructions as if they are stored in the lexicon along with all other constructions.
- However, some approaches to morphology claim that no constructions below the word level are stored in memory; rather, an online process of distributed analogy is hypothesized to account for morphological productivity. [Gurevich 2006]
 - For those taking this view, the generalizations presented here can be interpreted as formalizing an online analogical process.

Paradigms (Binyanim)

Ρ	Traditional	Transitivity:	/g/º/n/º/b/ Verbs ³		
	Characterization ¹	always (often) ²	Hebrew	Gloss	
1	"Simple"	(Transitive)	ganav	'steal'	
2	"Refl., passive"	Intrans. (Passive)	nignav	'be stolen'	
3	"Intensive"	(Transitive)	ginev	'steal repeatedly' (lit.)	
4	"Intensive Passive"	Passive	gunav	'be stolen/taken stealthily' (lit.)	
5	"Causative"	(Transitive)	higniv	'smuggle in, insert stealthily'	
6	"Causative Passive"	Passive	hugnav	'be smuggled in/inserted stealthily'	
7	"Reflexive-passive"	Intrans. (Passive)	hitganev	'sneak (in, out, or away)'	

1. [Halkin 1970] 2. [Arad 2005] 3. [Bolozky 1996]

The root /g/^o/n/^o/b/ is one which manifests itself in all seven paradigms, though its P3 and P4 verbs are limited to literary usage.

Paradigm Semantics

- There is a lot of idiosyncrasy in the meanings of verbs within the various binyanim. That is, the verb's meaning is often not completely predictable from the root and paradigm.
- Mandelblit [1997] attacks this problem under the rubric of grammatical blending [Fauconnier & Turner 1996]
 - She concludes that the different paradigms arise from a construed causal relationship, which explains the prototypical semantics

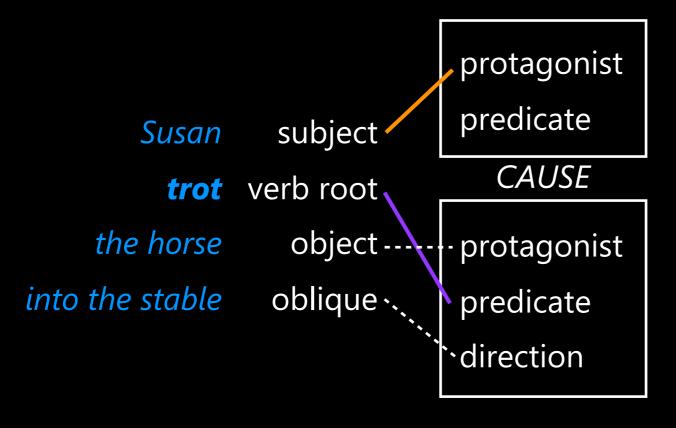
Paradigms: An English Analogy

protagonist

predicate

-direction

The horse trotted into the stable. *(basic)*



subject-

oblique -----

trot verb root-

horse

into the stable

Susan **trotted** the horse into the stable. *(causative)*

adapted from [Mandelblit 1997, p. 36]

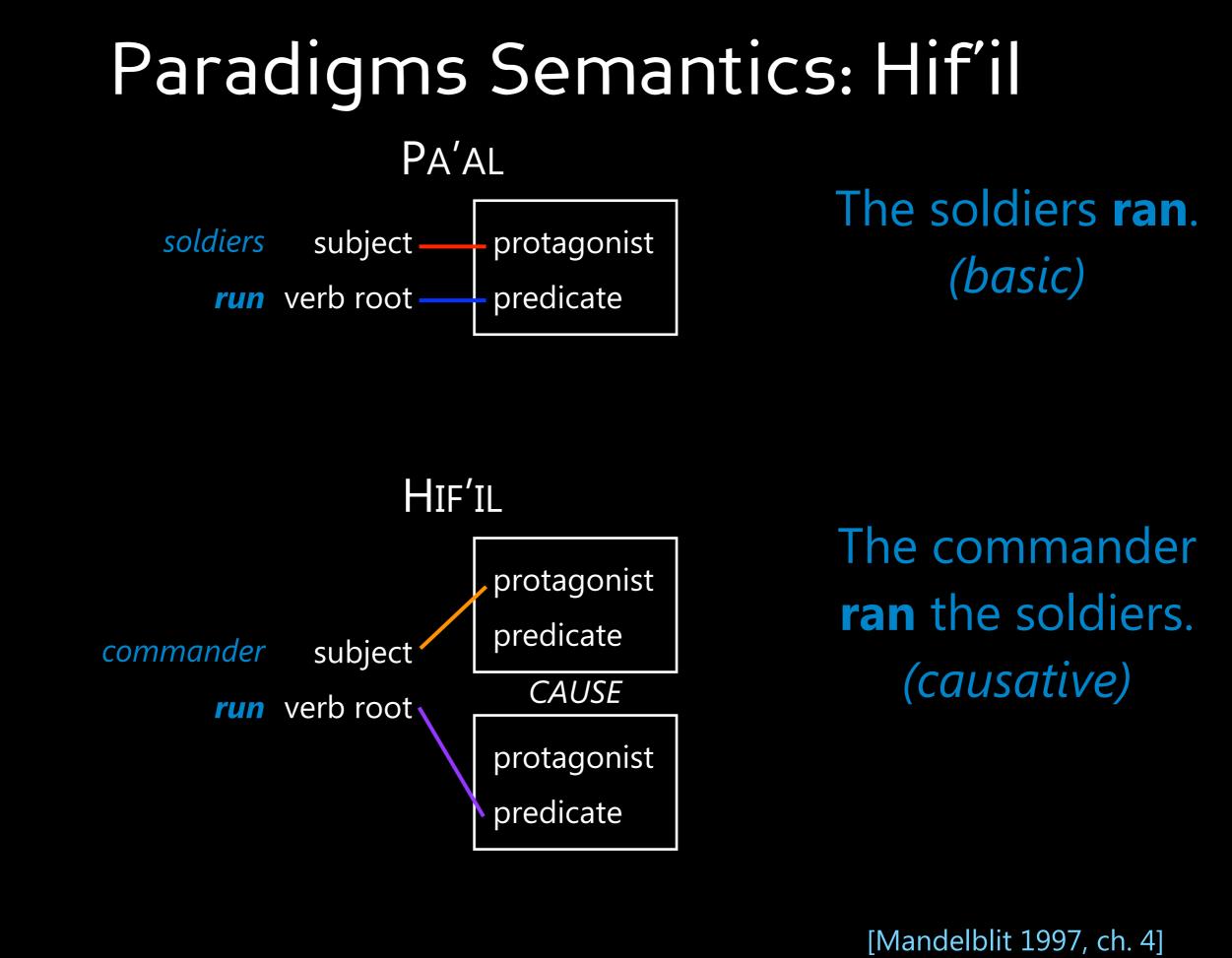
Paradigm Semantics

Mandelblit argues that the root contributes the "content" of the verb, and the paradigm picks out part of a causal sequence. For example: "The causative *hif'il* verbal pattern is used to mark a single sub-event (the *effected* subevent) within a conceived causal sequence of events. Marking other sub-events entails the usage of other *binyanim*."

Paradigm Semantics: Hif'il

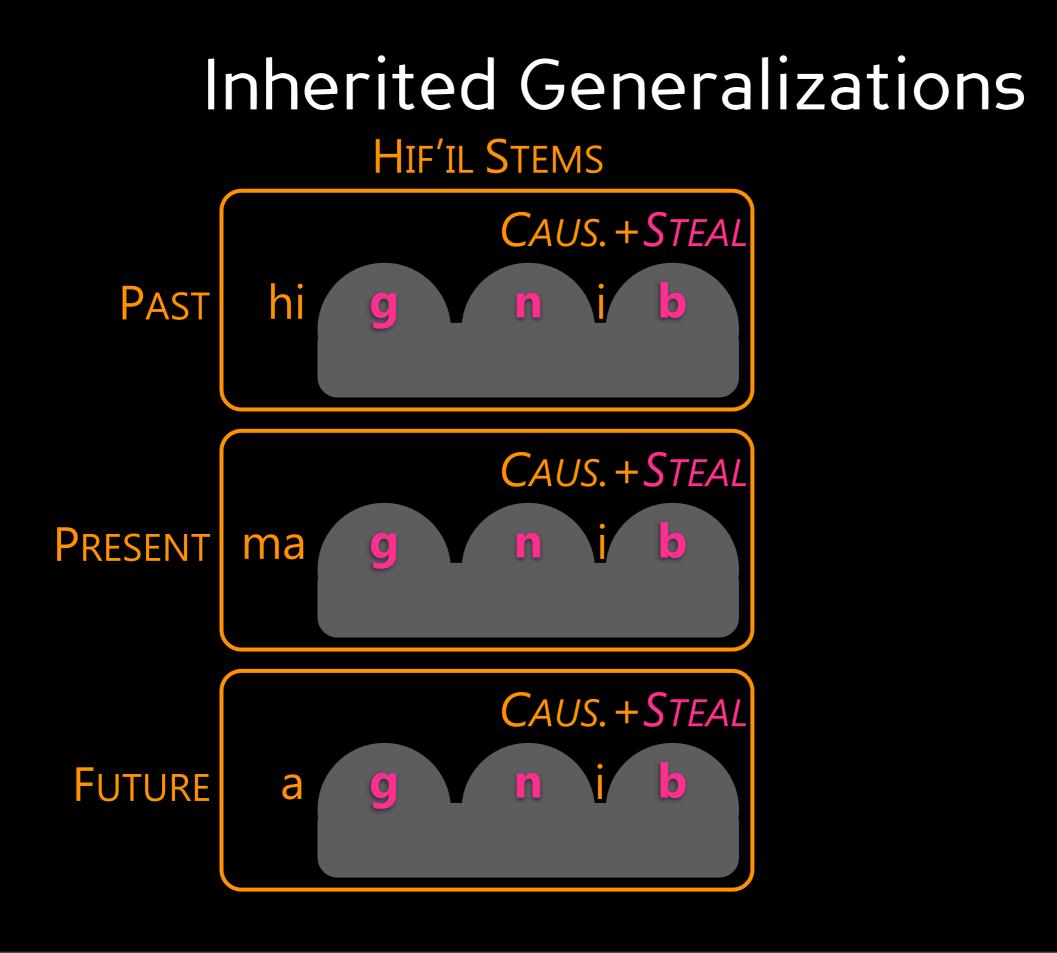
- (10) a. ha-xayal rats misaviv la-migraſ. the-soldier run.**PA'AL**.PAST.3.M.SG around to.the-courtyard 'The soldier ran around the courtyard.'
 - b. ha-məfaked herits ?et ha-xayal misaviv la-migra∫. the-commander run.нıғ'ı∟.раѕт.З.м.sg Acc the-soldier around to.the-courtyard 'The commander made the soldier run around the courtyard.'

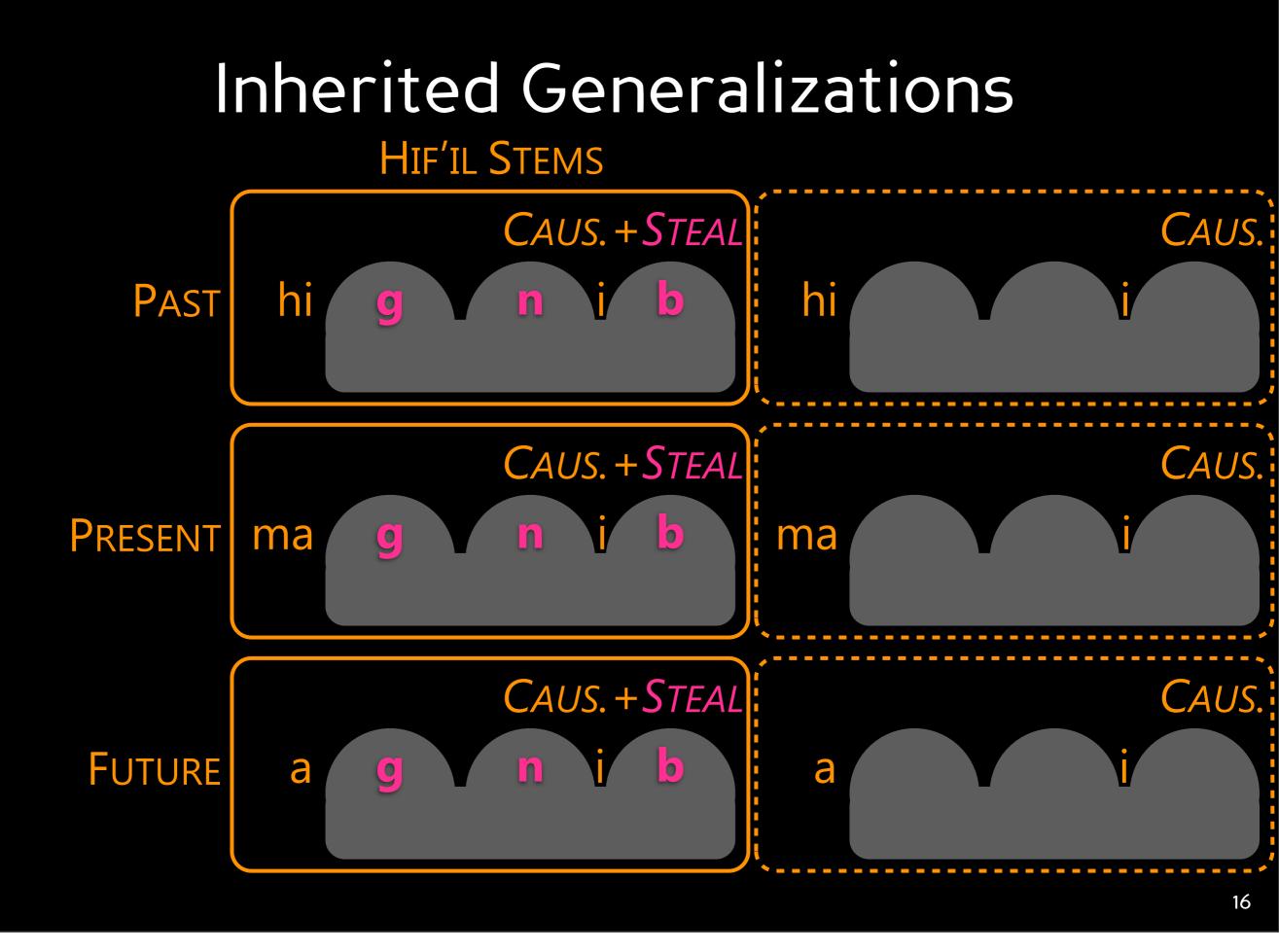
These verbs have the same root, /r/º/w/º/ts/ 'run'. In the causative hif'il sentence (b), *herits* 'cause to run' indicates that the root refers to the **effected event**—that is, what the soldiers are made to do. The causing event, i.e. *how* the commander makes them run, is unspecified.



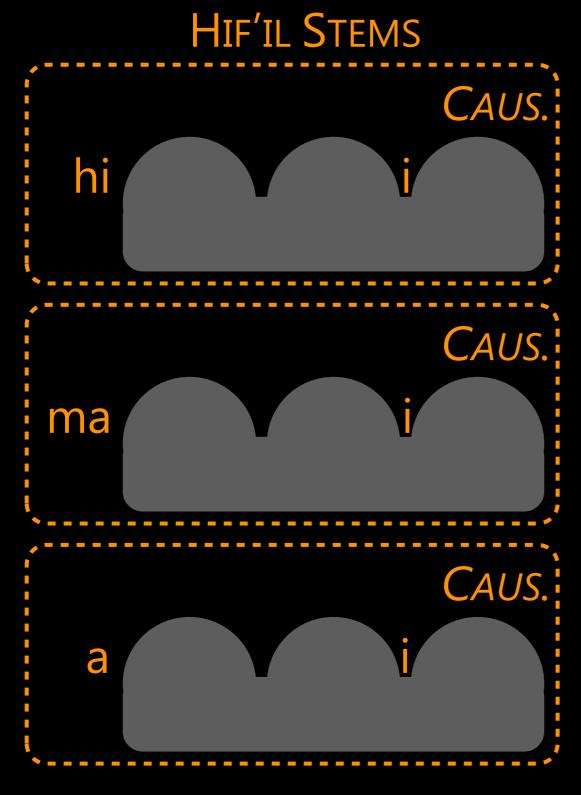
The space on the right (Input 1) illustrates the conceived causal relationship bound to specific participants and actions: *she* is understood to have taken some unspecified action—the *causing event*—which resulted in the horse trotting into the stable, trotting being the *effected event*. The space on the left (Input 2) shows how the Caused-Motion construction orders certain types of participants and predicates in an event

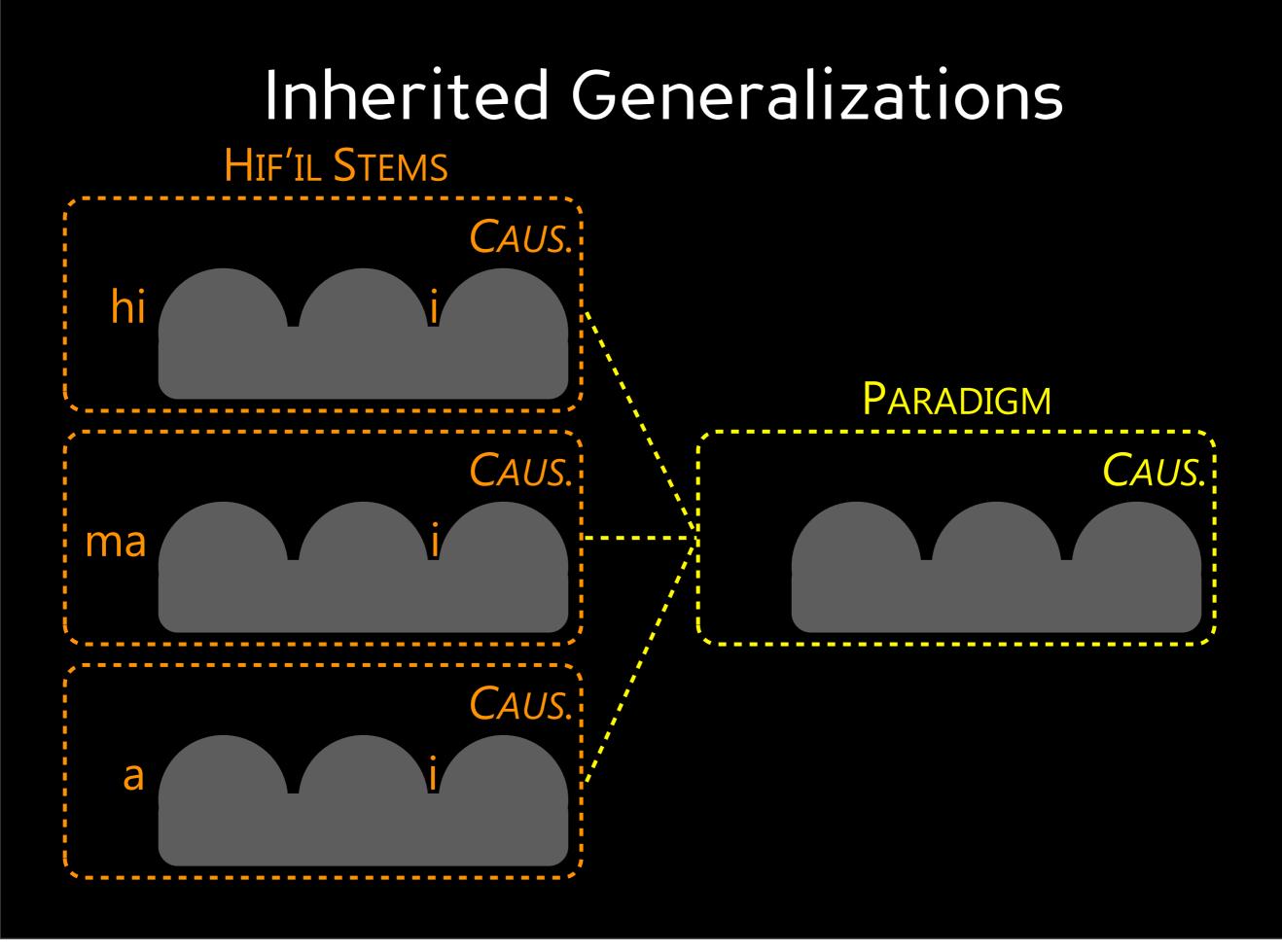
The space on the left (Input 2) shows how the Caused-Motion construction orders certain types of participants and predicates in an event sequence, associating them with syntactic categories. The Caused-Motion construction is said to be an *integrating* syntactic construction because it frames the sentence as a single event, even though the sentence has *unintegrated* semantics with two events in a causal relationship (depicted in Input 1). The blending operation results in the space at the bottom, with lexical items denoting some of the participants and predicates from Input 1 bound to syntactic positions from Input 2. **Those participants and predicates which are realized in the blend, with increased cognitive salience and overt representation in the sentence, are said to be** *profiled* **or** *highlighted***.**





Inherited Generalizations





Paradigm Semantics: Pi'el

- (10) a. ha-xayal rats misaviv la-migra∫.
 the-soldier run.pa'AL.PAST.3.M.SG around to.the-courtyard
 'The soldier ran around the courtyard.'
 - b. ha-məfaked herits ?et ha-xayal misaviv la-migra∫. the-commander run.нıғ'ı∟.раѕт.З.м.sg Acc the-soldier around to.the-courtyard 'The commander made the soldier run around the courtyard.'

Prototypical transitive pi'el verbs have a root denoting a causing event. Thus, they contrast with hif'il verbs much like *sneeze* in *Rachel sneezed the napkin off the table* contrasts with *trot* in *She trotted the horse into the stable*:

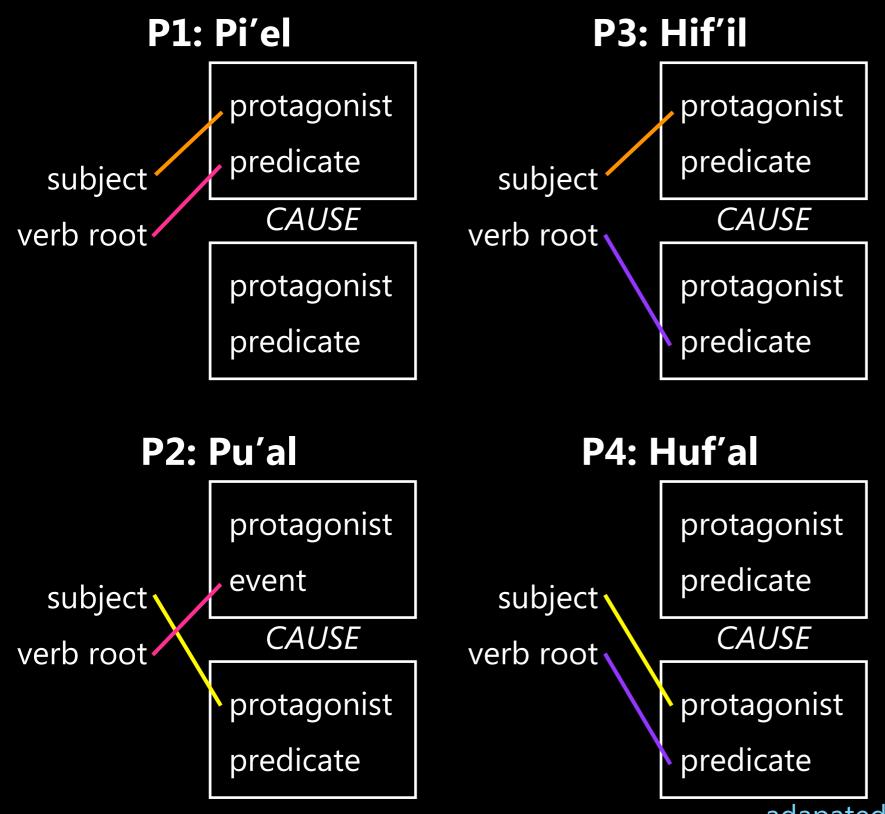
(14) ha-maʕasik piter ?et ha-ʕoved. the-employer fire.pɪ'el.past.3.м.sg acc the-worker 'The employer fired the worker.'

Paradigm Semantics: Huf'al and Pu'al

⁽¹⁰⁾ Huf'al is the passive counterpart of hif'il, and pu'al is the passive counterpart of pi'el:

- b. ha-məfaked herits ?et ha-xayal misaviv la-migra∫.
 the-commander run.нıғ'ı∟.раѕт.З.м.sg Acc the-soldier around to.the-courtyard
 'The commander made the soldier run around the courtyard.'
- (15) ha-xayal hurats (Sal yadei ha-mafaked).
 the-soldier run.нuf'AL.PAST.3.м.sg (on account.of the-commander)
 'The soldier was made to run (by the commander).'
- (14) ha-maĩasik piter ?et ha-ĩoved. the-employer fire.pi'el.past.3.m.sg acc the-worker 'The employer fired the worker.'
- (16) ha-Soved putar (Sal yadei ha-maSasik).
 the-worker fire.pu'AL.PAST.3.M.SG (on account.of the-employer)
 'The worker was fired (by the employer).'

Paradigms Semantics



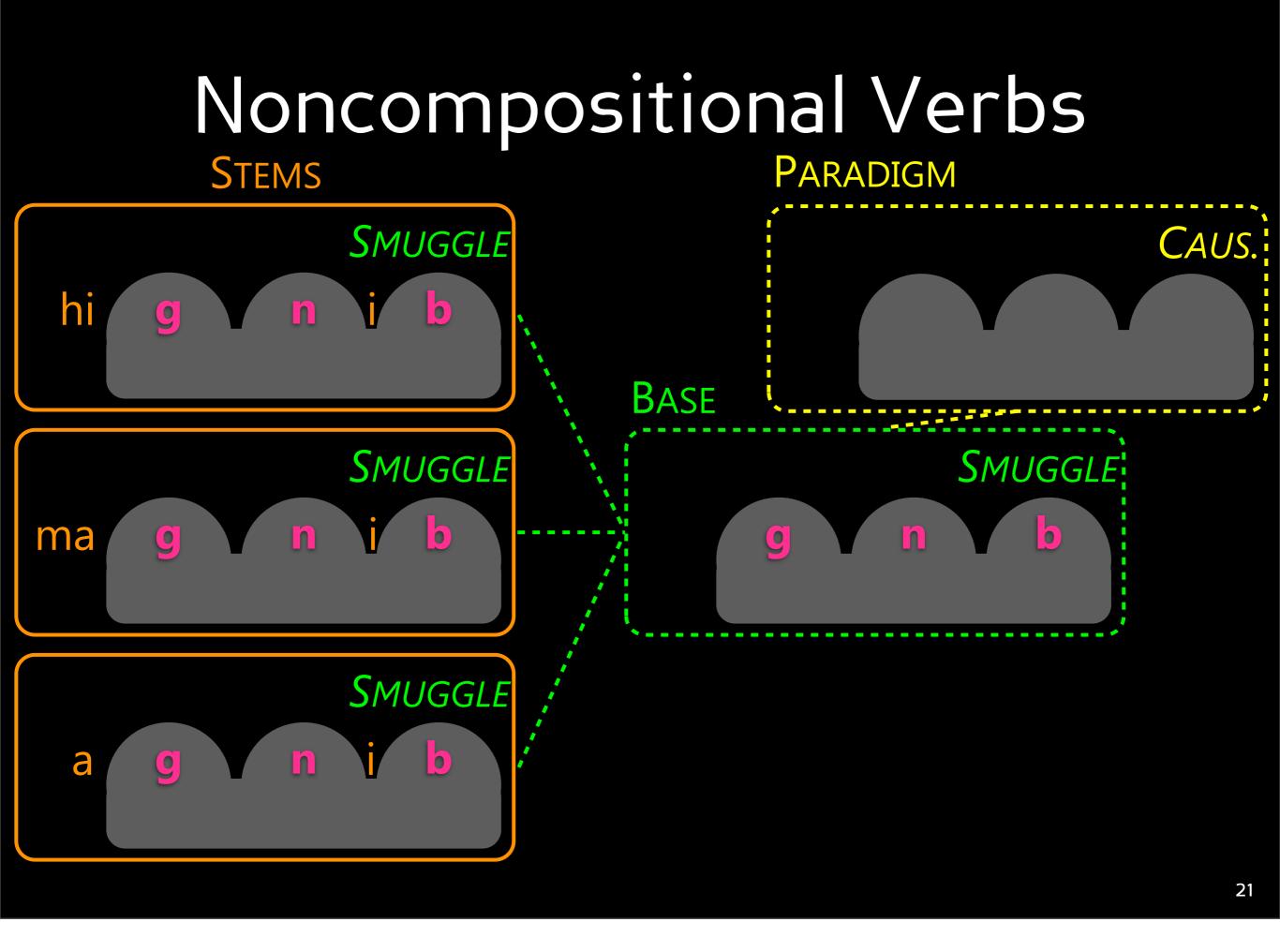
Blending schemas for paradigms P1 & P3 (active) and P2 & P4 (passive). The construed causal sequence on the RHS of each paradigm contains a causing event (top box) and an effected event (bottom box), each with agent and predicate. Paradigm constructions map one of the protagonists (agents) to the subject and one of the predicates to the verb root.

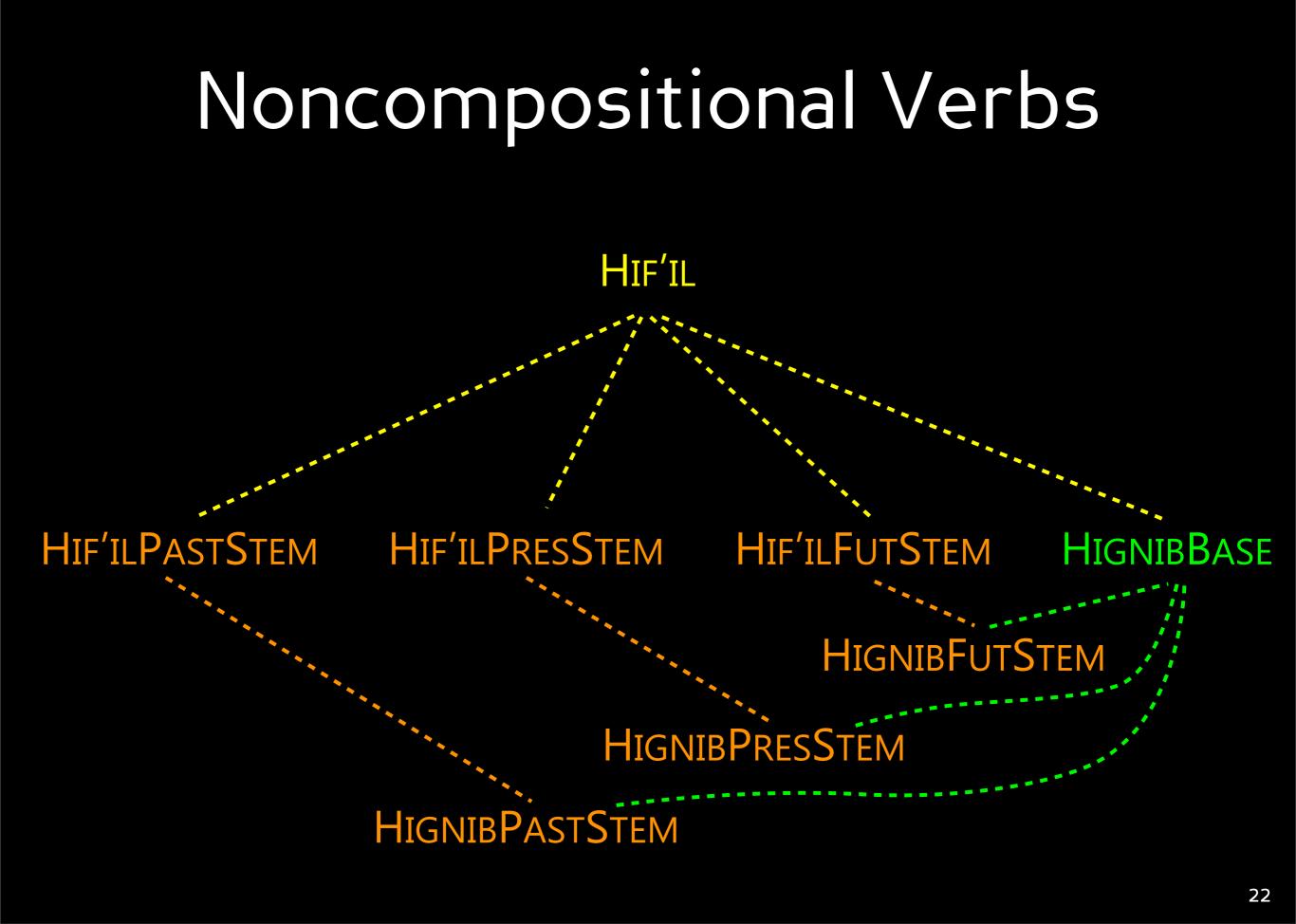
adapated from [Mandelblit 1997, p. 133]

The space on the right (Input 1) illustrates the conceived causal relationship bound to specific participants and actions: *she* is understood to have taken some unspecified action—the *causing event*—which resulted in the horse trotting into the stable, trotting being the *effected event*. The space on the left (Input 2) shows how the Caused-Motion construction orders certain types of participants and predicates in an event sequence, associating them with syntactic categories. The Caused-Motion construction is said to be an *integrating* syntactic construction because it frames the sentence as a single event, even though the sentence has *unintegrated* semantics with two events in a causal relationship (depicted in Input 1). The blending operation results in the space at the bottom, with lexical items denoting some of the participants and predicates from Input 1 bound to syntactic positions from Input 2. **Those participants and predicates which are realized in the blend, with increased cognitive salience and overt representation in the sentence, are said to be** *profiled* **or** *highlighted***.**

Noncompositional Verbs

- The story until now assumes the paradigm semantics is fully compositional given the root: the verb cxn may be deduced online
- To handle noncompositional verbs, we introduce a verb-specific base construction which pairs a particular root with a particular paradigm, and the associated semantics
 - Tense/other inflectional information does not alter the verb-specific meaning





This shows an inheritance hierarchy of constructions (à la HPSG): Hif'il is the most general and more specific cases inherit and elaborate upon its properties. HignibBase and its subtypes are idiosyncratic with respect to meaning, overriding the inherited prototypical causative meaning.

Embodied Construction Grammar

- Unification-based, semantically-rich formalism for describing lexical and syntactic—and now morphological—constructions [Bergen & Chang 2005] [Feldman 2006] [Feldman et al. 2009]
 - Part of the Neural Theory of Language project to develop computational simulations of language understanding
 - ECG grammars can represent embodied semantics: primitives include schemas/ frames, as well as metaphors and mental Spaces [Gilardi to appear]

Embodied Construction Grammar

- Facilitates (verifiably consistent) analyses of particular linguistic phenomena, e.g. motion-related constructions in English [Dodge 2010]
- Facilitates cognitive computational models of sentence processing [Bryant 2008] and language learning [Chang 2008] [Mok 2008]
- Tools include the probabilistic parser of [Bryant 2008] and a user-friendly interface for grammar engineering [Gilardi to appear]

ECG Analysis: Setup

- We want to formally specify a "construction" for Hebrew verbs
 - Small but very detailed decomposition of morphology and semantics
 - For our purposes, phonology is simplified to string concatenation
- Given this construction, a computer program can take an input word and list its possible analyses including semantic frames and their bindings

Schemas and Constructions

schema Causation

subcase of ComplexProcess
roles

causingProcess: Process effectedProcess: Process causalProtagonist: Entity affectedProtagonist: Entity

In ECG, meaning schemas are used to represent the frame semantics of a construction, and form schemas are used to decompose morphological forms. Schemas exist in an inheritance lattice and can define roles, which may be string-valued or may point to other schema instances.

Schemas and Constructions

schema Causation

subcase of ComplexProcess roles

causingProcess: Process effectedProcess: Process causalProtagonist: Entity affectedProtagonist: Entity

In ECG, meaning schemas are used to represent the frame semantics of a construction, and form schemas are used to decompose morphological forms. Schemas exist in an inheritance lattice and can define roles, which may be string-valued or may point to other schema instances. schema GNB subcase of Root roles r1 r2 r3 constraints $r1 \leftarrow "g"$ $r2 \leftarrow "n"$ $r3 \leftarrow "b"$

construction Root_GNB subcase of VerbRoot form: GNB meaning: Steal

Constructional Levels: Base

general construction Hif'il	
subcase of Paradigm	
constructional constituents	
root: Root	
form constraints	
root.r1 before root.r2 before root.r3	
meaning: Causation	
roles	
highlightedProtagonist: Entity	
highlightedProcess: Process	
constraints	
highlightedProcess ↔ root.m	
highlightedProcess ↔ effectedProcess	
highlightedProtagonist ↔ causalProtagonist	— hif'il-specific

Recall that paradigm hif'il highlights the *effected* process and the *causal* protagonist.

The Hif'il construction on the left specifies the compositional meaning and gives (underspecified) constraints on the form. HignibBase inherits from Hif'il for the root GNB, overriding the compositional meaning. The tense-specific stem constructions will inherit from HignibBase in turn.

Constructional Levels: Base

general construction Hif'il

subcase of Paradigm constructional constituents

root: Root

form constraints

root.r1 before root.r2 before root.r3

meaning: Causation

roles

highlightedProtagonist: Entity highlightedProcess: Process

constraints

highlightedProcess ↔ root.m

highlightedProcess ↔ effectedProcess

 $highlighted Protagonist \leftrightarrow causal Protagonist$

construction HignibBase subcase of Hif'il

constructional constituents

root: Root_GNB

meaning: Smuggle

an idiosyncratic meaning (overrides Causation)

- hif'il-specific

Recall that paradigm hif'il highlights the *effected* process and the *causal* protagonist.

The Hif'il construction on the left specifies the compositional meaning and gives (underspecified) constraints on the form. HignibBase inherits from Hif'il for the root GNB, overriding the compositional meaning. The tense-specific stem constructions will inherit from HignibBase in turn.

Conclusion

The aforementioned approach

- brings together the theoretical framework of Construction Grammar and studies of verbs in Modern Hebrew;
- integrates the form and meaning components of morphological structures in a single analysis; and
- employs and extends the Embodied Construction Grammar formalism so as to enable cognitive computational modeling of morphology.

Arad, M. (2005). Roots and patterns: Hebrew morpho-syntax. Dordrecht: Springer.

Bergen, B. K. (2003, March 18). Towards morphology and agreement in Embodied Construction Grammar. Manuscript. Retrieved October 18, 2007, from <u>http://www2.hawaii.edu/~bergen/papers/ECGmorph.pdf</u>.

Bergen, B. K., & Chang, N. (2005). Embodied Construction Grammar in simulation-based language understanding. In *Construction grammars: cognitive grounding and theoretical extensions* (pp. 147–190). John Benjamins.

Berman, R. A. (1978). *Modern Hebrew structure*. Tel-Aviv: University Publishing Projects.

Bolozky, S. (1996). 501 Hebrew verbs. Hauppauge, NY: Barron's Educational Series.

Booij, G. (2005). Compounding and derivation: evidence for Construction Morphology. In W. U. Dressler, F. Rainer, D. Kastovsky, & O. Pfeiffer (Eds.), *Morphology and its demarcations* (pp. 109–132). John Benjamins.

Booij, G. (2007). Construction Morphology and the lexicon. In F. Montermini, G. Boyé, & N. Harbout (Eds.), Selected proceedings of the 5th Décembrettes: Morphology in Toulouse (pp. 34–44). Somerville, MA: Cascadilla Press.

Bryant, J. (2008). Best-fit constructional analysis. Ph.D. dissertation, University of California, Berkeley.

Bybee, J. L. (2001). Phonology and language use. Cambridge University Press.

Bybee, J. L. (1985). *Morphology: a study of the relation between meaning and form*. Typological studies in language. Amsterdam: John Benjamins.

Chang, N. (2008). Constructing grammar: a computational model of the emergence of early constructions. Ph.D. dissertation, University of California, Berkeley.

Croft, W. (2001). Radical Construction Grammar: syntactic theory in typological perspective. Oxford University Press.

Dodge, E. (2010). Conceptual and constructional composition. Ph.D. dissertation, University of California, Berkeley.

Fauconnier, G., & Turner, M. (1996). Blending as a central process of grammar. In A. E. Goldberg (Ed.), Conceptual structure, discourse, and language (pp. 113-129). Stanford, CA: Center for the Study of Language and Information (CLSI), Cambridge University Press.

Feldman, J. A. (2006). From molecule to metaphor: a neural theory of language. MIT Press.

Feldman, J. A., Dodge, E., & Bryant, J. (2009). A neural theory of language and Embodied Construction Grammar. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis*. Oxford University Press.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: the case of 'let alone'. *Language*, 64(3), 501–538.

Finkel, R., & Stump, G. (2002). Generating Hebrew verb morphology by default inheritance hierarchies. In *Proceedings of the ACL-02 Workshop on Computational Approaches to Semitic Languages*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics.

- Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. University of Chicago Press.
- Goldberg, A. E. (2006). Constructions at work: the nature of generalization in language. Oxford University Press.
- Gurevich, O. (2006). *Constructional morphology: the Georgian version.* Ph.D. dissertation, University of California, Berkeley.
- Halkin, A. S. (1970). 201 Hebrew verbs. Hauppauge, NY: Barron's Educational Series.
- Inkelas, S. (2008, February). The morphology-phonology connection. Presented at the 34th meeting of the Berkeley Linguistics Society, Berkeley, CA.
- Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: the What's X doing Y? construction. *Language*, *75*(1), 1–33.
- Langacker, R. W. (1990). *Concept, image, and symbol: the cognitive basis of grammar*. Berlin: Mouton de Gruyter.
- Mandelblit, N. (1997). *Grammatical blending: creative and schematic aspects in sentence processing and translation*. Ph.D. dissertation, University of California, San Diego.

McCarthy, J. J. (1979). Formal problems in Semitic phonology and morphology. Ph.D. dissertation, MIT.

- Mok, E. (2008). *Contextual bootstrapping for grammar learning*. Ph.D. dissertation, University of California, Berkeley.
- Nathan, G. (2007). Phonology. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics* (pp. 611–631). Oxford University Press US.
- Orgun, C. O. (1996). *Sign-Based Morphology and Phonology with special attention to Optimality Theory* (Ph.D. dissertation). Berkeley, CA: University of California, Berkeley.
- Riehemann, S. Z. (1998). Type-based derivational morphology. *The Journal of Comparative Germanic Linguistics*, 2(1), 49–77.
- Roark, B., & Sproat, R. W. (2007). *Computational approaches to morphology and syntax*. Oxford surveys in syntax and morphology. Oxford: Oxford University Press.
- Rubba, J. (1993). *Discontinuous morphology in modern Aramaic* (Ph.D. dissertation). University of California, San Diego.
- Rubba, J. (2001). Introflection. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language typology and language universals: an international handbook* (Vol. 1, pp. 678–694). Berlin: Walter de Gruyter.
- Tomasello, M. (2003). Constructing a language: a usage-based theory of language acquisition. Cambridge, MA: Harvard University Press.

Acknowledgments

Thanks are in order...

To Jerry Feldman, Eve Sweetser, George Lakoff, John Bryant, and the rest of the Neural Theory of Language Group; Miriam Petruck; and Rutie Adler and her Hebrew Linguistics class; and Noah Smith, Lori Levin, and Scott Fahlman for their advice & feedback

And to all of you for listening!