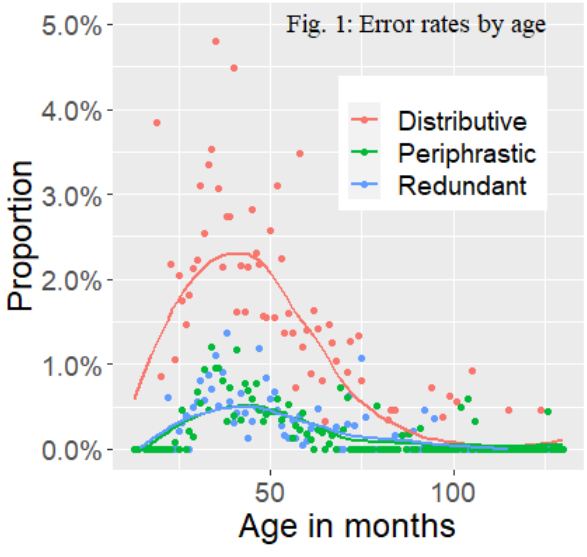
Errors of redundancy in child English and beyond

Johannes Hein, Imke Driemel, Fabienne Martin, Yining Nie & Artemis Alexiadou

Introduction. Children occasionally produce non adult-like forms during acquisition. Two common types of production are the redundant marking of a feature already spelled-out elsewhere and the overt marking of each semantic unit with a separate formative instead of 'compressing' semantic units into a single form (Guasti et al. 2022). These two types of 'undercompression errors' are respectively illustrated by (1b) and (1c) from English past tense formation, where they are known as overregularizations (Kuczaj 1977, 1978). Conducting a comprehensive corpus

(1) a. I **ate** my breakfast. b. I **ated** my breakfast. c. I **eated** my breakfast.

study on children's erroneous past tense productions in English, we found that *distributive* errors like *eat-ed* occur more often than *redundant* ones like *ate-d*. In *do*-support environments, erroneous double-marking of past tense (e.g. *did... ate*) is more frequent than erroneous *did... ated/eated*. The two types of errors illustrated in (1b/c) also occur in other domains, e.g. causatives and comparatives (Hein et al. 2022), possessives (Karmiloff-Smith 1981, Clark 1985), etc., which suggests a single unified treatment. **Claim.** We propose that both types of overregularization and their relative frequencies can be explained by children's well-established bias for a one-to-one mapping between form and meaning (Slobin 1985, Guasti et al. 2022), despite the apparent one-to-many relation in redundant errors. We implement the bias in Distributed Morphology as an occasional reduction of one-to-many relations between exponent and features by neglecting secondary feature specifications. Adopting Generalized Head Movement (Arregi and Pietraszko 2021, henceforth A&P) further allows us to also capture overtensing errors with an additional mistake, the omission of obliteration. The interaction of both mistakes captures the low frequency of *did... eated/ated* vs. *did... ate*. **Corpus study.** We targeted all UK and NA English-language corpora of typically developing children available through the ChiLDES database (MacWhinney 2000). Of the 44 irregular under the 100 most frequent verbs in the English ChiLDES corpora we excluded six from our investigation as their present and past forms are homographs (*cut*, etc.). We then ran a query for past tense forms of the 38 remaining verbs, including distributive and redundant error forms. Hits were automatically annotated for TAR(get), DIS(tributive), and RED(undant), excluding participles syncretic with past tense. Hits occurring as the complement of *did* were annotated as PER(iphrastic) errors. This resulted in 103,590 tokens total of which 100,674 (97.19%) were correct past tense forms. Among the remaining 2916 (2.81%) occurrences, we found 382 (0.37%) redundant errors, 1771 (1.71%) distributive errors, 354 (0.34%) periphrastic *do*-support (345 (0.33%) with *ate*, 9 (0.009%) with *ated/eated*) and 399 (0.39%) other errors not fitting the classification. The error rate over age is presented in Fig. 1.



Analysis. As past tense formation arguably involves some kind of head movement, our account is based on the most recent and most comprehensive account thereof, i.e. Generalized Head Movement (GenHM). GenHM distinguishes a head X's *morphological* features that are relevant for spellout from its *syntactic* ones. The former are bundled in the value (X_m) of a larger [M] feature. The result of an application of GenHM to a higher head X and the head Y of its complement is a hierarchically structured M-value [$_{X_m}$ X_m Y_m] which is shared between X and Y. Vocabulary Insertion targets the terminals of this complex M-value. Heads sharing a single M-value constitute a head chain. The shared

M-value is pronounced in the highest strong (*) position (a language-specific lexical property of certain heads), if any; otherwise, in the highest position. This is achieved by delinking (✗) the M-value from all other heads. Taking the Vocabulary Items in (2), Vocabulary Insertion applies to the result of GenHM between V and T in English. In a target derivation (3a), /ate/ wins over /eat/ for insertion into V_m due to its secondary [PST] feature. /-∅/ wins over /-ed/ due to its secondary feature list of roots (Embick 2003). We propose that a child may occasionally make the mistake to disregard secondary features. If that mistake occurs at V_m and T_m (3c), the child may insert /eat/ in V_m and /-ed/ in T_m, thus producing a distributive error. If it occurs only at T_m (3b), the child inserts /ate/ in V_m but /-ed/ in T_m, producing a redundant error. If the mistake occurs only on V_m, the child may insert /eat/ in V_m but /-∅/ in T_m, resulting in an omission error (not shown here). Negligence of secondary features essentially leaves the voca-

(2) a. /eat/ ⇔ [$\sqrt{\text{EAT}}$]
b. /ate/ ⇔ [$\sqrt{\text{EAT}}$] / __[PST]
c. /-ed/ ⇔ [PST]
d. /-∅/ ⇔ [PST] / __[{$\sqrt{\text{EAT}}$, $\sqrt{\text{BRING}}$, ...}]

(3) a. [$_{TP}$ T Adv [$_{VP}$ V* ...]] → V_m–T_m ⇓ ⇓ (1a) /ate/ /-∅/
b. [$_{TP}$ T Adv [$_{VP}$ V* ...]] → V_m–T_m ⇓ ⇓ (1b) /ate/ /-ed/
c. [$_{TP}$ T Adv [$_{VP}$ V* ...]] → V_m–T_m ⇓ ⇓ (1c) /eat/ /-ed/

bulary items specified for a single meaning component. When (2b) is stripped of the secondary [PST] feature only the primary feature [$\sqrt{\text{EAT}}$] remains. Negligence of secondary features thus turns a one-to-many mapping between form and features into a transparent one-to-one mapping which can be taken as a direct consequence of children's general bias towards one-to-one relations between form and meaning (Slobin 1985, Brighton et al. 2005, Guasti et al. 2022). Distributive errors are more frequent because if the drive towards one-to-one mapping triggers negligence of secondary features, it tends to trigger it wherever possible within a complex M-value. In *do*-support contexts such as polar questions, a head chain created by GenHM between V*, T and C is split at V* (A&P 2021: 261) such that there are two type-identical tokens of the shared M-value, one linked to V*, the other linked to C (4). As V_m is not linked to its V* head anymore, it will be realized as a form of *do* (in past contexts as the allomorph *did*), while T_m no longer linked to its T head will undergo obliteration (Arregi and Nevins 2012), resulting in /eat/ realizing V_m as the only compatible exponent (4a). We propose that a child may occasionally make the mistake of omitting the obliteration of T_m (4b). In such a case, T_m conditions the insertion of /ate/ into V_m and is itself realized by /-∅/. V_m, again, is exponed by the past allomorph *did*. The result is a redundant overtensing error *did... ate*. Adding a mistake of secondary feature negligence to this derivation gives rise to errors like *did... ated/eated*. Given the low probabilities of each mistake in isolation, however, the likelihood of combined occurrence is even lower, which explains the high frequency of *did... ate* vs. other periphrastic errors.

(4) a. [$_{CP}$ C [$_{TP}$ DP T] [$_{VP}$ V* ...]]] → V_m–T_m–C_m ⇓ ⇓ /did/ /-∅/; V_m–~~T_m~~–C_m ⇓ /eat/
b. [$_{CP}$ C [$_{TP}$ DP T] [$_{VP}$ V* ...]]] → V_m–T_m–C_m ⇓ ⇓ /did/ /-∅/; V_m–T_m–C_m ⇓ ⇓ /ate/ /-∅/

Outlook. The likelihood of a mistake can be tied to the stability of a VI's secondary feature representation which can be linked to the VI's input frequency. This predicts varying error rates across verbs, as observed in our data. We further show that this analysis readily extends to errors in causatives and comparatives, as these also arguably involve VIs with secondary features.

Sel. Refs.: Arregi & Pietraszko (2021). The Ups and Downs of Head Displacement. *LI 52*. • Kuczaj (1977). The acquisition of regular and irregular past tense forms. *J Verb Learn Verb Behav 16* • Slobin (1985). Crosslinguistic Evidence for the Language-Making Capacity. *In:* The Cross-linguistic Studies of Language Acquisition 2. Lawrence Erlbaum. • Embick (2003). Locality, Listedness, and Morphological Identity. *Studia Linguistica 57*.