Research Note

Young Children's Structure Production: A Revision of the Index of Productive Syntax

Evelyn P. Altenberg,[a] Jenny A. Roberts,[a] and Hollis S. Scarborough[b]

Purpose: The Index of Productive Syntax (IPSyn; Scarborough, 1990) is widely used to measure syntax production in young children. The goal of this article is to promote greater clarity and consistency in machine and hand scoring by presenting a revised version of the IPSyn (IPSyn-R) and comparing it with the original IPSyn (IPSyn-O).
Method: Longitudinal syntax production in 10 30- and 42-month-old typically developing children drawn from the Child Language Data Exchange System (MacWhinney, 2000) Weismer corpus was examined, using both the IPSyn-O and the IPSyn-R.

Results: The IPSyn-R provided nearly identical scores to the IPSyn-O with the exception of scores affected primarily by 1 modified noun phrase structure. Structures ranked as more advanced were produced less frequently. The results also reveal which of the IPSyn-R's 59 structures were most and least likely to be produced by this sample at these ages.
Conclusions: The qualitative and quantitative differences between the IPSyn-O and the IPSyn-R are relatively minor. The IPSyn-R can make it easier to score the IPSyn, both by clinicians and researchers, and facilitate the IPSyn's move to machine scoring of language samples.

Syntactic measurement is an important tool in understanding language development. Since the 1970s, measures of syntax derived from natural language samples have proven useful, despite some shortcomings, for describing young children's syntactic proficiency and growth (Altenberg & Roberts, 2016; Bernstein Ratner & MacWhinney, 2016; Hadley, Rispoli, & Hsu, 2016; Long & Channell, 2001). Among these, mean length of utterance (MLU; Brown, 1973) has predominated; other such measures include Developmental Sentence Scoring (Lee, 1974), Assigning Structural Stage (Miller, 1981), and Language Assessment, Remediation and Screening (Crystal, Fletcher, & Garman, 1976).

Subsequently, the Index of Productive Syntax (IPSyn; Scarborough, 1990) was created. It was designed to be a "summary scale of grammatical complexity that would be appropriate for the study of individual differences in language acquisition" (Scarborough, 1990, p. 1). It was presented primarily as a research tool, although with the hope "that other researchers and clinicians may find it helpful" (Scarborough, 1990, p. 13). Recent efforts have seen an increased focus on the value of the IPSyn for clinicians (e.g., Price et al., 2008), particularly in conjunction with computerized language sample analysis (LSA; Bernstein Ratner & MacWhinney, 2016; Long, 2001; Long & Channell, 2001).

Scarborough quantified the Assigning Structural Stage procedure (Miller, 1981), made some modifications to the content (by examining, e.g., the frequency of occurrence of each item and an item's sensitivity to age differences), and instituted a types-based scoring system that yielded numerical scores suitable for quantitative analyses. In brief, a language sample from a preschool child is reviewed for instances of the production of 56 listed syntactic structures (plus four "other" forms) within four subscales: noun phrases (NP), verb phrases (VP), questions/negations (Q/N), and sentence structures (SS). Items within subscales are developmentally ordered on the basis of the extant developmental literature and preliminary analyses of language samples for typically developing preschoolers; for example, the VP items range from V1 (any verb) to V16 (past tense copula). The child receives 0, 1, or 2 points per item, depending on whether the child produced zero, one, or at least two exemplars of that item. Hence, analysis of every utterance within

[a]Department of Speech-Language-Hearing Sciences, Hofstra University, Hempstead, NY
[b]Haskins Laboratories, New Haven, CT

Correspondence to Evelyn P. Altenberg: sphepa@hofstra.edu

a sample, a requirement of most prior measures, is unnecessary. An IPSyn score thus provides an overview snapshot of syntactic forms that are and are not yet in the child's sample and, given that children typically have wider competence than may be exhibited in a brief sample, serves as a measure of grammatical emergence rather than mastery. It is generally recommended that 50–100 utterances be used in conducting an LSA (Paul & Norbury, 2012).

The IPSyn has been judged to have "both face and content validity as a measure of syntactic growth" (Hewitt, Hammer, Yont, & Tomblin, 2005, p. 200), with Shulman and Capone (2010) citing it as one of the two most common measures of syntax development (MLU being the other). Paul and Norbury (2012, p. 312) indicate that it is a suitable tool for tracking progress for children receiving intervention. They note that a child's movement toward the normal range of ability on specific structures targeted by intervention would indicate the strength of the intervention program and provide information on "which syntactic goals have been met and which need additional intervention." Note, however, that the production or lack of production of two exemplars per item limits its utility in monitoring individual structures.

IPSyn scores have been found to distinguish typically developing children from various populations of children with language impairments, for example, children who are late talkers (e.g., Rescorla, Dahlsgaard, & Roberts, 2000), children with fragile X syndrome (e.g., Price et al., 2008), children with autism (e.g., Condouris, Meyer, & Tager-Flusberg, 2003), and children who "later developed reading disabilities" (Scarborough, 1990, p. 12).

Despite its advantages, users of the IPSyn still deal with a number of limitations. Concerns have been raised that "the operational definitions of many of its items do not accord well within formal linguistic frameworks" (Hadley & Short, 2005, p. 1357); that its sensitivity to age differences can be weak, especially outside the preschool period for which it was designed (Oetting, Newkirk, Hartfield, & Wynn, 2010); and that its Q/N subscale lacks utility when nonconversational (e.g., narrative) language samples are analyzed (Hewitt et al., 2005), as was acknowledged by its creator (Scarborough, 1990).

Shortcomings of a more practical sort motivated this study, however. Although satisfactory reliability has usually been achieved, coding accuracy requires a depth of syntactic knowledge that few untrained coders possess (Hassanali, Liu, Iglesias, Solorio, & Dollaghan, 2014). In our experience, even trained research assistants have been concerned with some limitations in the clarity, level of detail, and comprehensiveness of the coding manual, seeking answers to questions like, for instance, "Can there be something between the verb and noun phrase in N6 (two-word NP after verb)?" and "Can N9 (three-word NP) be credited if the noun phrase exemplar is longer than three words?" Our goal was to address these issues.

A well-known limitation of all measures derived from natural language samples is the time it takes to arrive at a score. Transcription is itself very time consuming, and for that reason, many clinicians do not routinely collect and analyze natural language samples (Pavelko, Owens, Ireland, & Hahs-Vaughn, 2016). Coding for MLU and IPSyn imposes even more time demands. For example, the time it took to score IPSyn averaged 30 min in Hassanali et al.'s (2014) study and 12–99 min (varying by scorer and sample complexity) in Long's (2001) study.

When applicable, automation of scoring can dramatically reduce processing time and make it less labor intensive. Not surprisingly, therefore, there has been considerable interest in automated transcription and, regarding our focus, in how well IPSyn scoring might be accomplished by computers (Bernstein Ratner & MacWhinney, 2016; Hassanali et al., 2014; Long, 2001; Long & Channell, 2001; Sagae, Lavie, & MacWhinney, 2005). However, MacWhinney (2014) raised concerns that resemble those of our hand scorers, noting that "some of the IPSyn rules are difficult to interpret unambiguously" and that such ambiguity may pose difficulties for machine scoring. Similarly, Altenberg and Roberts (2016) found, in the evaluation of one promising program for automated IPSyn scoring, the AC-IPsyn (Hassanali et al., 2014), that ambiguous interpretation of some items may have contributed to the reliability differences found between scores derived by hand scoring versus machine scoring. Whether scores derived by hand or machine are relied upon for clinical purposes, consistency in interpretation and scoring is imperative.

The push for automated LSA is undeniable, and some researchers have argued that "we have reached the point where LSA grammatical parsing and computation can and should be done using software" (Bernstein Ratner & MacWhinney, 2016, p. 83), as has already begun for IPSyn (Lubetich & Sagae, 2014). However, the reliability and validity of machine scoring programs for IPSyn, although promising, have not been firmly established; caution is probably merited in adopting them. Furthermore, even when such programs become established as valid and reliable and the IPSyn becomes more widely used, as is likely, it becomes even more essential that clinicians and researchers fully understand the IPSyn and what its scores are based on.

In addressing our goal of revising the IPSyn scoring criteria, therefore, we sought to make the revised IPSyn (IPSyn-R) guidelines clearer, less ambiguous, more detailed, and more complete than those for the original IPSyn (IPSyn-O), not only for the benefit of hand scorers but also for computer scoring purposes. In the remainder of this article, we first describe the changes to the coding manual and their rationales. We then present scores derived from transcripts of samples taken at ages 2;6 and 3;6 (years; months) for a small longitudinal sample and compare results obtained by applying the IPSyn-R criteria versus the original guidelines. The central question addressed here is: How does the IPSyn-R compare with the IPSyn-O, qualitatively and quantitatively? In addition, we explore how the IPSyn scores of our sample, in particular, the relationship between an item's rank and its likelihood of production, inform our understanding of the expected patterns of syntax production.

Method

IPSyn-R: Changes Made to the Original Coding Criteria

The IPSyn-R is provided in the Appendix. (See Scarborough, 1990, for the original.) All decisions about whether and how to change the scoring criteria were guided by our combined experiences in coding IPSyn and training scorers over many years and by our intent to preserve the nature of the original system while making the scoring guidelines more detailed and less ambiguous. Changes fell into one or more of three categories.

Wording Is Clarified

In many instances, item descriptions and coding directions were made clearer, but in ways that would leave the scoring of utterances unaffected. This includes items where restrictions were incorporated into the item's description. For example:

- Q8, originally defined as "yes/no question with inverted modal, copula, or auxiliary" has been expanded to read "yes/no question with inverted modal, copula, or auxiliary *BE, DO,* or *have.*"

- The exemplar must have a main verb to receive credit for V12, Q6, Q7, Q11, S8, or S18.

- The "phrasal" criterion has been replaced by a "structural" one when the form under consideration is more complex than a phrase. (This is the only revision to the general coding instructions.)

Tacit Guidelines Are Made Explicit

On the basis of our experience, most coders make some common, reasonable assumptions about scoring when the IPSyn-O directions are skimpy. For many items in IPSyn-R, these are now made explicit. For example:

- N9's description was changed from "three-word NP" to "three-word (or longer) NP."

- Q4 now specifies that "other words can intervene between the *wh-* word and the verb."

- Many of the specifications for second exemplar criteria have also been made more explicit, e.g., changing "lexical" to "lexical: different N."

Administration Is Streamlined

Three areas were addressed to maximize efficiency and clarity of administration by (a) eliminating or clarifying areas of potential confusion (including ensuring that structures are clearly distinct from one another), (b) maximizing internal consistency (primarily by the logical extension of crediting to some items), and (c) minimizing exclusions and second exemplar restrictions (see the "Notes" column of the Appendix). For example:

- For S10, "must begin a clause" was added to distinguish it from S5.

- N4 ("two-word [or longer] NP") now credits N1.

- V9's exclusion of "can't" or "won't" unless "*can/ will/could* modal is used" has been removed.

Language Samples

Language samples were drawn from the corpus donated to the Child Language Data Exchange System (CHILDES; MacWhinney, 2000) by Susan Ellis Weismer. Longitudinal samples were collected annually from the age of 2;6 with recordings made while the child was interacting with an examiner and "using a standard set of toys" (Weismer, p. 69). As described in detail by Moyle, Ellis Weismer, Lindstrom, and Evans (2007), all eligible participants had typically developing cognition and language, had normal hearing, and came from monolingual English-speaking homes, according to research assessments and parental reports.

For our analyses, we selected 20 samples, with 10 taken at age 2;6 and 10 at age 3;6 from the same children. These ages are of particular interest because they bracket a period that is thought to be a crucial time for identifying children with language impairments (Eisenberg & Guo, 2013). In addition, age differences between IPSyn-O scores have been found between these time points (e.g., Scarborough, 1990), and they are within the age range suggested by Oetting et al. (2010) for a close investigation of IPSyn structures.

We first identified children in the examiner–child corpus from whom transcripts were available at the age of 2;6 (30–32 months) and 3;6 (42–44 months) and that contained at least 100 child utterances. A small number of self-repetitions (3/100 utterances) was allowed in order to maximize the number of available transcripts. The average number of self-repetitions per transcript was one, for a total of 20 utterances with self-repetitions out of 2,000 utterances in our data set. Of the remaining five male and eight female participants, five female participants were selected randomly to balance gender. Mean MLU (and standard deviation) for the participants is 3.57 (0.64) at 30 months and 4.30 (0.75) at 42 months.

Procedure

CHAT (Codes for the Human Analysis of Transcripts) transcripts from CHILDES were imported into the Systematic Analysis of Language Transcripts program (Miller & Chapman, 2000). Utterances in each transcript were segmented as C-units (Moyle et al., 2007). The first two authors reviewed all transcripts and removed all incomplete and unintelligible utterances, utterances consisting only of filled pauses such as "oh, no" or sound effects (unless sound effects were judged to be used lexically, e.g., "He bammed the truck on the floor"), mazes (e.g., filled pauses, false starts, revisions; Miller, Andriacchi, & Nockerts, 2011), read text, and clearly memorized utterances from songs or other sources.

The first two authors each first scored 10 transcripts using the IPSyn-O (Scarborough, 1990) criteria and

independently scored 20% of the others so that reliability could be examined. The scorers differed on average by 2.5 points, and the mean percentage of point-to-point agreement was 92.8%. All the IPSyn-O exemplars selected were reviewed by both authors for adherence to the scoring criteria, with rare disagreements resolved by consensus; differences in interpretation of the guidelines accounted for most disagreements. The same transcripts were then scored using IPSyn-R criteria; again, all new exemplars were reviewed for adherence to the new criteria, with agreement reached by consensus.

Results

Scores at each age on the basis of IPSyn-O and IPSyn-R criteria are compared in Table 1. Subscale scores were examined in a Criterion × Age × Subscale repeated-measures analysis of variance with Greenhouse–Geisser correction. There were neither main effects of scoring criteria, $F(1, 9) = 2.727$, $p > .05$, $\eta_p^2 = .23$, or age, $F(1, 9) = 1.713$, $p > .05$, $\eta_p^2 = .16$, nor their interaction, $F(1, 9) = 0.000$, $p > .05$, $\eta_p^2 = 0$. There was also no interaction of subscale, criterion, and age, $F(1, 9) = 0.10$, $p > .05$, $\eta_p^2 = .01$. Because the subscales have different maximum scores, this main effect was meaningless and not tested, although interactions with age and criteria were of interest. A Subscale × Age effect reflected a large age difference, regardless of scoring criterion, for the SS subscale but not for the other subscales, $F(1.659, 19.80327) = 8.934$, $p = .004$, $\eta_p^2 = .498$. Scoring criterion also did not interact significantly with subscale, $F(1.788, 16.090) = 0.462$, $p > .05$, $\eta_p^2 = .049$.

Criterion Differences for Individual Items

A comparison of individual item scores provided by the two versions of IPSyn is provided in Table 2. For each structure at each age, the table lists the mean score for IPSyn-R and IPSyn-O, the absolute value of the differences between those means (maximum possible difference = 2), and the percentage of children who produced two exemplars (the maximum score) for the structure on the basis of IPSyn-R scoring criteria. Across all items, the mean absolute

difference between IPSyn-O and IPSyn-R per structure was 0.046 at 30 months and 0.039 at 42 months, out of a possible maximum of 2. Per transcript, the absolute difference in total scores under each criterion ranged from 0 to 3 points of a possible maximum score of 118 points. Because of the large number of comparisons in planned analyses of results for individual items, the false discovery rate (FDR) method, as described by Benjamini and Hochberg (1995) and Benjamini, Drai, Elmer, Kafkafi, and Golani (2001), was used to control for inflation of Type I error.

At the younger age, there were only 11 items for which mean scores were not equal for IPSyn-O and IPSyn-R, and for only three structures was the average difference greater than 0.1 point: N10 (adverb modifying verb or pronoun; $d = 1.44$, 1.7 vs. 1.0 points, respectively), N11 (any other bound morpheme on N or adjective; $d = 0.77$, 1.3 vs. 0.9 points), and V12 (regular past tense suffix; $d = 0.85$, 1.1 vs. 0.5 points). At the age of 42 months, there were eight items with different mean scores, of which only two involved differences between IPSyn-O and IPSyn-R scores of more than 0.2 point: N10 (adverb modifying verb or pronoun; $d = 2.12$, 1.6 vs. 0.6 points) and N11 (any other bound morpheme on N or adjective; $d = 0.70$, 0.6 vs. 0.1 point). Matched t tests with the FDR correction indicated that the differences were significant only for N10 at both ages.

Rank and Production

Scarborough (1990) numbered the structures within each subscale in developmental order, on the basis of prior studies of language development. To examine whether the ordering of structures in the IPSyn-R subscales corresponded to children's production of those structures in our language samples, we computed correlations between the items' ranks and two indices of performance—mean points earned and percentage of children who earned the full 2 points—within each subscale at each age.

As expected, strong inverse correlation coefficients were obtained, indicating that higher numbered items were indeed less likely to be produced. Using the FDR correction, all were significant at the .05 level. Correlations of item ranks with mean IPSyn-R points earned on the NP, VP,

Table 1. Comparison of IPSyn-R and IPSyn-O mean scores.

Subscale	Maximum points	M (SD) at the age of 2;6 years;months		M (SD) at the age of 3;6 years;months	
		IPSyn-R	IPSyn-O	IPSyn-R	IPSyn-O
NP	22[a]	19.0 (2.2)	20.1 (1.9)	18.3 (1.6)	19.8 (1.8)
VP	34	24.1 (2.5)	25.0 (2.3)	24.4 (2.8)	24.6 (2.7)
Q/N	22	13.4 (2.0)	13.6 (2.2)	13.3 (3.7)	13.7 (3.8)
SS	40	18.0 (3.8)	18.1 (3.9)	23.0 (3.9)	23.2 (4.0)
Total score	118[a]	74.5 (7.7)	76.8 (7.5)	79.0 (8.2)	81.3 (8.5)

Note. $n = 10$. IPSyn-R = revised Index of Productive Syntax; IPSyn-O = original Index of Productive Syntax; NP = noun phrases; VP = verb phrases; Q/N = questions and negations; SS = sentence structures.

[a]Maximum points were 24 for NP and 120 total in the original scoring criteria.

Table 2. Results by structure, per transcript.

Structure and abbreviated description		30-month-olds (*n* = 10)			42-month-olds (*n* = 10)			30-month-olds (*n* = 10)	42-month-olds (*n* = 10)
		Mean, IPSyn-O	Mean, IPSyn-R	Abs. value of mean difference (max. = 2)	Mean, IPSyn-O	Mean, IPSyn-R	Abs. value of mean difference (max. = 2)	% of children with maximum score, IPSyn-R	% of children with maximum score, IPSyn-R
N1	Noun	2	2	0	2	2	0	100	100
N2	Pronoun	2	2	0	2	2	0	100	100
N3	Modifier	2	2	0	2	2	0	100	100
N4	Two-word NP	2	2	0	2	2	0	100	100
N5	Article before noun	1.9	1.9	0	2	2	0	90	100
N6	Two-word NP after verb	2	2	0	2	2	0	100	100
N7	Plural suffix	1.4	1.4	0	1.9	1.9	0	70	90
N8	Two-word NP before verb	1.9	1.9	0	1.9	1.9	0	90	90
N9	Three-word NP	1.9	1.9	0	1.8	1.8	0	90	90
N10	NP adverb	1.7	1.0	0.7	1.6	0.6	1.0	20	0
N11	Other bound morpheme, N or adj.	1.3	0.9	0.4	0.6	0.1	0.5	30	0
V1	Verb	2	2	0	2	2	0	100	100
V2	Particle or preposition	2	2	0	2	2	0	100	100
V3	Prep. phrase	2	2	0	2	2	0	100	100
V4	Copula linking two Ns	1.8	1.8	0	1.9	1.9	0	90	90
V5	Catenative	2	2	0	2	2	0	100	100
V6	Auxiliary *BE, DO, HAVE*	2	2	0	1.8	1.8	0	100	90
V7	Progressive *–ing*	1.8	1.8	0	1.1	1.1	0	80	50
V8	Adverb	2	1.9	0.1	2	2	0	90	100
V9	Modal before V	1.9	1.9	0	1.9	1.9	0	90	90
V10	Third-person sing. pres.	1.5	1.5	0	1.7	1.7	0	80	80
V11	Past tense modal	0.7	0.7	0	1.4	1.4	0	30	60
V12	Regular past tense	1.1	0.5	0.6	0.5	0.4	0.1	20	0
V13	Past tense auxiliary	0.8	0.8	0	0.6	0.6	0	30	30
V14	"Medial" adverb	1.9	1.8	0.1	2	1.9	0.1	80	90
V15	Ellipsis (& emphasis, IPSyn-O)	1.3	1.2	0.1	0.8	0.8	0	50	20
V16	Past tense copula	0.1	0.1	0	0.7	0.7	0	0	30
V17	Other bound morpheme	0.1	0.1	0	0.2	0.2	0	0	10
Q1	Intonation	2	2	0	2	2	0	100	100
Q2	Routine, etc.	2	2	0	2	2	0	100	100
Q3N	Simple negation	2	2	0	1.5	1.5	0	100	60
Q4	*Wh*-question + verb	1.6	1.5	0.1	1.7	1.7	0	60	80
Q5N	Neg. between subject + verb	1.8	1.8	0	1.4	1.4	0	90	60
Q6	*Wh*-Q w/ inverted modal, copula, aux	1.3	1.1	0.2	1.7	1.6	0.1	40	70
Q7N	Negation of copula, modal, aux	1.8	1.9	0.1	1.5	1.4	0.1	90	60
Q8	Yes/no Q w/ inverted copula, modal, aux	0.8	0.8	0	1.1	1.1	0	30	60
Q9	*Why, when, which, whose*	0.2	0.2	0	0.2	0.2	0	0	0

(table continues)

Table 2. *(Continued).*

Structure and abbreviated description		30-month-olds (*n* = 10)			42-month-olds (*n* = 10)			30-month-olds (*n* = 10)	42-month-olds (*n* = 10)
		Mean, IPSyn-O	Mean, IPSyn-R	Abs. value of mean difference (max. = 2)	Mean, IPSyn-O	Mean, IPSyn-R	Abs. value of mean difference (max. = 2)	% of children with maximum score, IPSyn-R	% of children with maximum score, IPSyn-R
Q10	Tag question	0	0	0	0.5	0.3	0.2	0	10
Q11	Q w/ negation + inversion	0.1	0.1	0	0.1	0.1	0	0	0
S1	Two words	2	2	0	2	2	0	100	100
S2	Subject–verb	2	2	0	2	2	0	100	100
S3	Verb–object	2	2	0	2	2	0	100	100
S4	Subject–verb–object	2	2	0	2	2	0	100	100
S5	Conjunction (any)	1.6	1.7	0.1	1.9	1.9	0	80	90
S6	Any two Vs	2	2	0	2	2	0	100	100
S7	Conjoined phrases	0.6	0.6	0	1.2	1.2	0	10	50
S8	Infinitive	1.5	1.5	0	1.7	1.7	0	70	80
S9	*Let/Make/Help/Watch*	1.1	1.1	0	0.8	0.8	0	40	30
S10	Subordinating conj. (adverbial, IPSyn-O)	0.9	0.7	0.2	1.6	1.4	0.2	20	60
S11	Mental state V (propositional comp. IPSyn-O)	0.4	0.4	0	1.3	1.3	0	20	50
S12	Conjoined clauses	0.5	0.5	0	0.9	0.9	0	10	30
S13	*If* or *wh*-clause	0.3	0.3	0	1.2	1.2	0	10	50
S14	Bitransitive predicate	0.1	0.1	0	0.2	0.2	0	0	10
S15	Three or more (nonaux) Vs	0.5	0.5	0	1.5	1.5	0	10	60
S16	Relative clause	0.3	0.3	0	0.3	0.3	0	10	10
S17	Infinitive clause: new subject	0	0	0	0.2	0.2	0	0	0
S18	Gerund	0.2	0.2	0	0.3	0.3	0	0	0
S19	Fronted or center subord. clause	0	0	0	0.1	0.1	0	0	0
S20	Passive or tag (Other, IPSyn-O)	0.1	0.1	0	0	0	0	0	0
Mean				0.046			0.039		

Note. Portions of this table draw heavily upon material from within: Scarborough, H. (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11*(1),1–22. © Cambridge University Press 1990, published by Cambridge University Press, reproduced with permission. IPSyn-O = original Index of Productive Syntax; IPSyn-R = revised Index of Productive Syntax; max. = maximum; NP = noun phrase; N = noun; adj. = adjective; Prep. = prepositional; V = verb; sing. = singular; Neg. = negative; Q = question; w/ = with; aux = auxiliary; conj. = conjunction; comp = complement; subord. = subordinate.

Q/N, and SS subscales were −.75, −.80, −.89, and −.92, respectively, at the age of 30 months and −.73, −.76, −.92, and −.90, respectively, at the older age. Correlations of item ranks with mean IPSyn-O points earned on the NP, VP, Q/N, and SS subscales were −.68, −.78, −.91, and −.92, respectively, at the age of 30 months and −.69, −.75, −.90, and −.89, respectively, at the age of 42 months.

Similarly, when the mean percentage of children who produced two exemplars was correlated with the structure's rank within the subscale, strong effects were also seen. For IPSyn-R, correlations for the NP, VP, Q/N, and SS subscales were −.77, −.83, −.88, and −.89, respectively, at the age of 30 months and −.73, −.77, −.90, and −.92, respectively, at the age of 42 months. For IPSyn-O, the corresponding values were −.77, −.80, −.90, and −.89, respectively, at the age of 30 months and −.71, −.74, −.90, and −.91, respectively, at the age of 42 months.

Discussion

The qualitative and quantitative comparisons of the IPSyn-O and IPSyn-R indicate that the differences between them are relatively minor. Most of the qualitative differences, for example, changes to wording only and making assumptions explicit, were unlikely to affect IPSyn scores, and in fact, there was no significant difference between the two in total scores. The average absolute mean difference per individual structure under each criterion was 0.043 out of 2.00, a relatively small difference, further supporting the conclusion that scoring aligns closely under the two criteria. Nonetheless, there was a significant difference between the scores for N10 (adverb modifying adjective or pronoun) under each criterion. Note that, under the IPSyn-R, "right (t)here" was disallowed as an N10 exemplar; this resulted in lower scores for this structure. N10 was the only structure with a significant criterion difference.

As with the IPSyn-O, the correlations for each subscale of the IPSyn-R, in terms of both mean score and percentage of children with two exemplars, were all negative. That is, the higher the structure's subscale rank, the less likely it was to be produced. Although this is not surprising, given that Scarborough originally developed the IPSyn rank orderings on the basis of developmental research, it is worth noting that these orderings are supported by our data. The fact that this is the case for both the IPSyn-O and IPSyn-R speaks both to their similarity and to the likelihood that the rank orders are capturing meaningful production realities, providing support to the validity of the IPSyn. Overall, the results suggest that the differences between the IPSyn-R and the IPSyn-O do not affect the characteristics of the IPSyn-O that have made it such a valuable tool; rather, the revisions made enhance its usability.

The subscale and total scores for our data at the age of 30 months are higher than those found by Scarborough, suggesting that the Weismer participants were more advanced syntactically than those examined by Scarborough (1990) at the age of 30 months. This may be a factor in why the difference between the two age groups is not significant; that

is, the children may have been starting to plateau, suggesting that future research with a larger subject pool may be more likely to find an age difference. However, it is important to note that there was no difference between the two age groups regardless of whether the IPSyn-O or the IPSyn-R was used, further indicating that the IPSyn-R yields very much the same outcomes as the IPSyn-O.

Streamlining the IPSyn

The child development literature contains investigations, from various perspectives, of many of the structures incorporated within the IPSyn (e.g., Diessel & Tomasello, 2005, looking at relative clauses, and Theakston & Rowland, 2009, looking at auxiliary *BE*). However, although IPSyn total scores and subscale scores are widely used, there has been little examination, within the framework of the IPSyn, of the individual items that constitute it. Hadley (1998), Scarborough and Dobrich (1990), and Oetting et al. (2010) are exceptions to this; however, each of these studies reports on the production of only a subset of the IPSyn's structures for the particular populations and age groups they examine. Thus, Oetting et al. point to "the need for a large cross-sectional and multidialectal study of items on the measure using children who are between the ages of 2 and 4 years. This type of study would allow researchers to evaluate the diagnostic sensitivity and specificity of the items using children whose ages are ideally suited for this work" (p. 337).

As has always been the case, the presence or absence of two exemplars is not sufficient to measure an item's productivity for an individual child. Nonetheless, when the patterns of the subscales are examined, the picture that emerges is potentially useful, given the strong relationship found between subscale items' rank and their production. A recurrent theme in the literature has been the challenge of ensuring that language samples are accurate representations of a child's abilities; in particular, it has been pointed out that there may not be an opportunity, in a given sample, for a child to produce a specific structure (e.g., Balason & Dollaghan, 2002; Tomasello & Stahl, 2004). Syntax production may also be impacted by the size of a child's lexicon (e.g., Rowland & Fletcher, 2006) and by the nature of the discourse (Southwood & Russell, 2004). Thus, there are a number of factors that can affect the likelihood of a particular structure being produced; among these is the possibility that the structure has not yet been acquired. The study described here was not designed to distinguish between these factors. What it does tell researchers and clinicians is the relative likelihood of a specific structure being produced in a 100-utterance language sample by the typically developing children in our sample. Nonetheless, it is clear that some structures (e.g., pronoun, prepositional phrase, verb–object) were widely used at the age of 30 months by the children in our sample, as all produced two exemplars of them. Other structures (e.g., tag questions, infinitive clause with new subject) were either not yet emerging at the age of 42 months or not elicited in the language samples analyzed here, as none or only one of the 30- or 42-month-olds

produced even one exemplar of them. These results suggest that scoring for the appearance of some structures, for example, two-word NP and subject–verb, may not be particularly useful in discriminating between typical samples, as all children in this sample produced at least two instances of them. Scoring for other structures, for example, tag questions and fronted or center subordinate clauses, may also not be particularly useful in distinguishing among the transcripts at these ages, as only a very small percentage of children in either age group produced even one exemplar of these structures. It is the other structures, those primarily in the middle range of productivity, such as the progressive -*ing*, the infinitive, and mental state verbs, that have the most potential for distinguishing one sample from another. These middle-range structures may thus be most promising in terms of distinguishing typically developing from non–typically developing children.

Our examination of the individual structures indicated that subscale rank orders correlated negatively with item production. It also pointed to ways that the IPSyn might be streamlined in the future to focus on those structures that distinguish most effectively between individual transcripts. A more explicit description of the IPSyn, our primary goal here, provides an essential foundation for such future modifications.

Limitations, Clinical Implications, and Future Directions

The data here are limited by the number of participants and by the limitations inherent in all language sampling research. Furthermore, given the size of the sample used here as well as the fact that this was not a random sample, these data should not be considered to be normative. However, data such as that collected here can provide the seeds for a normative database using the IPSyn-R (see, e.g., Nippold, Vigeland, Frantz-Kaspar, & Ward-Lonergan, 2017). Eventual machine scoring of IPSyn samples would further facilitate the establishment of a normative database that could help clinicians better determine appropriate structures for assessment and intervention. The IPSyn is an excellent progress monitoring tool, as it provides both quantitative information that can be easily charted and information about areas targeted for intervention. Clinicians targeting VP development, for example, could monitor both whether the verb subscale is increasing over time and which structures within a subscale are emerging after intervention, given the strong relationship found between subscale items' rank and their production, while keeping in mind the limitation inherent in requiring only two productions of an item.

The field would benefit from research that investigates the IPSyn from the perspective of its individual structures with larger groups of participants, samples of different lengths, other age groups, different kinds of discourse, and children who speak nonmainstream dialects as well as children with communication disorders. For example, comparing child with adult frequencies of syntactic structures, as was done with *wh*-questions by Rowland and Fletcher

(2006), would be a good first step in teasing apart the influence of frequency on production scores.

The information on production contributes to our understanding of both the productive use and diagnostic utility of the IPSyn's structures and suggests that an abbreviated version of the IPSyn that focuses on those structures for which there is more variability at these ages may be a feasible research and clinical tool. Relevant research can be useful in further refining some categories. Note, for example, that mental state verbs (S11), one of the structures in this middle range, have been shown to be used less by late talkers than by typically developing children (Lee & Rescorla, 2008) and also show a relatively large age difference here. Guo, Owen Van Horne, and Tomblin's (2011) findings suggest that combining *BE, DO,* and *HAVE* auxiliaries may not capture production realities of children whose language is not typically developing. Klee and Gavin (2010) provide frequency counts of various Language Assessment, Remediation and Screening (LARSP) structures for 152 preschoolers, a valuable resource for such a project.

Future research is also needed to explicitly measure the reliability of the IPSyn-R as well as the accuracy of any programs designed for its machine scoring. (See Altenberg & Roberts, 2016, and Long & Channell, 2001, for examples of the latter.) The first two authors regularly gave graduate students in their courses an IPSyn assignment, in recent years with the IPSyn-R. They observed that classes that used the IPSyn-R were able to work more independently and required less support than those that had used the IPSyn-O. Furthermore, four professionals (two with PhDs in linguistics, one with a PhD in communication disorders, and one communication disorders doctoral candidate) with no experience using the IPSyn were asked to compare the clarity of the descriptions of IPSyn-O versus IPSyn-R using a survey in which items from each version of the IPSyn were randomized. Three rated IPSyn-R as overall clearer; one rated them as equally clear. Although this limited assessment did not address accuracy or reliability, it suggested greater clarity for the IPSyn-R descriptions.

In summary, the IPSyn-R, in conjunction with over 25 years of research conducted using the IPSyn as a tool, confirms and contributes to the IPSyn's value as a research and clinical instrument. Although the quantitative differences between the IPSyn-R and the IPSyn-O are minimal, its modifications provide practical advantages for its use. With the clear current impetus toward machine scoring programs, the lack of ambiguity in the interpretation of IPSyn guidelines becomes imperative. It is our hope that the IPSyn-R will make it easier for clinicians and researchers to reliably use the detailed syntactic information that the IPSyn provides.

Acknowledgments

References

Altenberg, E. P., & Roberts, J. A. (2016). Promises and pitfalls of machine scoring of the Index of Productive Syntax. *Clinical Linguistics & Phonetics, 30,* 433–446.

Balason, D. V., & Dollaghan, C. A. (2002). Grammatical production in 4-year-old children. *Journal of Speech, Language, and Hearing Research, 45,* 951–969.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., & Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behavioural Brain Research, 125,* 279–284.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289–300.

Bernstein Ratner, N., & MacWhinney, B. (2016). Your laptop to the rescue: Using the Child Language Data Exchange System Archive and CLAN utilities to improve child language sample analysis. *Seminars in Speech and Language, 37,* 74–84.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology, 12*(3), 349–358.

Crystal, D., Fletcher, P., & Garman, M. (1976). *The grammatical analysis of language disability.* London, UK: Arnold.

Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language, 81*(4), 882–906.

Eisenberg, S. L., & Guo, L.-Y. (2013). Differentiating children with and without language impairment based on grammaticality. *Language, Speech, and Hearing Services in Schools, 44,* 20–31.

Guo, L.-Y., Owen Van Horne, A. J., & Tomblin, J. B. (2011). The role of developmental levels in examining the effect of subject types on the production of auxiliary in young English-speaking children. *Journal of Speech, Language, and Hearing Research, 54,* 1658–1666.

Hadley, P. A. (1998). Early verb-related vulnerability among children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1384–1397.

Hadley, P. A., Rispoli, M., & Hsu, N. (2016). Toddlers' verb lexicon diversity and grammatical outcomes. *Language, Speech, and Hearing Services in Schools, 47,* 44–58.

Hadley, P. A., & Short, H. (2005). The onset of tense marking in children at risk for specific language impairment. *Journal of Speech, Language, and Hearing Research, 48,* 1344–1362.

Hassanali, K., Liu, Y., Iglesias, A., Solorio, T., & Dollaghan, C. (2014). Automatic generation of the Index of Productive Syntax for child language transcripts. *Behavior Research Methods, 46,* 254–262.

Hewitt, L. E., Hammer, C. S., Yont, K. M., & Tomblin, J. B. (2005). Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. *Journal of Communication Disorders, 38,* 197–213.

Klee, T., & Gavin, W. J. (2010). *LARSP reference data for 2- and 3-year-old children.* Christchurch, New Zealand: University of Canterbury Research Repository.

Lee, E. C., & Rescorla, L. (2008). The use of psychological state verbs by late talkers at ages 3, 4 and 5 years. *Applied Psycholinguistics, 29,* 21–39.

Lee, L. (1974). *Developmental sentence analysis.* Evanston, IL: Northwestern University Press.

Long, S. H. (2001). About time: A comparison of computerized and manual procedures for grammatical and phonological analysis. *Clinical Linguistics & Phonetics, 15,* 399–426.

Long, S. H., & Channell, R. W. (2001). Accuracy of four language analysis procedures performed automatically. *American Journal of Speech-Language Pathology, 10,* 180–188.

Lubetich, S., & Sagae, K. (2014). Data-driven measurement of child language development with simple syntactic templates. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers,* 2151–2160.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2014). *Re: IPSyn coding [Online forum comment].* Retrieved from https://groups.google.com/forum/#!msg/chibolts/uoH3ckm9oiQ/Lm3jjrqFjjYJ

Miller, J. F. (1981). *Assessing language production in children: Experimental procedures.* Baltimore, MD: University Park Press.

Miller, J. F., Andriacchi, K., & Nockerts, A. (2011). *Assessing language production using SALT software.* Middleton, WI: SALT Software LLC.

Miller, J. F., & Chapman, R. (2000). *Systematic Analysis of Language Transcripts (SALT).* Madison: University of Wisconsin, Language Analysis Lab.

Moyle, M. J., Ellis Weismer, S., Lindstrom, M., & Evans, J. (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research, 50,* 508–528.

Nippold, M. A., Vigeland, L. M., Frantz-Kaspar, M. W., & Ward-Lonergan, J. M. (2017). Language sampling with adolescents: Building a normative database with fables. *American Journal of Speech-Language Pathology, 26,* 908–920.

Oetting, J. B., Newkirk, B. L., Hartfield, L. R., & Wynn, C. G. (2010). Index of Productive Syntax for children who speak African American English. *Language, Speech, and Hearing Services in Schools, 41,* 328–339.

Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence* (4th ed.). St. Louis, MO: Mosby Elsevier.

Pavelko, S. L., Owens, R. E., Ireland, M., & Hahs-Vaughn, D. L. (2016). Use of language sample analysis by school-based SLPs: Results of a nationwide survey. *Language, Speech, and Hearing Services in Schools, 47*(3), 246–258.

Price, J. R., Roberts, J. E., Hennon, E. A., Berni, M. C., Anderson, K. L., & Sideris, J. (2008). Syntactic complexity during conversation of boys with Fragile X syndrome and Down syndrome. *Journal of Speech, Language, and Hearing Research, 51,* 3–15.

Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language, 27,* 643–664.

Rowland, C. F., & Fletcher, S. L. (2006). The effect of sampling on estimates of lexical specificity and error rates. *Journal of Child Language, 33,* 859–877.

Sagae, K., Lavie, A., & MacWhinney, B. (2005). Automatic measurement of syntactic development in child language.

In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics* (pp. 197–204). Stroudsburg, PA: Association for Computational Linguistics.

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11,* 1–22.

Scarborough, H. S., & Dobrich, W. (1990). Development of children with early language delay. *Journal of Speech and Hearing Research, 33,* 70–83.

Shulman, B. B., & Capone, N. C. (2010). *Language development: Foundations, processes, and clinical applications.* Sudbury, MA: Jones and Bartlett.

Southwood, F., & Russell, A.F. (2004). Comparison of conversation, freeplay, and story generation as methods of language sample elicitation. *Journal of Speech, Language, and Hearing Research, 35,* 343–353.

Theakston, A. L., & Rowland, C. F. (2009). The acquisition of auxiliary syntax: A longitudinal elicitation study. Part I: Auxiliary BE. *Journal of Speech, Language, and Hearing Research, 52,* 1449–1470.

Tomasello, M., & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language, 31,* 101–121.

Weismer, S. E. *Clinical corpora. CHILDES database* (pp. 69–71).

IPSyn-R

Structure	Description	Credit	Notes
N1	Proper noun, common noun (Lexical: different N)		
N2	Pronoun, functioning as an entire noun phrase [Cannot be functioning as modifier] (Lexical: different pronoun)		

N3	Modifier, including adjectives (including predicate adjectives),
	    possessives, quantifiers
	(Lexical: different modifier)
			NOT: Modifier in isolation (e.g., *this*) that child could be
			    using as pronoun

N4	Two-word (or longer) NP (Phrasal: one or both words of NP different)		
N5	Article, used before a noun Article need not be directly before the noun as long as both are part of the same noun phrase. (Lexical: different article. Or phrasal: different NP)		
N6	Two-word (or longer) NP (as in N4) after verb or preposition There can be intervening structures between the verb and noun phrase. (Phrasal: different NP)	N4	

N7	Plural suffix
	[Words that are never used in the singular, although they end
	    in -*s* and "look" plural, e.g., *pants*, are not credited]
	(Context: different N)
		N1	ONLY as 2nd exemplar: words that are usually pluralized,
			    e.g., *blocks*, *grapes*.

N8	Two-word (or longer) NP before verb There can be intervening structures between the NP and verb. (Phrasal: different NP)	N4	
N9	Three word (or longer) NP (Det/Mod + Mod + N) (Phrasal: different NP)	N4	NOT: *lots of toys*; conjoined nouns, e.g., *Mom and Dad*, *the boy and the girl*.
N10	Adverb modifying adjective or pronoun (Lexical: different adverb)	V8	NOT: *alldone* and *allgone*; *right (t)here*; *yes*, *yup*, *yeah* Although *not* is an adverb, credit Q3 instead.
N11	Any other bound morpheme on N or adjective (Lexical: different bound morpheme)		ONLY as 2nd exemplar: -*y* suffix (e.g., *nutty*, *sleepy*, *stinky*) NOT: compounds (e.g., *blackboard*, *seatbelt*).
N12	(N12 is eliminated)		
V1	Verb (Lexical: different V)		
V2	Particle or preposition (Lexical: different particle or preposition)		
V3	Prepositional phrase (Preposition + NP) (Phrasal: different PP)	V2	
V4	Copula linking two nominals or a nominal and a predicate adjective (Lexical: different form of copula. Or structural: same copula with different structure)	V1	ONLY as 2nd exemplar: Contracted *is* ('*s*): e.g., *He's silly*. NOT: *How are you?*
V5	Catenative (pseudo-auxiliary) preceding a verb (Lexical: different aux or main V)		
V6	Auxiliary *BE*, *DO*, *HAVE* Note that contraction of auxiliary is okay (Lexical: different aux or main V)	V5	NOT: "*don't* + V" unless *do/does/did* auxiliary used.

IPSyn-R

Structure	Description	Credit	Notes
V7	Progressive suffix (only when used as verb in sentence) Okay in isolation if judged to be a verb. [*-ing* words used as adjectives, e.g., *a swimming pool*, or *-ing* words used as nouns, e.g., *Swimming is fun*, are not credited here. For the latter, credit S18.] (Context: *-ing* added to different V)	V1	
V8	Adverb Can modify preposition and conjunction as well as verb. (Lexical: different adverb)		NOT: *alldone* and *allgone* (young children typically see them as one unit); *I think so*; *(t)here*; *yes, yup, yeah* Although *not* is an adverb, credit Q3 instead.
V9	Any one-word modal preceding a verb (Lexical: different modal. Or context: different V following modal)	V5	
V10	Third-person singular present tense suffix (*-s/-es* suffix on verb) [Words such as *does* and *says*, which look like they have the third-person *-s* but are irregular in their pronunciation, are excluded. Can be accepted only if child pronounces word as verb root plus *-s*, e.g., "*dooz*" (for *does*), "*saze*" (for *says*).] (Context: suffix added to different V)	V1	
V11	Past tense modal: *would, could, should, might* (Lexical: different past tense modal. Or context: different V following modal)	V9	
V12	Regular past tense suffix Suffix must be on main verb of a clause with no auxiliary verb. [Words with *-ed* suffixes are sometimes used as adjectives, e.g., *He's scared*; these are not credited as V12.] (Context: different V)	V1	
V13	Past tense of *BE*, *DO*, or *HAVE* auxiliary (Lexical: different aux. Or context: different V following aux)	V6	
V14	"Medial" adverb (adverb in middle of clause, typically before verb) (Lexical: different adverb)	V8	NOT: *alldone* and *allgone*; *yes, yup, yeah* Although *not* is an adverb, credit Q3 instead.
V15	Copula (C), modal (M), or auxiliary (A) used for ellipsis Note: Credit any/all structures in the credit column that are relevant (Lexical: different copula, modal, form of *DO* or *HAVE*. Or structural: different clause)	V4-C V6-A, V9-M[a]	
V16	Past tense copula: *was, were* (Lexical: different form of copula. Or structural: same copula in different clause)	V4	
V17	Any bound morpheme on verb or on adjective (to make adverb); must be a morpheme type that is not credited on any other IPSyn item. [Words like *hardly, really, repeat, butter, number*, etc., look like they have a familiar prefix or suffix (e.g., *-ly, re-, -er*, as in *quickly, rewrite, taller*) but the "suffixes" cannot be segmented out.] (Lexical: different root word or different bound morpheme)		
Q1	Intonationally marked question (Structural: different Q)		
Q2	Routine question with or without a verb, or *wh-* pronoun alone (Structural: different Q)		
Q3	Simple negation (neg + X): neg = *no(t), can't, don't*; X = NP, VP, PP, Adj, Adv, etc. [The *no* cannot be an answer to a yes/no question; it must be negating something.] (Structural: different simple negation)		

IPSyn-R

Structure	Description	Credit	Notes
Q4	Question with an initial *wh-* question word followed by verb; other words can intervene between the *wh-* question word and the verb. (Lexical: different *wh-* question word. Or structural)	Q1 Q2	If verb in 1st exemplar is *DO* or *GO*, 2nd exemplar can be neither *DO* nor *GO* NOT: *What's this? What's that?* Allow: *What is this? What is that?*
Q5	Negative morpheme (*n't*, *no*, *not*) between subject and verb Note that contraction of *not* (*n't*) is okay. (Lexical: different negative morpheme. Or phrasal: different VP)	Q3	ONLY as 2nd exemplar: *I dunno/nunno/don't know*
Q6	*Wh-* question with inverted modal, copula, or auxiliary *BE*, *DO*, or *HAVE*. Sentence must have main verb. Exclude *wh-* word in subject position because no opportunity for inversion, e.g., *What is happening?* (Phrasal: different VP)	Q4	If verb in 1st exemplar is *DO* or *GO*, 2nd exemplar can be neither *DO* nor *GO* NOT: *What's this/that? What is this/that? How are you?*
Q7	Negation of copula, modal, or auxiliary *BE*, *DO*, or *HAVE*. Sentence must have main verb. (Phrasal: different VP)	Q5	NOT: *I dunno/nunno/don't know.* If verb in 1st exemplar is *DO* or *GO*, 2nd exemplar can be neither *DO* nor *GO*
Q8	Yes/no question with inverted modal, copula, or auxiliary *BE*, *DO*, or *HAVE* (Structural: different relevant question)	Q1 Q2	
Q9	*Why*, *when*, *which*, *whose* used as a question word (not as a conjunction) (Lexical: different question word taken from list)	Q1	
Q10	Tag question, with tag containing verb and subject (Phrasal: different tag)	Q1 Q2	
Q11	Question with negation AND inverted copula/modal/auxiliary *BE*, *DO*, or *HAVE*. Sentence must have main verb. (Structural: different relevant Q)	Q6 Q7 Q8	NOT: Tag question with negative tag, e.g., *I need that, don't I?* Credit Q10 instead
S1	Two-word combination (Lexical: at least one different word)		
S2	Subject–verb sequence (Phrasal: different sequence)	S1	
S3	Verb–object sequence (Phrasal: different sequence)	S1	
S4	Subject–verb–object sequence [Predicate adjectives, e.g., *It is red*, are not objects] (Phrasal: different sequence)	S2 S3	
S5	Conjunction (Lexical: different conjunction)		
S6	Sentence with two verbs Verbs cannot be auxiliary verbs. (Phrasal: different VP)	S1	
S7	Phrases joined by a coordinating conjunction (Phrasal)	S1 S5	
S8	Infinitive: *to* + verb; there must also be a main verb (Lexical: different infinitive V)	S6 V5	NOT: Phonologically simplified forms, e.g., *gotta*, *gonna*, *hafta*, *wanna*, *oughta*, before infinitive verb
S9	*Let/Make/Help/Watch* introducer. There needs to be a second verb after the *let/make/help/watch* introducer. [Nonimperative forms: e.g., "*That makes me think about him*," do not credit S9.] (Structural)	S6	
S10	Subordinating conjunction. Must begin a clause. (Lexical: different subordinating conjunction)	S5	

Appendix (p. 4 of 4)

IPSyn-R

Structure	Description	Credit	Notes
S11	Mental state verb or verb of communication followed by a nominal clause acting as its object. The nominal clause has *that* as its subordinating conjunction (not a *wh-* conjunction); the *that* is optional. (Structural: different subordinate [nominal] clause)	S6	
S12	Conjoined clauses, each of which can stand alone. Conjunction must be present. Both clauses must have a subject and verb; however, if first or second clause is imperative, *you* may be understood. [If *if* can be replaced by *whether*, credit S13 rather than S12.] (Structural: different conjoined clauses)	S5 S6	
S13	*If* clause or nominal *wh-* clause [If *if* cannot be replaced by *whether*, credit S12 rather than S13.] (Structural: different *if* or *wh-* clause)	S6 S10	
S14	Bitransitive predicate (same thing as dative) The indirect object can be placed either before or after the direct object. (Structural)	S3	ONLY as 2nd exemplar: *Gimme that*
S15	Sentence with 3 or more verbs. May include infinitive but cannot include auxiliary verbs. (Structural)	S6	
S16	Relative clause, marked or unmarked (Structural: different relative clause)	S6	
S17	Infinitive clause; subject of infinitive clause must be different from subject of immediately preceding verb. (Structural: different infinitive clause)	S8	
S18	Gerund: verb + *-ing* used as a noun phrase. Sentence must also have main verb. (Lexical: different gerund. Or structural: different clause)		
S19	Fronted or center-embedded subordinate clause (Structural: different subordinate clause)	S6	
S20	Full or truncated passive construction; or tag comment/intrusion containing a clause [Do not credit a classic tag question here (i.e., one based on the structure of the main sentence, for example: *She was working, wasn't she?*); credit Q10 instead.] (Structural)	S11[a]	

Note. This table has heavily drawn upon material from within: "Index of Productive Syntax," by H. Scarborough, 1990, *Applied Psycholinguistics*, *11*(1), pp. 1–22 © Cambridge University Press 1990, published by Cambridge University Press, reproduced with permission. Criteria for 2nd exemplars are indicated in parentheses at the end of each item's description. Grammatical information designed to help the novice user is included in square brackets under "Description." The "Notes" column specifies exclusions and 1st exemplar restrictions. IPSyn-R = revised Index of Productive Syntax; N = noun; NP = noun phrase; Q = question; neg = negative; VP = verb phrase; adj = adjective; adv = adverb; V = verb; aux = auxiliary; PP = prepositional phrase; Det = determiner; Mod = modifier.

[a]Credit relevant exemplars only.