A Syntactic Taxonomy of Young Children's Utterances: A Descriptive Account

Bob Morris Jones

bmj@aber.ac.uk

August 2025

The aim of this work is to present a very general introductory taxonomy of the syntax of young children's utterances. To achieve this introductory presentation, only the barest of explanations are given about the syntactic patterns.

The study is based on a database of the Welsh of young children for whom Welsh is their first language. Details are available at https://users.aber.ac.uk/bmj/abercld/cronfa3_7/sae/intro.html.

The term utterance primarily identifies a string of speech sounds whose boundaries can be described in phonological terms which give an utterance an identifying prosodic contour. However, more relevant for the purposes of this work is that an utterance can also be given a description in terms of its syntactic constituency. The term utterance thus provides a general descriptive label for several syntactic patterns accounted for below, although in places the terms clause and sentence are also used.

1 Finite clauses and reduction

Taking a very general view, the taxonomy of the utterances of young children is made up of four possibilities:

  i.   Finite sentences, that is, any utterance which contains a finite verb
 ii.   Subject and predicate utterances
iii.   Predicate phrase only utterances
 iv.   Fragments

The possibilities in (ii) to (iv) are derived from finite sentences by reduction: the absence of a finite verb gives (ii), the absence of a finite verb and a subject gives (iii). In (iv) we have utterances which contain one or more phrases which are not clearly identifiable as forming (i), (ii) or (iii). Such utterances will be labelled as fragments. To help achieve a general picture, a predicate phrase is taken to include not only a verb and any complements but also aspect markers.

There is far more to the syntax of the utterances of young children than the constituency outlined above. For example, utterances can contain adjuncts, answer words, tags, parenthetical phrases, vocatives, negative words, preverbal particles and co-ordinating conjunctions. Further, utterances can have different words orders. But to achieve an overall and general description of young children's utterances, this account will focus exclusively on the constituencies given above but all additional constituents and word orders are retained in the examples.

2 The Syntactic Taxonomy: finiteness and reduction

2.1 Finite sentences

Finite clauses are identifiable by the presence of a finite verb. The latter can convey tense, factuality (factual or counterfactual) and agreement features (person and number). The finite verb can be an inflected lexical verb, an auxiliary verb or the copula.

In identifying a finite clause in these terms we are not concerned with the details of its major constituents. Examples are as follows.

2   a. gei   di   hwn.
        have.FUT.2SG   you.SG   this.M
        'you can have this.'

    b. fedra'   i godi   hon.
        can.PRES.1SG   I lift   this.F
        'I can lift this.'

    c. dyn   ni   wedi colli 'r   bws!
        be.PRES.1PL   we   PERF lose   the   bus
        'we have missed the bus!'

    d. ma'   'wnna   'n2   drwm.
        be.PRES.3SG   that.M   PRED   heavy
        'that is heavy.'

    e. chwaden 'dy   'wn?
        duck   be.PRES.3SG   this.M
        'is this a duck?'

    f. 's   'da   ni   ddim ieir.
        be.PRES.3SG   with   we   NEG   chickens
        'we haven't got any chickens.'

Other syntactic types of utterances are created through reduction: that is, by omitting one or more of the major constituents of a finite clause.

2.2 Reduction: subject-predicate utterances

The omission of the finite verb in some instances can produce utterances which have the linear sequence [Subject + Predicate]. In Welsh, 'clipping' of the finite element is made possible by the fact that the finite form occurs in initial position in normal-order utterances. Examples are given in (3).

3 a. gin   i un1 tewach.
       with   I one fatter
       'I ('ve got) one fatter

   b. lot lot o   tywod   'da   ti.
       lot lot of sand   with   you.SG
       'you ('ve got) a lot of sand.'

   c. sand   fanna.
       sand   there
       'sand there.'

   d. Mami   'n   neud e.
       Mammy   PROG   do   it
       'Mammy doing it.'

   e. nos   wedi dod   nawr.
       night PERF come   now
       'night (has) come now.'

   f. cowboi arall isie   iste fanna.
       cowboyother   want   sit   there
       'another cowboy want(s) sit there.'

   g. 'wn   yn2   drwm.
       this.M   PRED   heavy
       'this (is) heavy.'

2.3 Reduction: predicate phrases

The omission of the finite verb and the subject produces utterances which are made up of a predicate phrase alone. The predicate phrase is made up of lexemes which can head a predicate phrase in both finite sentences and subject-predicate phrases. Examples are given in (4).

4 a. rhaid   i'   fi gal   brwsh yn1 fanna.
      necessity for I   have brush   in   there
      'I must have a brush there.'

   b. a   cnoi   trwyn fi.
      and bite   nose   I
      'and bite my nose.'

   c. a   'di   bod yn1 Sir+fo:n.
      and PERF   be   in   Anglesey
      'and been in Anglesey.'

   d. isie   tebot.
      want   teapot
      'want a teapot.'

   e. yn2   brysur iawn.
      PRED busy   very
      'very busy.'

   f. gynno ni   hefyd.
      with   we   also
      'we've got also.'

   g. dim i'   fod yn2   llawn.
      NEG to be   PRED full
      'not to be full.'

2.4 Reduction: fragments

The taxonomic category of fragments is negatively defined. Fragments contain a variety of phrases which cannot be given the syntactic structure of one of the other general syntactic patterns. Examples are given in (5).

5 a. fan+hyn.
         'here'
   b. iawn.
         'right.'
   c. diar, diar.
         'dear, dear.'

d. nage fel 'a!

   no   like that

   'not like that.'

e. beth?

   'what?'

f. tan dolig.

   'until Christmas,'

g. 'ben y   mynydd   fanna?

   head the mountain there

   'on top of the mountain there.'

h. dim hwnna.

   NEG that.M

   'not that.'

i. neb   i' siarad.

   no-one to speak

j. hwn yfe?

   this.M Q

   'this?'

k. un i' ceir gal   mynd 'n+o:l a   'mla'n.

   one for cars have go   back   and forward

   'one for cars to go back and to.'

l. Plwmsan a Wynff a Glyn a   'i   daid   a   bob   dim1.

   Plwmsan a Wynff a Glyn and CL.3SG.M grandfather and every nothing

   'Plwmsn and Wynf and Glyn and his grandfather and everything.'

The fragment can be may be equivalent to a subject, object or adjunct in a finite clause. But these details are not considered in this study.

2.5 Predicative demonstratives

In Welsh there are utterances which cannot be accounted for by the above framework. Predicative demonstratives are utterances which are headed by the lexemes *dyma* 'here/this is', *dyna* 'there/that is' and *dacw* 'yonder is'. In speech, they are respectively shortened to *'ma2, 'na2* and, in the case of southern dialects, *'co*. Illustrations of predicative demonstrative utterances are given in the following examples.

6 a. 'na2   beth fi'   'n   trial neud.
    that+is what be.PRES.1SG PROG try do
    'that's what I'm trying to do.'

  b. 'na2   neis.
    that+is nice
    'that's nice.'

  c. co   hwn!
    yonder+is this.M
    'look at this.'

  d. 'na2   'i   lyged e.
    that+is CL.3SG.M eye he
    'that's his eye.'

  e. 'ma2   fe man+'yn.
    here+is he here
    'here he is here.'

  f. dyna   nw   'n   mynd.
    that+is they PROG go
    'there they go.'

Predicative demonstratives are treated as a separate syntactic type of utterance.

2.6 *Piau* clauses

Piau clauses convey ownership and can contain a finite constituent, namely, the copula. Examples are as follows.

7 a. fi sy   bia hwnna.
    I   be.PRES.3SG   own   that.M
    'that's mine.'

  b. fi o'dd   bia   fo.
    I   be.IMPF.3SG own   it
    'it was mine.'

  c. y   fuwch yma sy   bia &d [//] y lloi   yma i+gyd.
    the cow   here be.PRES.3SG own &d   the calves here all
    'all these calves are the cow's.'

But in the case of the present tense, the copula can be omitted.

8 a. Mam bia 'r Fiat.
   Mother own the Fiat
   'the Fiat is Mum's.'

  b. chdi bia hwnna.
   you.SG own that.M
   'that is yours.'

  c. cowboi bia hwn.
   cowboy own this.M
   'this is a cowboy's.'

Predicative demonstratives do not following the canonical Welsh order of finite constituent + subject + predicate. For the examples given above we have [NP + Finite element + *piau* + XP] or [NP + *piau* + XP]. For the purposes of simplicity, those examples of a *piau* clause which contains the copula are considered to be finite sentences and those which lack the copula are considered to be subject-predicate utterances.

2.7 Complex sentences

Complex sentences contain more than one clause and, for the purposes of this study, their taxonomic category is based on the main clause, which can be a possibility listed in section 1.

1 a. mi1 'ydodd Mrs+williams wbath bod ni 'n gorod dod. [Finite]
   PT say.PERF.3SG Mrs+Williams something be we PROG have+to come
   'Mrs Williams said something that we have to come.'

  b. <a ci> [>] fi 'n gweld o' fi 'n mynd. [Subject-pred.]
   and dog I PROG see be.IMPF.1SG I PROG go
   'and my dog seeing I was going.'

  c. jocan o'dd toy gyda ni. [Predicate]
   joke be.IMPF.3SG toy with we
   'pretend we had a toy.'

2 a. w i 'n aros yma tan dw i 'n mynd adre. [Finite]
   be.PRES.1SG I PROG stay here until be.PRES.1SG I PROG go home
   'I stay here until go home'

  b. sand yn mynd i+mewn i' fo os nei di neud fel 'na. [Subject-pred.]
   sand PROG go into to it if do.FUT.2SG you.SG do like that
   'sand going into it if you do like that.'

   c. dest 'di mynd am bod 'o 'di pwyso 'i   wats.   [Predicate]

       just PERF go   for be he PERF press   CL.3SG.M watch

       'just gone because he had pressed his watch.'

   d. a   'wn   os ti'   ddim isie.   [Fragment]

       and this.M if be.PRES.2SG NEG   want

       'and this if you don't want.'

3 a. w   i 'n   gwbod sut mae   hware hwnna.   [Finite]

       be.PRES.1SG I PROG know   how be.PRES.3SG play   that.M

       'I know how to play that.'

   b. 'na2   pam by'   nw 'n   cwmpo.   [Pred. det.]

       that+is why be.FUT.3SG they PROG fall

       'that's why they will be falling.'

   c. neb   yn   gwbo' lle   oedd   y double+decker.   [Subject-pred.]

       no-one PROG know   where be.IMPF.3SG the double-decker

       'no-one knowing where the double decker was,'

   d. gwbod lle   ma'n   nw.   [Predicate]

       know   where be.PRES.3PL they

       'know where they are.'

3 Other Matters

There are other types of utterances which occur in the database which are not included in the above taxonomy.

3.1 Songs and English utterances

Utterances which are expressions which are taken form songs and utterances which are completely English are not included in the taxonomy. This is not to say that utterances which are complete English expressions are not of interest. But they are not considered in this work.

3.2 Extralinguistic matters

We include under this heading exclamations, verbal pauses and noises, illustrated respectively in the following examples,

9 a. oo!

 'oh'

 b. ymm +...

 'uhm …'

 c. doing@i doing@i doing@i doing@i <doing@i doing@i> [>] [= onomatopia]!

The verbal pauses which occur alone indicate an unfinished utterance — in more detail, an utterance which has not started.

3.3 Unanalysable data

There are over 105,776 utterances which are produced by first language speakers of Welsh in the database (see table 1) but not all provide the data for this work. The following types of utterances are excluded.

First there are utterances which have not been clear to the transcribers and are entirely missing data, represented by xxx in the data lines of the corpus and XX[xx] in the syntactic description.

Second, there are also examples of utterances which contain some missing data but which is sufficient to conceal what is said and therefore conceal the syntax of the utterance.

10 a. xxx [= 3 sill] i+lawr!

 xxx   down

 'xxx down.'

 b. ie xxx [= 6 sill].

 yes xxx

 'yes xxx.'

 c. xxx [= 1 sill] hwnna xxx [= 2 sill].

 xxx   that xxx

 'xxx that xxx.'

Third, there are examples of utterances which contain English phrases and which overall cannot be said to have Welsh syntax.

11 a. <is that> [% Saesneg] [?] tywod?

          sand

 'is that sand.'

b. <come on> [% Saesneg] 'ta1.

                              then

'come on then,'

c. efo   'r   car 'di   <broken down,,> [% Saesneg] ia.

with   the car PERF   yes

'with the car broken down, yes.'

Such utterances are of interest and are open to analysis. But that will not be undertaken in this work. In some instances, the English material forms a phrase which fits into the overall syntax of the utterance.

12 a. ti'   gwel' <racing car> [% Saesneg] 'ma?

    be.PRES.2SG see   here

    'you see this racing car?'

b. fi sy   'n   neud <yorkshire pudding> [% Saesneg] [?].

    I   be.PRES.SG PROG make

    'it's me who makes Yorkshire pudding.'

c. ti'   isie   cal   <cup o' tea> [% Saesneg]?

    be.PRES.2SG want have

    'do you want to have a cup of tea?'

Such utterances have overall Welsh syntax and are included in the taxonomy.

Fourth, there are examples which could be fragments but are unfinished and cannot therefore be said to be fragments with certainty.

13 a. mewn fan+hyn +...

    in   here

    'in here …'

b. fanna +...

    'there …'

c. &sle &twt &w honna +...

    &   &   &   this.F

    '& & & this …'

Although the unanalysable data are not exploited in a statistical profile of types of utterances, they do provide data for the analysis of the phrases which are within them (not undertaken here).

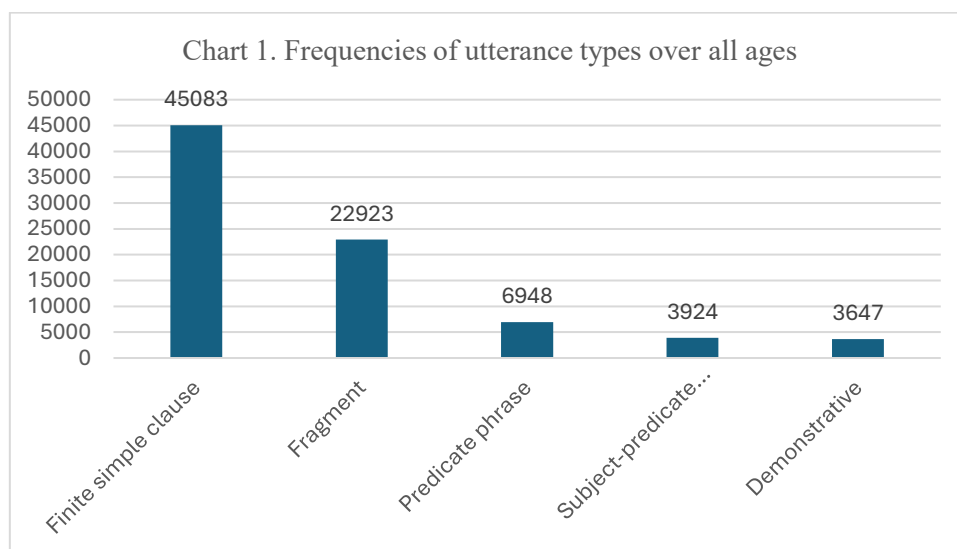# 4 Frequencies and longitudinal comparisons

4.1 Overall frequencies

Table 1 gives the frequencies for data which figure in the taxonomy and data which do not. There are a total of 105,776 utterances produced by first language speakers of Welsh. Of these, 78.02% provide analysable data.

Table 1. Overall frequencies of the syntactic types

Finite clause	45083		54.63%
Fragment	22923		27.78%
Predicate phrase	6948		8.42%
Subject-predicate sequence	3924		4.75%
Demonstrative	3647		4.42%
		82525	78.02%
Missing data entirely	8309		
Extralinguistic	6362		
Fragments with missing data	3994		
English utterance entirely	2437		
Fragments, unfinished	1799		
Songs	204		
Fragments with English phrase(s)	146		
		23251	21.98%
		105776	

Chart 1 gives a graphic display of the details about the analysable data in table 1.

Chart 1. Frequencies of utterance types over all ages

Utterance type	Frequency
Finite simple clause	45083
Fragment	22923
Predicate phrase	6948
Subject-predicate…	3924
Demonstrative	3647

Of the analysable data, finite sentences make up just over half of all utterances. The second most frequent type of utterance are fragments, accounting for over one quarter. All other utterances are comparatively low frequency. The striking feature of this table is the high frequencies of fragments. Together, finite clauses and fragments make up 82.41% of the total of analysable data. Another feature is that utterances which are made up of a predicate phrase are more frequent than the syntactically more-complex subject-predicate utterances.
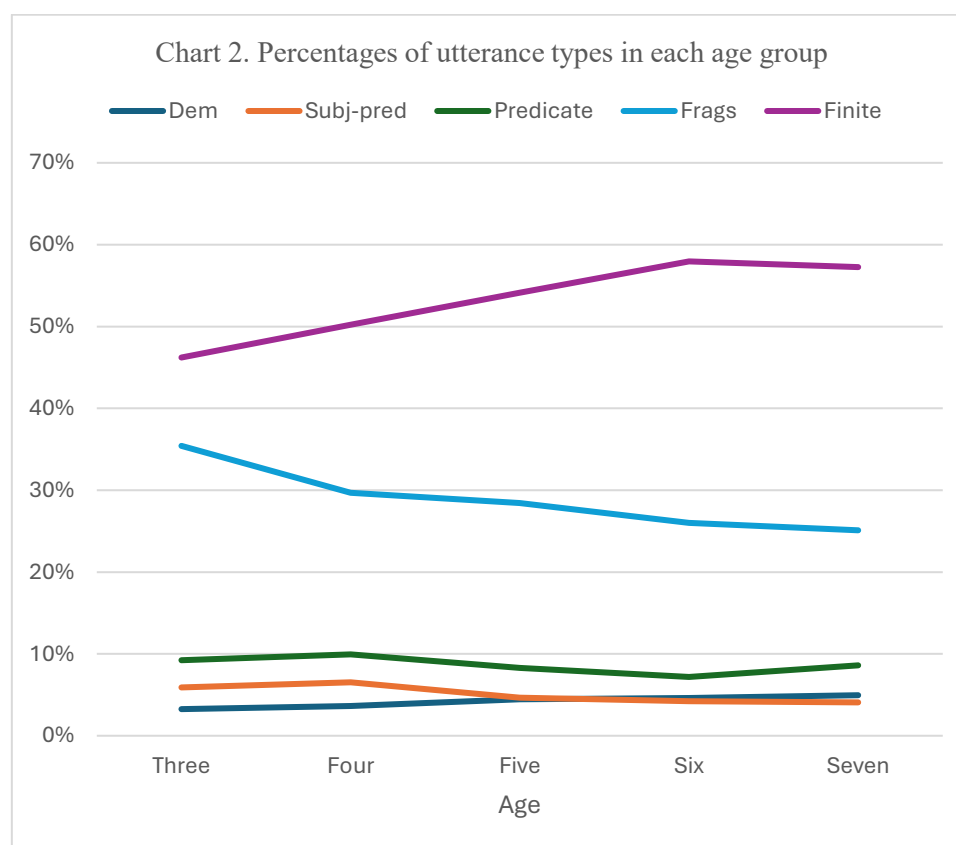
4.2 Age groups

In this section, frequencies are presented which confirm the hierarchical profile given in table 1 and which also provide longitudinal comparisons. The statistics are based on (i) frequencies of the syntactic types for the individual age groups, (ii) average use of the syntactic types in the age groups and (iii) the number of speakers who use the syntactic patterns in the individual age groups.

Table 2 gives the frequencies and percentages for the syntactic patterns in each age group.

Table 2. Frequencies and percentages by age and utterance types
(percentage base is the total for each age group)

Age	Finite	%	Fragments	%	Predicate	%	Subj-pred	%	Demonst.	%	Totals
Three	2727	46%	2090	35%	544	9%	348	6%	192	3%	5901
Four	5355	50%	3164	30%	1060	10%	696	7%	387	4%	10662
Five	15507	54%	8144	28%	2378	8%	1331	5%	1279	4%	28639
Six	9971	58%	4475	26%	1236	7%	729	4%	793	5%	17204
Seven	11523	57%	5050	25%	1730	9%	820	4%	996	5%	20119

A more graphic presentation of the percentages in table 2 is given in chart 2.

Chart 2. Percentages of utterance types in each age group

The table confirms the hierarchical order in table 1, which in descending order is: (i) finite clauses, (ii) fragments, (iii) predicates and (iv) subject-predicates and demonstratives. Chart 2 also shows that the percentages of finite clauses increase over the age range and, in contrast, the percentages of fragments decrease. It is reasonable to assume that these two trends are related: the reduction in the percentages of fragments is caused by the use of a finite clause instead.

Table 3 gives average use of the utterance types per speaker in each year group, that is, the total number of examples in each age group is divided by the total number of speakers, including speakers who have not used any examples.

Table 3. Average use of the utterance types in each year group. 'All' = all users in an age group											
		Finite		Fragments		Predicate		Subj + Pred		Demonst.	
Age	All	Freq.	Aver.	Freq.	Aver.	Freq.	Aver.	Freq.	Aver.	Freq.	Aver.
Three	38	2727	71.76	2090	55.00	544	14.32	348	9.16	192	5.05
Four	63	5355	85.00	3164	50.22	1060	16.83	696	11.05	387	6.14
Five	143	15507	108.44	8144	56.95	2378	16.63	1331	9.31	1279	8.94
Six	75	9971	132.95	4475	59.67	1236	16.48	729	9.72	793	10.57
Seven	74	11523	155.72	5050	68.24	1730	23.38	820	11.08	996	13.46

A graphic representation of the details in table 3 is given in chart 3.

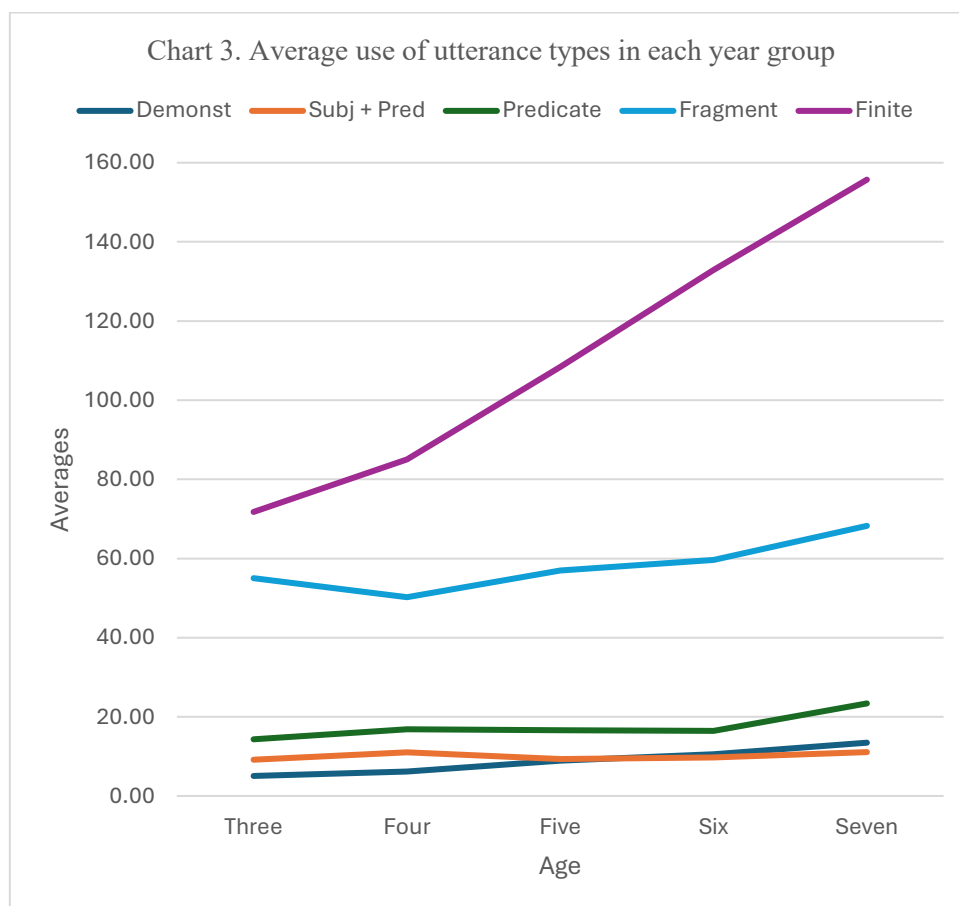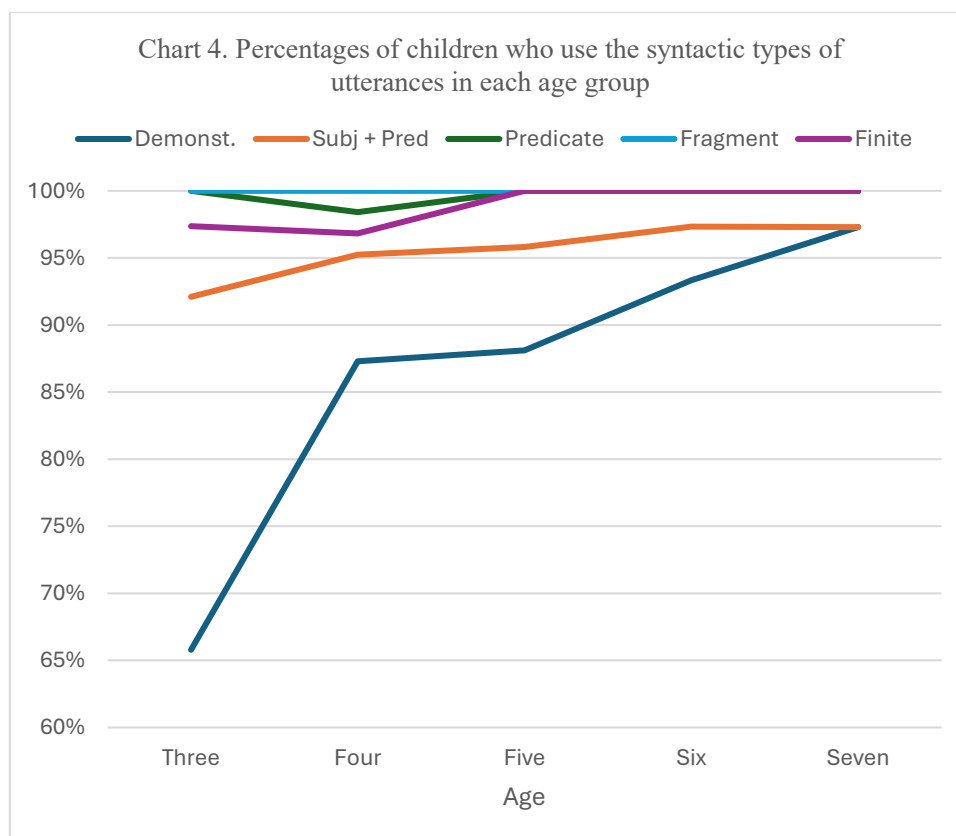Chart 3. Average use of utterance types in each year group

Chart 3 also confirms the hierarchy of utterance types. In addition, the average use of utterance types increases over the years. The greatest increase is seen in the use of finite clauses. The increases in the use of other utterance types are much smaller. It is interesting to note that the proportion of fragments in each age group decreases in the case of fragments (chart 2) but their frequencies are sufficient to provide an increase in their average use (chart 3).

Table 4 gives the number speakers in each age group who use the syntactic types, expressed as a percentage of the total number of speakers in each age group.

Table 4. Longitudinal percentages of actual users of utterance types
'All' = all users in age group. 'Users' = actual users of utterance type in age group

Age	All	Finite		Fragments		Predicate		Subj + Pred		Demonstrat.	
		Users	%s	Users	%s	Users	%s	Users	%s	Users	%s
Three	38	37	97%	38	100%	38	100%	35	92%	25	66%
Four	63	61	97%	63	100%	62	98%	60	95%	55	87%
Five	143	143	100%	143	100%	143	100%	137	96%	126	88%
Six	75	75	100%	75	100%	75	100%	73	97%	70	93%
Seven	74	74	100%	74	100%	74	100%	72	97%	72	97%

Chart 4 gives a graphic representation of the details in table 4.

Chart 4. Percentages of children who use the syntactic types of utterances in each age group

This chart shows that fragments are used by all speakers at all ages. This is also mainly so in the case of predicates and finites. It is especially striking to see that there are a very small number of children of three and four years of age who do not use a finite utterance in their performance. There are two reasons for this. One is reticence. Circumstances at the time of recording brought about instances of a single child in conversation with a researcher. Attempts at prompting conversation by the latter only succeeded in bringing about one-word responses. The other reason is language switching. Some children switched to English, leaving a smaller number of Welsh utterances which did not contain a finite clause. It is noticeable, too, that the numner of users of subject-predicate utterances increases over the years. But the most striking increase is that of predicative demonstratives which move from 66% at three years of age to 97% at seven years of age.

It is interesting to compare charts 3 and 4. Chart 4 shows that finites, fragments and predicates are quite similar in terms of the numbers of children who use them. But chart 3 shows that there are much bigger differences between these types in terms of their average use by speakers (including by speakers who have zero use). The longitudinal development is in terms of average use and not so much in terms of numbers of speakers, that is, roughly the same number of speakers increase their use of these syntactic types.