Type/Token Ratios: what do they really tell us?

Brian Richards

Journal of Child Language / Volume 14 / Issue 02 / June 1987, pp 201 - 209
DOI: 10.1017/S0305000900012885, Published online: 17 February 2009

J. Child Lang. 14 (1987), 201–209. Printed in Great Britain

Type/Token Ratios: what do they really tell us?*

BRIAN RICHARDS
University of Bristol

(*Received* 4 *October* 1986)

ABSTRACT

Type/Token Ratios have been extensively used in child language research as an index of lexical diversity. This paper shows that the measure has frequently failed to discriminate between children at widely different stages of language development, and that the ratio may in fact fall as children get older. It is suggested here that such effects are caused by a negative, though non-linear, relationship between sample size (i.e. number of tokens) and Type/Token Ratio. Effects of open and closed class items are considered and an alternative Verbal Diversity measure is examined. Standardization of the number of tokens before computing Type/Token Ratios is recommended.

In an investigation into the language development of 480 children between the ages of three and eight, Templin (1957) compares the total number of different words (types) in 50 consecutive utterances with the total number of words in the same 50 utterances (tokens). She concludes:

> This ratio is approximately one different word for slightly over every two words uttered. The ratio shows little variation over the age range tested and among subsamples, sex, and SES groups.
>
> (Templin 1957: 115)

In theory, Type/Token Ratio (TTR) weights range of vocabulary for size of speech sample. This is necessary because it is reasonable to assume that the more words sampled, the greater the probability of finding more different words. The larger the resulting TTR the less repetitive the vocabulary usage; if a speech sample contains 20 words and they are all different we obtain the 'ideal' TTR: 20/20 = 1·00. On the other hand the sample in which the same

[*] Data referred to in this paper originates from research supported by the ESRC (C00428325005). I would like to thank Peter Robinson, Gilbert Davies and Bencie Woll for their comments on an earlier draft, and Mary Gutfreund for her assistance and helpful suggestions. Address for correspondence: School of Education Research Unit, University of Bristol, 22 Berkeley Square, Bristol BS8 1HP, UK.

word is repeated 20 times yields a figure of $1/20 = 0{\cdot}05$. TTR, therefore, looks like a useful index of lexical diversity independent of sample size, and it has been used in this function in a wide range of studies, both as an input variable, and as a child language variable. Variation in TTR has figured, for example, in investigations into the 'motherese' register (e.g. Drach 1969, Broen 1972, Phillips 1973), the fine-tuning hypothesis (Cross 1977), and the relationship between input and children's referential or expressive orientation (Della Corte, Benedict & Klein 1983). As a measure of early child language it has been used in studies of individual differences (e.g. Lieven 1978), longitudinal case studies (e.g. Fletcher 1985), and, more recently, as a predictor of foreign language learning aptitude and attainment (Skehan 1986).

Miller (1981) and Fletcher (1985) have suggested that Templin's fairly uniform figure for TTR of approximately 0·5[1] could be used as a base-line for assessment. According to Fletcher:

> ... if the TTR falls well above or below 0·50, we can conclude that the lexical diversity is not normal. Why this should be, of course, we will not know without investigating further. As a measure, TTR is inevitably crude: but a low TTR is a flag for possible restrictions on the range of use of vocabulary by the child in his syntactic structures.
>
> (Fletcher 1985:47)

Miller also sees TTR as a potential measure for diagnosing language impairment:

> The consistency of this measure makes it enormously valuable as a clinical tool. For example, if a normal hearing child's TTR is significantly below 0·50 we can be reasonably certain the sparseness of vocabulary use is NOT an artefact of SES but is probably indicative of a language-specific deficiency.
>
> (Miller 1981:41)

In order that TTRs calculated by others should be comparable with those of Templin's sample, Miller recommends close duplication of Templin's methodology in the collection of data and counting of words, including the use of a set of 50 consecutive utterances as a basis for all calculations.

However, one thing about the TTRs obtained from Templin's data is puzzling. This is the fact that they do appear to be so consistent. After all, the youngest children were three, and the oldest were eight. The number of DIFFERENT words in 50 utterances of the older children was nearly double that of the three year olds, yet lexical diversity as measured by TTR was slightly

---

[1] Actual mean values for each age group, calculated by Miller (1981) from Templin's figures range from 0·43 to 0·47.

smaller. If children are extending their vocabulary, why is there no accompanying development in variety of usage?

The answer to this question is twofold. Firstly, there is a basic theoretical problem to do with variation in sample size. Secondly, there is a methodological problem resulting from the way sample size was standardized (number of UTTERANCES rather than number of TOKENS).

The theoretical problem was highlighted by an attempt by the author of this paper to quantify the range of main verbs which co-occur with various auxiliaries. The aim was to identify children whose usage of auxiliaries tended to be stereotyped (Richards, in prep.). One approach might have been to compare the total number of tokens of a certain auxiliary with the number of types of main verbs with which it co-occurred. A ratio analagous to the TTR discussed here could then be calculated. However, an inspection of the data showed that a combination of factors (small numbers of tokens, and variation in the number of tokens) rendered such a ratio meaningless. The figure obtained had more to do with the number of tokens than with the development of the auxiliary verb in question. It is hardly surprising that this should be so when samples are small, but it can also be predicted that TTRs calculated from a large number of tokens will generally be lower than those calculated from a smaller number. This is because, while obtaining more words also means obtaining more different words, each additional different word sampled is a move towards exhausting the subject's repertoire of commonly used vocabulary.

Suppose a child's total vocabulary consists of N types. It is only possible for that child to attain the 'ideal' TTR of 1·0 if the number of tokens in a speech sample is equal to, or less than N. If the number of tokens is double the child's entire range, then the maximum possible TTR is $N/2N = 0{\cdot}5$. Similarly, if there are four times as many tokens as the number of items in the vocabulary, then the TTR can attain a value no greater than $N/4N = 0{\cdot}25$, and so on.

The effect can be demonstrated by a simple experiment. If a continuous passage of prose, or a child language transcript is analysed for TTR, basing the calculation on the first ten words, then on the first 20 words, then on the first 30, and so on, increasing the number of words by ten each time, then it is possible to plot the number of tokens in intervals of ten against TTRs for cumulatively larger samples. The result of such an exercise can be seen in Fig. 1. The prose text used was the beginning of Chapter 1 of Chomsky 1965 (pp. 3–4). The child language sample was a 45 minute recording of Sian (2;2.2), one of the subjects in the auxiliary verb study referred to above. Sian was a talkative child and her language development was unusually advanced for her age. At this stage her MLU based on structured utterances only (see Wells 1985) was 3·9 morphemes.

Variation in criteria for identifying types and tokens can clearly be

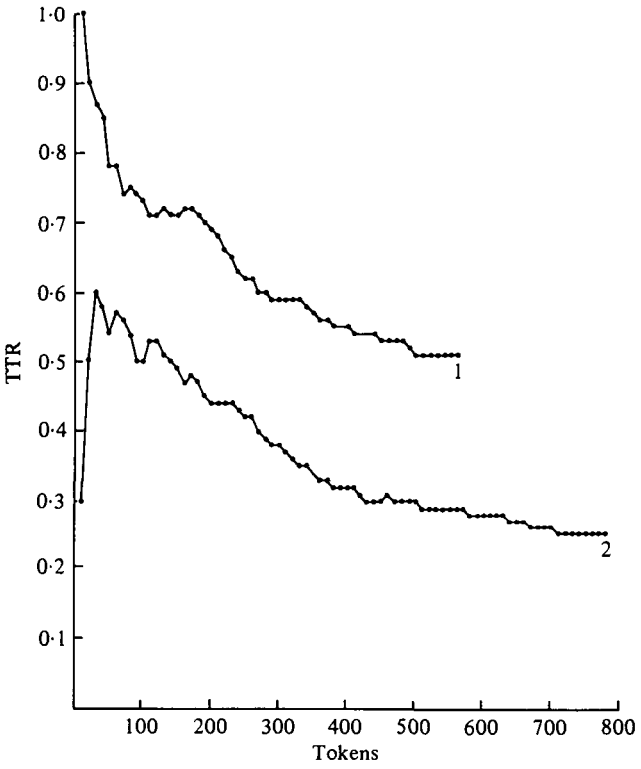

Fig. 1. TTR for cumulatively larger samples plotted against number of tokens for (1) a passage of continuous prose, and (2) 45 minutes of spontaneous speech from a 2-year-old.

problematic. Rules for calculating TTR in this experiment were adapted from Templin (1957:160), Miller (1981:42) and Fletcher (1985:47):

(1) Compound nouns and hyphenated words count as one word.

(2) Expressions functioning as a single unanalysed unit for the child (e.g. *all right*) are counted as one word.

(3) Parts of verbal groups are treated as separate words. Thus *have been playing* counts as three tokens: HAVE, BE, and PLAY.

(4) Inflections and bound morphemes do not count as separate words. Inflected and non-inflected forms of the same stem count as a single type.

(5) Contractions of subject and predicate (*I'm, she's*) are treated as two words. Full and contracted forms are treated as a single type.

(6) Contractions of the verb and negative particle (e.g. *can't*) count as one token. These verb forms count as separate types from the corresponding affirmative forms.

(7) Homonyms, for example *do* as an auxiliary, and *do* as a main verb are counted as separate types.

(8) Interjections (*oh*, *ouch*!) are excluded.

The curves in Fig. 1 show the extent to which, if the ability to vary vocabulary is in effect held constant, TTR is predicted by sample size. But they also illustrate the effect of contextual variables while the sample size is still fairly small. The studied style of the academic results in ten types in the first ten tokens and 18 types in the first 20. The ratio then falls away rapidly as the number of tokens increases, only to rise again slightly when there is a change of theme. The 2-year-old, on the other hand, produces a more erratic curve for the first 175 words, the trough at the beginning corresponding with repetitive recitation of colour names, and the later peaks being caused by the spontaneous labelling of a series of pictures of different animals.

The graph also demonstrates that if sample size is held constant, the prose of the Professor of Linguistics consistently has a higher TTR than the spontaneous speech of the 2-year-old. If, on the other hand, sample size varies extensively, it is possible to produce TTRs which appear to demonstrate that the speech of the 2-year-old shows greater lexical diversity than the prose of the Professor of Linguistics!

What Sian's graph cannot tell us about is the ontogenesis of TTR. Templin's study was cross-sectional in design and showed no increase in TTR in older children, but two pieces of research which use this measure longitudinally show a fall in TTR as children get older, despite evidence of language progress on other measures. Over a period of six months Lieven's data from Kate and Beth show a decline from 0·59 to 0·46 for Kate, and from 0·33 to 0·12 for Beth (Lieven 1978, Table 1:176). Similarly, Fletcher (1985) found that Sophie's TTR fell from 0·34 at 2;4, to 0·31 at 3;0, and again to 0·26 at 3;5. This brings us to the methodological question of how best to standardize the size of samples. In Lieven's study TTRs covary with the number of utterances on which calculations were based. As the two children grew older, so their rate of production of comprehensible utterances increased. TTRs from later samples were therefore based on a larger number of tokens. The trend for Sophie (Fletcher 1985) cannot be explained by variation in the number of utterances; Fletcher based all calculations on 100 utterances. Interestingly, though, he recalculated the TTR for the sample at 3;5 from 50 utterances for greater comparability with Templin's figures. The result corresponds with the prediction made above: TTR rose from 0·26 to 0·39.

The data from Sophie shows that standardizing the number of utterances is not sufficient if TTR is to be a valid developmental measure. Standardizing according to recording time is also unsatisfactory. Using the Bristol child language corpus (Wells 1985), Skehan (1986) used a standard number of timed samples from each child. TTRs and several other vocabulary measures

were calculated for 19 children at 39 months. TTRs proved to be negatively correlated with five out of six of the measures,[2] but the only statistically significant effect was TTR × Number of tokens ($r = -0{\cdot}61$, d.f. = 17, $P < 0{\cdot}01$). Again, this is consistent with the prediction made above.

The problem is that the linguistically more advanced child tends to produce longer utterances (both in morphemes AND words) and more frequent utterances per timed unit. Since we have shown that, if the ability to vary lexis is in effect held constant, TTR is NEGATIVELY associated with the number of tokens, any development in areas which increase the number of tokens (e.g. MLU, or rate of speech) will automatically REDUCE development of TTR. This effect may be so great as to cause an apparent DECREASE in lexical diversity over time.

At this point it is worth considering the relationship between the various word classes and the shape of the above curves. It is reasonable to assume that words belonging to the closed classes, in particular, are responsible for the initial steep decline in the graphs, since as a relatively small set of frequently used items they are often more likely to re-occur while the sample size is still small than words of the open classes. Equally, one would expect TTRs calculated from open class items to be generally higher than those based on closed class items. In order to test this, Sian's transcript was reanalysed such that separate TTRs were produced for open and closed classes (see Fig. 2). The open/closed distinction was derived from Quirk, Greenbaum, Leech & Svartvik (1985:67) but splits adverbs into a closed group such as *there*, *here*, and *now*, and an open group consisting of those with an adjectival base such as *quickly* and *easily* (Quirk *et al.* 1985:73). Numerals were assigned to the open classes, and the non-contracted negative particle *not* and the infinitive marker *to* to the closed classes.

It can be seen that the predictions are borne out by the data from Sian. Above 30 tokens, TTRs are consistently higher for open class items, but the pattern of development as the number of tokens increases is different for each class. Since the ratio of open class words to closed class words will vary from child to child and from sample to sample TTRs calculated from all tokens contain another source of variation which casts additional doubt on their validity. In addition, one might question the inclusion of function words at all, if the aim is to obtain an index of vocabulary development. Of course, the classes of word which are validly included will depend on the questions being asked by researcher or clinician. Crystal (1982), for example, calculates separate TTRs for 'major' (contentives) and 'minor' lexical items (func-

[2] There is a related finding in Nelson (1973). TTRs were calculated for 18 children at 2;0 in an attempt to adjust estimates of vocabulary size for variations in length of recordings, talkativeness, and utterance length. Nelson found TTR to be unrelated to vocabulary measures as well as to other child language indices.

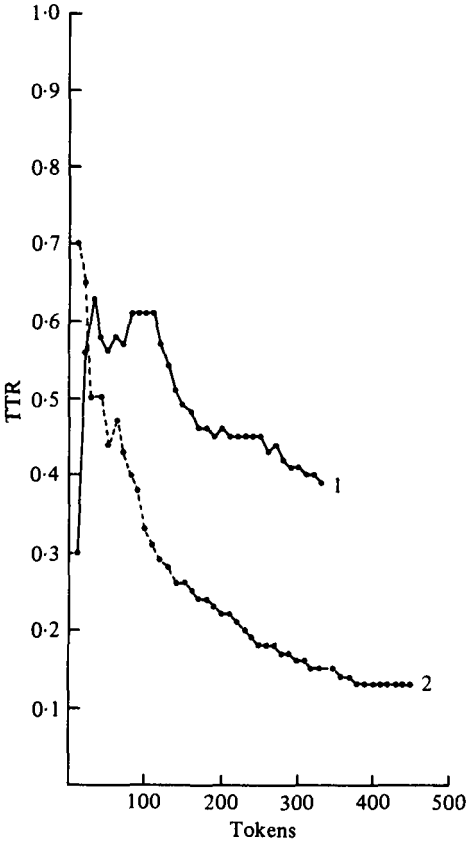

Fig. 2. TTR for cumulatively larger samples plotted against number of tokens for (1) open class items and (2) closed class items in the speech of a 2-year-old.

tors). What remains clear, however, is that figures from both closed and open classes are affected by sample size.

What do TTRs really tell us? Apparently they may tell us more about the number of tokens from which they are calculated than about range of vocabulary usage, and in some cases the number of tokens is itself determined by factors such as talkativeness and mean length of utterance. One remedy is to do what Cross (1977) has done and standardize the number of tokens from each child. How large this number should be, and whether the words should be consecutive or randomly extracted, can only be determined by further experimentation. Nevertheless, the data from Sian suggests that ratios may be unstable before 200 words, and that it takes a sample of between 400 and 500 tokens to minimize effects of sample size. In practice, there will be children who fail to produce whatever standard number of tokens has been

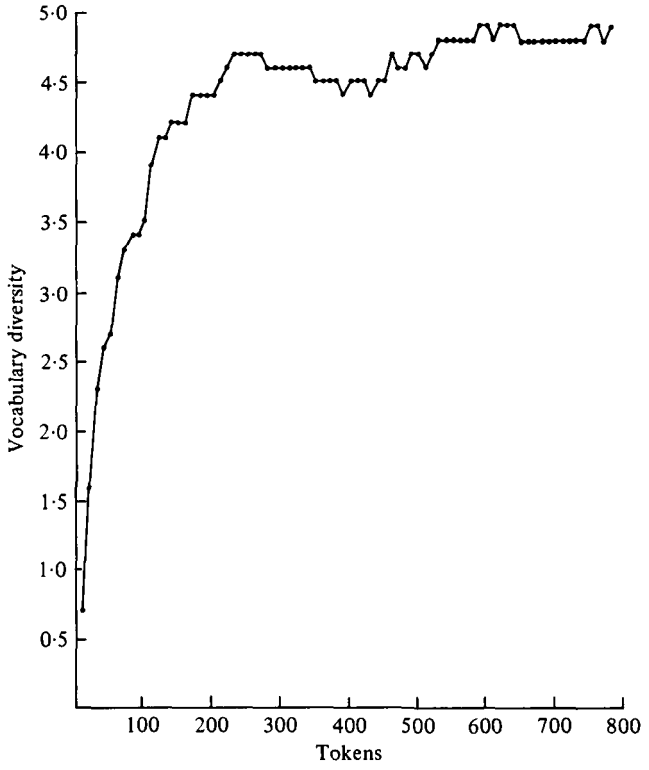

Fig. 3. Scores for cumulatively larger samples on Carroll's Vocabulary Diversity measure plotted against number of tokens in the speech of a 2-year-old.

decided on. If this is the case, it is possible to predict the TTR of a larger sample by regression, using the procedure described above to relate TTR to varying numbers of tokens. Since the relationship between the two is one of diminishing effects, the logarithm of the number of tokens can be used to obtain a more linear relationship between the two variables (Blalock 1979). An alternative may be to use J. B. Carroll's vocabulary diversity measure. According to Carroll:

> A measure of vocabulary diversity that is approximately independent of sample size is the number of different words divided by the square root of twice the number of words in the sample.
>
> (Carroll 1964:54)

This measure was tested on the combined open and closed class vocabulary in Sian's transcript. However, while TTR was negatively correlated with sample size, this index proves to be positively related to the number of tokens

(Fig. 3). The diversity figure rises sharply as the sample increases in size, and only begins to level out when approximately 250 tokens have been reached. If this is generally the case with child language samples, then verbal diversity has no clear advantage over TTRs based on a standard sample size.

As a research tool, TTR calculated from a uniform number of tokens may have a useful role to play. Before it can be of real value as a clinical tool, however, norms need to be established which are more soundly based than the figure of approximately 0·5 which was obtained from Templin's study.

REFERENCES

Blalock, H. M. (1979) *Social statistics*. Second edition. London: McGraw-Hill.

Broen, P. (1972). The verbal environment of the language-learning child. ASHA *Monograph* **17**.

Carroll, J. B. (1964) *Language and thought*. Engelwood Cliffs, NJ: Prentice Hall.

Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge, MA: M.I.T. Press.

Cross, T. G. (1977) Mothers' speech adjustments: the contributions of selected child listener variables. In C. Snow & C. A. Ferguson (eds), *Talking to children: language input and acquisition*. Cambridge: C.U.P

Crystal, D. (1982) *Profiling linguistic disability*. London: Edward Arnold.

Della Corte, M., Benedict, H. & Klein, D. (1983) The relationship of pragmatic dimensions of mothers' speech to the referential-expressive distinction. *Journal of Child Language* **10**. 35–43.

Drach, K. M. (1969) The language of the parent: a pilot study. In K. M. Drach, B. Kobashigawa, C. Pfuderer & D. Slobin, *The structure of linguistic input to children*. University of California at Berkeley: Language Behaviour Research Laboratory, Working Paper No. 14.

Fletcher, P. (1985) *A child's learning of English*. Oxford: Blackwell.

Lieven, E. V. M. (1978) Conversations between mothers and young children: individual differences and their possible implication for the study of child language learning. In N. Waterson & C. Snow (eds), *The development of communication*. Chichester: Wiley.

Miller, J. F. (1981) *Assessing language production in children: experimental procedures*. London: Edward Arnold.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development* **38**. No. 149.

Phillips, J. R. (1973) Syntax and vocabulary of mothers' speech to young children: age and sex comparisons. *Child Development* **44**. 182–5.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985) *A comprehensive grammar of the English language*. London: Longman.

Richards, B. J. (in prep.) Individual differences and the development of the auxiliary verb system in young children.

Skehan, P. (1986) Early lexical development and the prediction of foreign language learning success. Paper presented at the CILT/ESRC Conference on Second Language Acquisition, Windsor, Berkshire.

Templin M. C. (1957) *Certain language skills in children*. Minneapolis: University of Minnesota Press.

Wells, C. G. (1985) *Language development in the pre-school years*. Cambridge: C.U.P.