MATHEMATICAL COGNITION, 1997, *3* (1), 63–85

From 834 to Eighty Thirty Four: The Reading of Arabic Numerals by Seven-year-old Children

Richard Power
Information Technology Research Institute, University of Brighton, Brighton, UK

Maria Felicita Dal Martello
Dipartimento di Psicologia Generale, University of Padua, Padua, Italy

A group of seven-year-old Italian children was asked to transcode from arabic numerals to verbal numerals. The stimuli were written arabic numerals such as 365; the responses were spoken Italian numerals such as *tre cento sessanta cinque*. Several regular error patterns were observed. The less-advanced subjects fragmented complex stimuli, so that 365 would be transcoded by a numeral of the form *thirty six five*, or ignored part of the stimulus. Children who had just learned the word *cento* often used it in initial position, so that 200 was transcoded as *cento due*. Later, they experimented with various ways of subdividing the digit string and introducing a multiplicand, producing such errors as 834 = *eighty thirty four*, 803 = *eighty thousand and three*. The development of transcoding ability was explained by an asemantic model using production rules.

People in our culture are familiar with two notations for naming the natural numbers: verbal numerals (e.g. three hundred and sixty-five) and arabic numerals (365). Verbal numerals are used mainly in conversation; arabic numerals are preferred for performing calculations and for writing down large numbers. By the age of ten, most children have mastered both the arabic system and a verbal system and can translate from one to the other.

Investigations of numeral transcoding usually employ two tasks, which we may call "reading" and "dictation". In a reading test, the experimenter presents a written arabic numeral and the subject responds by producing a spoken verbal numeral. In a dictation test, conversely, the experimenter presents a spoken verbal numeral which the subject tries to write down in arabic digits.

The present article investigates the performance of seven-year-old Italian children on the reading task. The subjects were beginning their second year of

Requests for reprints should be sent to Richard Power, Information Technology Research Institute, University of Brighton, Lewes Road, Brighton BN2 4AT, UK. Email: rjdp@itri.bton.ac.uk
We thank Brian Butterworth and the referees for helpful comments.

elementary school, by which time most children start to form ideas about the structure of complex numerals and hence to produce revealing errors. We hoped that such errors would provide evidence both about the transcoding mechanism and about the way in which it is acquired.

Theories of Numeral Transcoding

During the early 1980s there was a resurgence of interest in numeral transcoding, beginning with studies on acalculia patients by Deloche and Seron (1982) and Seron and Deloche (1984). On the basis of these and other studies, McCloskey and colleagues proposed an influential general architecture for number processing (McCloskey, Caramazza, & Basili, 1985). In McCloskey's model, number processing is based on an abstract internal representation of numbers which supports the calculation system. Transcoding from arabic to verbal numerals occurs through two stages: first, perception of the arabic numeral, leading to its formulation in the abstract internal code; second, production of the verbal numeral. The internal representation of the number thus serves as the output of the perception system and as the input to the production system. Transcoding from verbal to arabic numerals occurs in an analogous manner.

An alternative to McCloskey's theory was proposed by Deloche and Seron (1987), who argued that transcoding could be performed by an *asemantic* process: that is, by a system that derives a verbal numeral from an arabic numeral (or vice-versa) directly, without constructing an abstract internal representation of the number. More recent evidence favours a multi-route model (Cohen, Dehaene, & Verstichel, 1994), in which transcoding can occur either through a semantic route or through an asemantic route. Cipolotti and Butterworth (1995) describe the remarkable case of a patient (S.A.M.) who made frequent errors on transcoding problems even though he could perceive and produce numerals reliably when asked to solve calculation problems. Cipolotti (1995) also reports a patient (S.F.) with a specific difficulty in reading arabic numerals aloud (i.e. in transcoding to spoken verbal numerals), even though he could both understand arabic numerals and produce verbal numerals in other contexts.

These studies, and others reviewed by Seron and Noel (1995), suggest that people normally transcode between verbal and arabic numerals by means of an asemantic system; the semantic route is activated only when the number has to be utilised in some other task (e.g. memory or calculation).

Models of Asemantic Transcoding

Asemantic transcoding implies a syntactic mapping from the input code (e.g. arabic numerals) to the output code (verbal numerals). Such a mapping would presumably not be feasible for languages that use a non-decimal system, such as

Suppire or Sora (Welmers, 1973). In Suppire, the number 379 is named by a numeral of the form *eighty four and twenty two and ten and nine*, or $(80 \times 4) + (20 \times 2) + (10 + 9)$; in Sora, the same number is named by *twelve six twenty twelve seven*, or $[(12 + 6) \times 20] + (12 + 7)$. Of course, these cultures do not use arabic numerals; however, anyone wishing to read aloud 379 in Suppire or Sora would obviously have to do some mental arithmetic, so following the semantic route.

Deloche and Seron (1987) propose an asemantic transcoding algorithm that interprets the digits of the arabic numeral in groups of three. First, the arabic numeral is broken up into triplets, starting from the right; this grouping may already be indicated by punctuation marks (e.g. commas in Britain, points in Italian). The left-most group is then loaded into a three-digit frame, filling empty positions with zeros if the group has only one or two digits. A set of rules specialised for numbers up to 999 is then applied to the digits in the frame to produce the first few words of the verbal numeral. Depending on the number of groups remaining, a suitable multiplicand word is then produced, and the next group is loaded. This process continues until no groups are left.

For example, suppose that the task is to read the arabic numeral 12345 in English. First, the numeral is divided into two groups, 12 and 345. Next, the group 012 is loaded into the frame, the extra zero being added to fill the empty place. Specialised rules applied to this three-digit group yield the word *twelve*. Since there is one remaining group, the multiplicand *thousand* is produced. Now the group 345 is loaded into the frame and the words *three hundred and forty-five* are produced. Here the process halts since no groups remain.

Theoretically this model is odd because of the arbitrary importance assigned to triplets. It is true that for numbers up to a billion, the major division in an arabic numeral can be found by segmenting the digit string into triplets and detaching the left-most group. However, this is not the whole story. First, within each group of three digits, the first digit can be divided from the other two: thus the structure of 345 is clearly 3(45) rather than (34)5. More seriously, for very large arabic numerals such as 2,300,000,000,000 (*two thousand three hundred billion*), the major division should come after 2300, not after 2. In other words, the procedure for transcoding this numeral should be to transcode 2300, then produce the multiplicand *billion*, then transcode the rest (in this case, only zeros). Note that a satisfactory transcoding algorithm must get the *structure* of the verbal numeral right as well as the words. This structure is shown by the intonation and rhythm with which the verbal numeral is spoken.

Another theory of verbal numeral production has been proposed by McCloskey, Sokol, and Goodman (1986). This theory is semantic rather than asemantic: it assumes that the input is a semantic representation of the form [3] 10EXP2 [6] 10EXP1 [5] 10EXP0 (for the number 365). However, such a semantic representation is almost a notational variant of arabic numerals: The only difference is that orders of magnitude are shown by symbols (e.g. 10EXP2) rather than by serial position. The model works by associating a syntactic frame with each order of magnitude. For a number in the hundreds (like 365) the syntactic frame is as follows:

{{ONES:_} MLT:H TENS:_ ONES:_}

Having selected a frame, the numbers for each order of magnitude are filled in:

{{ONES:3} MLT:H TENS:6 ONES:5}

Finally, elements of the frame are expressed by words: ONES:3 by *three*, MLT:H by *hundred*, TENS:6 by *sixty*, and ONES:5 by *five*, yielding the American-English numeral *three hundred sixty five*.

This theory has similar drawbacks to that of Deloche and Seron. First, it fails to articulate the structure 3(65) within the group of three digits: indeed, the syntactic frame erroneously brackets the verbal numeral as *(three) hundred sixty five* rather than *(three hundred) sixty five*. Secondly, the authors do not show how the method could be extended to large numbers. They suggest a recursive process exploiting “the fact that verbal numerals are made up of one or more units of the form [(ONES) hundred TENS ONES] MULTIPLIER”, but, as we have shown above, this is true only for numbers below a thousand billion.

As an alternative to these theories, we propose a model based on the analysis of numerals in Power and Longuet-Higgins (1978). The fundamental idea is that the constituents of complex verbal numerals are usually themselves verbal numerals. This allows us to name a number such as 365 by applying a *recursive* rule that refers to the names of simpler numbers:

> The English numeral for 365 is composed of the English numeral for 300, followed by the conjunction *and*, followed by the English numeral for 65.

Similar rules apply to the constituents:

> The English numeral for 300 is composed of the English numeral for 3, followed by the word *hundred*.

> The English numeral for 65 is composed of the English numeral for 60, followed by the English numeral for 5.

Eventually we arrive at numbers for which there are lexical rules:

> The English numeral for 3 is *three*.

To state these rules in a more concise and general way, we will introduce two conventions.

- The expression "eng(X)" denotes the English numeral for the arabic numeral X. Thus eng(3) denotes the English numeral for 3, i.e. *three*.
- In an expression like "eng(ABC)", the letters A, B, C are variables ranging over arabic digits. Thus ABC represents any three-digit arabic numeral.

With these notational conventions, the rules for naming 365 can be specified as follows:

eng(ABC) = eng(A00) *and* eng(BC)
eng(A00) = eng(A) *hundred*
eng(AB) = eng(A0) eng(B)
eng(3) = *three*
eng(60) = *sixty*
eng(5) = *five*

Note that the first three of these rules have been generalised. The rule for eng(ABC) is applicable not only to 365 but to most three-digit arabic numerals. We say "most" because there is a class of exceptions, namely those matching the pattern A00. Application of the ABC rule to 300 would lead either to infinite recursion or to an incorrect English numeral such as *three hundred and and*. We can avoid this error in two ways. Either we pose a condition on the ABC rule (i.e. that BC cannot be 00), or we order the rules so that exceptional cases (like A00) are encountered first. Preferring (for simplicity) the latter method, we give in Table 1 a set of rules for asemantic transcoding from arabic to British-English numerals.[1]

Table 1 introduces one new convention: The symbol α in the rule "eng(0α) = eng(α)" matches any *sequence* of arabic digits. The rule therefore stipulates that a leading zero in a string of arabic digits should be deleted. Note that the rules do not mention the numeral *zero* itself (or the more colloquial *nought*); strictly speaking, zero is not a natural number, and the word *zero* is not a constituent of any other numeral. (We use it for reading decimals, e.g. *nought point five*, but that is another matter.) Of course it would do no harm to add the rule "eng(0) = *zero*" at the top of Table 1, but this rule would not interact in any way with the rest of the table: it would not be invoked by any of the subsequent recursive rules.

To illustrate the transcoding procedure, we trace the process of reading (in British English) the arabic numeral 23004. Denoting the English numeral by eng(23004), we match the rules in Table 1 to this pattern, starting from the top of the list. (Note that at each stage we match the rules to the *whole* of the current

[1] To save space, some obvious lexical rules are omitted.

TABLE 1
Rules for Direct Transcoding from Arabic to English Numerals

eng(0α)	=	eng(α)	eng (A000)	=	eng(A) *thousand*
eng(1)	=	*one*	eng(A0BC)	=	eng(A000) *and* eng (BC)
eng(2)	=	*two*	eng(ABCD)	=	eng(A000) eng (BCD)
eng(19)	=	*nineteen*	eng(AB000)	=	eng(AB) *thousand*
eng(20)	=	*twenty*	eng(AB0CD)	=	eng(AB000) *and* eng (CD)
eng(90)	=	*ninety*	eng(ABCDE)	=	eng(AB000) eng(CDE)
eng(AB)	=	eng(A0) eng(B)	eng(ABC000)	=	eng(ABC) *thousand*
eng(A00)	=	eng(A) *hundred*	eng(ABC0DE)	=	eng(ABC000) *and* eng(DE)
eng(ABC)	=	eng(A00) *and* eng(BC)	eng (ABCDEF)	=	eng(ABC000) eng(DEF)

digit string, so that transcoding occurs top-down rather than left-to-right.) The first rule to match is:

eng(AB0CD) = eng(AB000) *and* eng(CD)

Instanciating this rule, with AB = 23 and CD = 04, we obtain

eng(23004) = eng(23000) *and* eng(04)

We must now reapply the rules in order to transcode 23000 and 04. The first rule to match eng(23000) is

eng(AB000) = eng(AB) *thousand*

with AB = 23. Matching eng(23) to the rule

eng(AB) = eng(A0) eng(B)

we obtain

eng(23) = eng(20) eng(3)

Lexical rules now apply, yielding eng(20) = *twenty* and eng(3) = *three*. To complete the English numeral, it remains to transcode 04. At this point the zero-deleting rule

eng(0α) = eng(α)

matches, yielding eng(04) = eng(4); once this leading zero has been deleted, the lexical rule eng(4) = *four* can be applied, the final result being *twenty three thousand and four*. The full derivation is shown by the tree in Fig. 1, which assigns the correct syntactic structure to the verbal numeral.

Acquisition of the Transcoding System

In a previous study (Power & Dal Martello, 1990), we used a rule-based model in order to explain the errors made by seven-year-old Italian children in the dictation task (i.e. the task of writing down verbal numerals in Arabic digits).

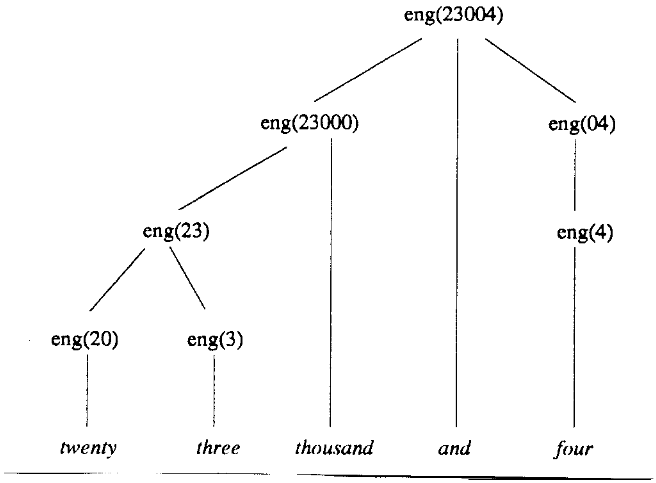

FIG. 1. Derivation of the English numeral for 23004.

Our main finding was that for numbers above 100, many children regularly wrote down *tre cento sessanta cinque* (i.e. *three hundred sixty five*) as 30065; similarly, they wrote down *due mila tre cento* (*two thousand three hundred*) as 2000300. To explain this error, we suggested that the children lacked a correct rule for combining the components of a sum. For instance, when writing down the arabic form of *three hundred sixty five*, they correctly formed the components 300 and 65, but proceeded to combine them wrongly by applying *concatenation* rather than *overwriting*. This finding has been confirmed by a similar study on French children (Seron, Van Lil, & Noel, 1995); Cipolotti, Butterworth, and Warrington (1994) also report on an acalculia patient D.M. who consistently made the same error.

In explaining such errors, we were not addressing the overall architecture of the number-processing system, but the detailed composition of the transcoding rules. In Power and Dal Martello (1990) we assumed, in line with McCloskey's proposed architecture, that the arabic numerals would be generated from an underlying semantic representation; however, our results have no bearing on the issue of whether the transcoding process is semantic or asemantic, and we could have given an analogous explanation using asemantic rules. Table 2 gives a set of rules for direct transcoding of American-English numerals to arabic numerals; we use American rather than British English because the American system is closer to Italian (neither language uses the *and* form).

Table 2 introduces some further notational conventions.

- "ara(X)" denotes the arabic numeral corresponding to the verbal numeral X. For instance, ara(*one*) denotes the arabic numeral for *one*, so we can write ara(*one*) = 1.

TABLE 2
Direct Transcoding from American English to Arabic Numerals

ara(*one*)	=	1
ara(*two*)	=	2
ara(*nineteen*)	=	19
ara(*twenty*)	=	20
ara(*ninety*)	=	90
ara(α *thousand* β)	=	ara(α *thousand*) # ara(β)
ara(α *thousand*)	=	ara(α) & 000
ara(A *hundred* α)	=	ara(A *hundred*) # ara(α)
ara(A *hundred*)	=	ara(A) & 00
ara(*twenty* A)	=	2 & ara(A)
ara(*ninety* A)	=	9 & ara(A)

- In a rule like "ara(*twenty* A) = 2 & ara(A)", the variable A ranges over single words.
- The operation "&" concatenates two strings of arabic digits. Thus "ara(*twenty-three*) = 2 & ara(*three*)" means: to form the arabic numeral for *twenty-three* , concatenate 2 with the arabic numeral for *three*.
- In a rule like "ara(α *thousand*) = ara(α) & 000", the variable α ranges over sequences of words. For instance, when matching ara(*twenty-three thousand*), α would take the value *twenty-three* . The rule says that to form the arabic numeral for *twenty-three thousand*, you concatenate the arabic numeral for *twenty-three* with the digit sequence 000.
- The operation X # Y, where X and Y are sequences of arabic digits, overwrites the trailing zeros of X with the sequence Y. (If X has insufficient trailing zeros, the operation cannot be applied.) For example, 1000 # 3 = 1003; 1000 # 23 = 1023; and 1000 # 234 = 1234. The effect of the operation is thus to form the sum of X and Y.

The advantage of these rule-based models is that they allow us to propose precise models of the transcoding system at different stages of development. (Alternatively, in the case of acalculia patients, they allow us to define the nature of the damage.) Nearly all the children studied in Power and Dal Martello (1990) could dictate reliably for numbers below 100, and also for numbers that were multiples of 100. However, most of them dictated *tre cento sessanta cinque* as 30065. Thus their behaviour exactly fitted the following model:

- All the lexical rules in Table 2 were present.
- The rule ara(*twenty* A) = ara(*twenty*) & ara(A) was present, along with its counterparts for the words *thirty* to *ninety*.
- The rule ara(A *hundred*) = ara(A) & 00 was present.

- For the pattern ara(A *hundred* α), the children were applying the incorrect rule "ara(A *hundred* a) = ara(A *hundred*) & ara(α)": instead of the overwriting operation #, they were using the concatenation operation &. Thus instead of 300 # 65 = 365, they obtained the result 300 & 65 = 30065.

In the present study, our aim is to construct an analogous model of the system that transcodes from arabic numerals to verbal numerals. From the errors made by seven-year-old children, we identify which of the rules in Table 1 have been learned; we also show that the children apply some incorrect rules that might plausibly evolve into the correct ones.

Method

Materials. The stimuli were arabic numerals written on separate cards. They were presented in groups, beginning with single-digit numbers (1, 2, 3, etc.) and progressing in gradual stages to numbers just below a million. The groups were defined according to the length and composition of the stimulus: Group 1 covered the numbers 1–9, Group 2 the numbers 11–99, and so forth. All groups except the first were split into three subgroups (labelled a, b, c). The criteria for this subdivision are shown in Table 3 which also specifies the stimuli in each subgroup.

TABLE 3
Arabic Numerals Used as Stimuli

1	Single-digit numerals
	1 2 3 4 5 6 7 8 9
2a	Two-digit numerals in the range 10–19
	10 11 12 13 14 15 16 17 18 19
2b	Two-digit numerals in the sequence 20, 30, …, 90
	20 30 40 50 60 70 80 90
2c	Other two-digit numerals
	21 33 45 57 69 72 84 96
3a	Three-digit numerals ending in 00
	100 200 300 400 500 600 700 800 900
3b	Other three-digit numerals without internal zero
	111 240 371 495 566 618 727 834 999
3c	Three-digit numerals with internal zero
	101 208 307 404 501 609 702 803 905
4a	Four-digit numerals ending in 000
	1000 2000 3000 4000 5000 6000 7000 8000 9000
4b	Other four-digit numerals without internal zero
	1200 2480 3612 4149 5838 6222 7921 8875 9999
4c	Four-digit numerals with internal zero
	1005 2019 3802 4009 5604 6068 7003 8804 9087

Test materials were also prepared for five- and six-digit numbers (Groups 5 and 6); however, most subjects found these too difficult.

A zero is classified as “internal” if it has at least one non-zero digit both in the string on its left and the string on its right. In the numeral 10200300, for example, the first three zeros are internal and the final two zeros are external. Numerals containing internal zeros were assigned to a separate category since they were expected to pose special problems.

Subjects. The subjects were 9 boys and 6 girls from the second-year class of the Manin elementary school, Padua. The average age of the children was 7 years 3 months.

Procedure. The experiment was carried out during school hours. After being introduced to the class, the experimenter took the children one at a time to an adjoining room where the test was administered. Each subject was presented with a series of arabic numerals, printed on separate cards, and asked to read them out in words. Although general encouragement was provided, the children were not told whether their answers were right or wrong. The test items were presented group by group, starting with Group 1 and ending with 6c (if the subject managed to get that far). The experimenter did not present every item in every group. If the child answered the first three or four items in a group correctly, the experimenter proceeded to the next group: the aim was to discover the zone where the child began to make errors, and to explore that zone thoroughly. When the experimenter judged that the problems were so far beyond the child’s ability that the errors were no longer enlightening, the test was halted. Some of the children wanted to stop before this point; others insisted on continuing to the very end. In either case, the subject’s wishes were respected.

Results

Overall

Overall performance is summarized by Table 4, which reveals the following:

- Responses within each stimulus group were consistent: In most cases subjects either got every item right (R), or got every item wrong (W). Considering only the first three items in each group, the consistent response patterns RRR and WWW were obtained in 132 of the 158 groups attempted (84%), compared with the random expectation of 25% (RRR and WWW are two out of eight possible permutations).
- Nearly all the children had mastered the reading of one- or two-digit numerals, but they began to err in Groups 3–4. Five- or six-digit numerals proved too difficult; most of the subjects did not even attempt them.

TABLE 4
Overall Performance

	Group															
Subject	*1*	*2a*	*2b*	*2c*	*3a*	*3b*	*3c*	*4a*	*4b*	*4c*	*5a*	*5b*	*5c*	*6a*	*6b*	*6c*
Alberto	A	A	A	A	A	A	A	A	A	S	A	A	S	A	S	N
Bruno	A	A	A	A	A	A	A	A	A	A	A	S	S	N	N	N
Clara	A	A	A	A	A	A	A	A	A	N	A	S	S	A	S	N
Nino	A	A	A	A	A	A	A	A	S	S	N	A	S	N	N	N
Daria	A	A	A	A	A	A	A	A	N	S						
Emilio	A	A	A	A	A	S	A	S	N	N	N					
Franco	A	A	A	A	A	S	S	A	N	S						
Guido	A	A	A	A	A	A	N	N	N	N						
Marco	A	A	A	A	A	A	S	N								
Lisa	A	A	A	A	A		S									
Paola	A	A	A	A	N	N	N	N	N	N	N	N				
Silvio	A	A	A	A	N	N	N	S	N							
Teresa	A	A	A	A	S	N	N	N	N							
Rita	A	A	S	S	A											
Vittorio	A	A	S	S												

Note: A: all correct; S: some correct; N: none correct.

Table 5 shows a set of rules for direct transcoding from arabic to Italian numerals. In contrast to Table 1 (the English counterpart), we have given all the lexical rules, so that readers who do not know Italian can understand the children's responses.

Syntactically, English and Italian numerals differ only in a few small respects:

- Italian has single-word numerals for 100 and 1000 (*cento* and *mille*).
- Italian has no counterpart to the British-English conjunction *"and"* in such numerals as *one hundred and two* (*cento due*).
- Italian has a plural form for *thousand* (*mila*).

Group 2 (10–99)

There were no mistakes at all for Groups 1 and 2a, so all the children had mastered the lexical rules from ita(1) to ita(19). Two children were unreliable for Groups 2b and 2c, giving an inconsistent mixture of correct and incorrect responses. For instance, Rita read 50 as *trenta cinque* (thirty five), then proceeded to read 57 correctly as *cinquanta sette*.

TABLE 5
Rules for Direct Transcoding from Arabic to Italian Numerals

ita(0α)	=	ita(α)	ita(20)	=	*venti*
ita(1)	=	*uno*	ita(30)	=	*trenta*
ita(2)	=	*due*	ita(40)	=	*quaranta*
ita(3)	=	*tre*	ita(50)	=	*cinquanta*
ita(4)	=	*quattro*	ita(60)	=	*sessanta*
ita(5)	=	*cinque*	ita(70)	=	*settanta*
ita(6)	=	*sei*	ita(80)	=	*ottanta*
ita(7)	=	*sette*	ita(90)	=	*novanta*
ita(8)	=	*otto*	ita(AB)	=	ita(A0) ita(B)
ita(9)	=	*nove*	ita(100)	=	*cento*
ita(10)	=	*dieci*	ita(A00)	=	ita(A) *cento*
ita(11)	=	*undici*	ita(ABC)	=	ita(A00) ita(BC)
ita(12)	=	*dodici*	ita(100)	=	*mille*
ita(13)	=	*tredici*	ita(A000)	=	ita(A) *mila*
ita(14)	=	*quattordici*	ita(ABCD)	=	ita(A000) ita(BCD)
ita(15)	=	*quindici*	ita(AB000)	=	ita(AB) *mila*
ita(16)	=	*sedici*	ita(ABCDE)	=	ita(AB000) ita(CDE)
ita(17)	=	*diciassette*	ita(ABC000)	=	ita(ABC) *mila*
ita(18)	=	*diciotto*	ita (ABCDEF)	=	ita(ABC000) ita(DEF)
ita(19)	=	*diciannove*			

Group 3 (100–999)

Group 3a. Of the children who were reliable below 100, three made errors for Group 3a (100, 200, …, 900). The error was always the same: The word *cento* was placed first, so that for example 200 was read as *cento due* (hundred two) rather than *due cento*. Eleven such errors were observed, all corresponding to the rule:

ita(A00) = *cento* ita(A)

This finding suggests that when children first encounter the word *cento* (or its equivalent in other languages), they expect it to behave syntactically like *venti*, *trenta*, and the other -*ty* words. In Italian, as in English, all two-word numerals below a hundred have the form BIG SMALL: for instance, *twenty-three* has the form 20 + 3 where 20 is larger than 3. The first Italian numeral to depart from this convention is *due cento* (two hundred), which has the form SMALL BIG.

Considering the results for the whole of Group 3, the three children who failed on Group 3a (Paola, Silvio, Teresa) did not give a single correct response to stimuli in the range 200–999. Out of a total of 50 responses, they placed *cento* at the beginning of the numeral in 18 cases, and in 32 cases left it out altogether.

Group 3b. Seven of the children had mastered Group 3b, and three had given up. The other five subjects made a total of 28 errors (Table 6), which fell into the following categories.

Misplaced "Cento". This refers to the error, already mentioned, of placing *cento* at the beginning of the numeral for a number in the range 200–999. The equivalent error in English would be to say *hundred two* instead of *two hundred*. However, Italian differs from English in that *cento* occurs at the beginning of numerals in the range 100–199: *one hundred and two* translated into Italian is simply *cento due*. If (as seems likely) numerals in this range are learned before numerals in the range 200–999, the preconception that *cento* comes at the beginning of a numeral would be reinforced. It would be interesting to have comparative results for English children on this point.

Consecutive -ty *Words* (or the "Humpty Dumpty" error). This occurs when two words for numbers in Group 2a (20, 30, …, 90) are produced in succession: for instance, *eighty thirty four* as the numeral for 834. Such numerals are always ill-formed. They were produced several times by two children, Franco and Silvio.

Fragmentation. This is the strategy of splitting the arabic numeral into parts and expressing each part separately, a strategy often used by adults in order to read telephone numbers. For instance, Teresa read 495 as *quaranta nove e cinque*, equivalent to *forty nine and five*. This error occurred 37 times throughout Group 3; for 36 of these cases, the first two digits were grouped together, following the rule

$$\text{ita(ABC)} = \text{ita(AB) ita(C)}$$

as opposed to

$$\text{ita(ABC)} = \text{ita(A) ita(BC)}$$

TABLE 6
Errors for Group 3b

Type of Error	*Frequency*	*Example*
Misplaced "cento"	9	371 *cento tre*
Consecutive -*ty* words	8	834 *ottanta trenta quattro*
Fragmentation	6	371 *trenta sette e uno*
Ignoring part of stimulus	4	566 *cinquanta sei*
Misconstruing the stimulus	1	371 *tre cento diciassette*

Ignoring Part of the Stimulus. This means leaving one or more digits unexpressed. For instance, Silvio read 727 as *settanta sette*, equivalent to *seventy seven*, omitting the middle digit. Like fragmentation, this is obviously a convenient strategy for children who lack reliable rules for transcoding arabic numerals with three digits.

Misconstruing the Stimulus. This seems to be a performance error, a slip of the kind that an adult might also make once in a while. The only error of this type in Group 3b was made by Emilio, who gave correct answers to all the other questions in Group 3: he read 371 as *tre cento diciassette*, equivalent to *three hundred and seventeen*, presumably misreading the stimulus as 317.

Group 3c. Six children had mastered Group 3c (three digits with internal zero), and two had given up. The remaining seven subjects made 47 errors, with the distribution shown in Table 7. Most of these errors had exactly the same form: A stimulus of the form A0B was ready by saying the word for A0 followed by the word for B. For instance, 307 was read as *trenta sette* (thirty seven). Although we have classified these as fragmentation errors, another interpretation is possible: they might be due to ignoring part of the stimulus (the middle digit).

Group 4 (1000–9999)

Group 4a. For Group 4a (1000, 2000, …, 9000), six children gave correct answers and six made a total of 23 errors, distributed as shown in Table 8.

Wrong Multiplicand. Two children gave responses of the form *due milioni* (two million) or *due cento* (two hundred). A third child, Guido, gave consistent responses of the form *due cento mila* (two hundred thousand). Initially perplexing, these errors fit into a fairly consistent system that Guido applied to all three-digit and four-digit numerals; we will return to this case in the discussion.

TABLE 7
Errors for Group 3c

Type of Error	*Frequency*	*Example*
Fragmentation	31	609 *sessanta nove*
Wrong multiplicand	9	404 *quattro mila quattro*
Misplaced “cento”	4	404 *cento quaranta quattro*
Misconstruing the stimulus	3	307 *tre cento settanta*

TABLE 8
Errors for Group 4a

Type of Error	*Frequency*	*Example*
Wrong multiplicand	9	2000 *due milioni*
Misplaced multiplicand	7	2000 *mille due*
Misplaced wrong multiplicand	9	2000 *cento due*

Misplaced Multiplicand. This consists, as before, in the placement of the multiplicand word at the beginning of the Italian numeral. All errors of this type had the form *mille due* (thousand two). In every case, the singular form *mille* was used rather than the plural form *mila*. The correct Italian numeral for 2000 is *due mila*, not *due mille*; the response *mille due* is a well-formed numeral meaning 1002.

Misplaced Wrong Multiplicand. All responses in this category had the form *cento due* (hundred two). The two children who made this error (Paola and Teresa) also made the error of misplacing *cento* in Group 3a: in fact, their responses for Groups 3a and 4a were identical.

Group 4b. Only three children had mastered Group 4b. Eight children made a total of 49 errors, shown in Table 9.

Ignoring Part of Stimulus. Children unable to transcode a four-digit numeral often contented themselves with transcoding part of it and ignoring the rest. Daria gave eight responses consistently ignoring the final digit: For instance, she named 3612 as *tre cento sessantuno* (three hundred and sixty-one). Paola went even further, reading 3612 as *sessantuno* (sixty-one), and 8875 as *ottantotto* (eighty-eight), with no regularity on which digits were dropped.

TABLE 9
Errors for Group 4b

Type of Error	*Frequency*	*Example*
Ignoring part of stimulus	24	5838 *cinque mila ottanta tre*
Incorrect digit grouping	9	5838 *cinquantotto mila trentotto*
Fragmentation	8	5838 *cinquantotto trentotto*
Wrong multiplicand	8	5838 *cinque milioni e ottanta tre*

Incorrect Digit Grouping. This error, made consistently by Guido, consisted in assigning the structure (AB)(CD) rather than (A)(BCD). Thus 3612 was read as *trenta sei mila dodici* (thirty-six thousand twelve).

Fragmentation . All these errors were made by Fabio, who consistently split the four digits into two pairs. For instance, he read 7921 as *settanta nove ventuno* (seventy nine twenty-one).

Wrong Multiplicand. All these errors were made by Emilio, who used *milioni* instead of *mila*, and also ignored part of the stimulus by dropping the final digit. He therefore read 3612 as *tre milioni sessantuno* (three million sixty-one); this pattern was maintained regularly through eight responses.

Group 4c. For Group 4c (four digits with internal zero), only Bruno responded correctly; eight children made a total of 40 errors, as shown in Table 10. Although the number of errors is too small to draw firm conclusions, there was evidence suggesting that some configurations of internal zeros pose special problems. In particular, we obtained seven errors in which arabic numerals of the form AB0C were read by the Italian numeral for A0BC. Daria, for example, responded correctly (except for a slip) to three stimuli of the form A00B and three of the form A0BC, but all three stimuli of the form AB0C were misconstrued as A0BC. Thus 5604 was read as *cinque mila sessanta quattro* (five thousand and sixty four), even though Daria had responded correctly in Group 3c, reading 307 as *tre cento sette*, not as *trenta sette*.

Discussion

Almost all the children could reliably transcode arabic numerals for numbers below a hundred: only two subjects made any errors at all. The focus of the study lies in their attempts to read arabic numerals with three or four digits. Only two of the children were able to transcode these reliably; the others made errors that provide evidence on how the transcoding system is acquired.

TABLE 10
Errors for Group 4c

Type of Error	*Frequency*	*Example*
Fragmentation	17	9087 *novanta e ottanta sette*
Wrong multiplicand	7	4009 *quattro milioni e nove*
Misconstruing AB0C as A0BC	7	3802 *tre mila ottanta due*
Incorrect digit grouping	6	5604 *cinquanta sei mila quattro*
Ignoring part of stimulus	3	7003 *settanta tre*

The special feature of the rules for three- and four-digit arabic numerals is that they introduce the multiplicand words *cento* and *mila* (hundred, thousand). The Italian (or English) numeral is formed by splitting the arabic string into parts, transcoding these parts separately, and adding the correct multiplicand word in the correct position. Errors therefore fall into two main classes: Responses that do not express the multiplicand at all, and responses that express it wrongly.

Omitting the Multiplicand Word

Children who did not know the multiplicand words, or were unsure how to use them, adopted two strategies: Either they fragmented the stimulus into parts, each no longer than two digits, or they ignored part of the stimulus.

Fragmentation. This category accounted for 62 of the 197 errors observed (31%). For three-digit arabic numerals we obtained 37 fragmentation errors, of which 36 divided the string ABC into AB and C (rather than A and BC). For four-digit numerals we obtained 25 errors, all of which divided ABCD into AB and CD. (Any other division would, of course, have yielded a three-digit part that would have to have been further divided.) For all 62 errors, the stimulus was fragmented into just two parts, and these parts were expressed in left-to-right order. The term-by-term strategy (Seron & Deloche, 1983), which would transcode 365 (for instance) as *tre sei cinque* (three six five), was never employed.

This regular pattern of results suggests that the children were applying the following strategy:

- If you do not know how to transcode an arabic numeral, split it into parts that you can transcode.
- Fragment the stimulus from left to right, making each fragment as large as possible. For instance, do not bite off only one digit if you can cope with two.
- Transcode each fragment separately, maintaining the same left-to-right order.

Note that this strategy, which fits 61 of the 62 errors observed, always yields the wrong grouping of a three- or four-digit arabic numeral. To transcode a three-digit numeral ABC into Italian (or English), the correct grouping is (A)(BC), not (AB)(C); to transcode a four-digit numeral ABCD, the correct grouping is (A)(BCD), not (AB)(CD).

Partial Transcoding. Considering only the responses with no multiplicand word, we observed 9 errors due to ignoring part of the stimulus. Most of these errors were made by Paola, who responded to arabic numerals of the form ABCD by selecting two digits, usually but not always BC, forming them into a two-digit numeral, and ignoring the rest.

For both these errors—fragmentation and partial transcoding—we believe that the most plausible explanation lies not in specific transcoding rules, but in general strategies governing the application of the rules. In other words, the children erred not because they had incorrect rules for transcoding three- or four-digit arabic numerals, but because they had no rules at all; in consequence, they fell back on general principles that would be applicable to almost any task of language production.

Expressing the Multiplicand Wrongly

Most of the children showed awareness that Italian numerals for numbers in the range 100–9999 should contain *cento*, *mille*, or *mila*; of 197 errors in this range, 123 (62%) contained one of these words (or the higher multiplicand word *milioni*). In employing multiplicand words, four types of error were made, either singly or in combination: selecting the wrong multiplicand word; placing it in the wrong position; dividing the arabic numeral wrongly; and ignoring the final digit.

Wrong Multiplicand. This category was the commonest of the four, accounting for 58 errors, including 9 in which the multiplicand was also wrongly placed, and 8 in which the final digit was ignored.

For three-digit arabic numerals, two mistakes were made: Instead of the correct *cento*, subjects employed either a *-ty* word or *mila* (thousand). In the former case, the "Humpty Dumpty" error resulted: 834 was construed as $(8 \times 10) + (30 + 4)$, with a multiplicand of 10 rather than 100, leading to the response *ottanta trenta quattro* (eighty thirty four). In the latter case, the response was simply *otto mila trenta quattro* (eight thousand and thirty four).

For four-digit arabic numerals, there were three mistakes: Instead of *mila*, subjects employed either *cento*, or *milioni*, or *cento mila*. Thus 2019 was read in three different ways:

Clara	*due cento diaciannove*	two hundred and nineteen
Emilio	*due milioni e diciannove*	two million and nineteen
Guido	*due cento mila diciannove*	two hundred thousand and nineteen

As we have seen, all these errors can be explained by specific rules: The "Humpty Dumpty" error would arise from

ita(ABC) = ita(A0) ita(BC)

Employing *mila* instead of *cento* could arise from either of the following:

ita(ABC) = ita(A000) ita(BC)
ita(ABC) = ita(A) *mila* ita(BC)

Note that in all these cases the rule is nearly correct: The mere addition of an extra zero converts the mal-rule

ita(ABC) = ita(A0) ita(BC)

into the correct rule

ita(ABC) = ita(A00) ita(BC)

Similarly, replacing *mila* by *cento* converts the mal-rule

ita(ABC) = ita(A) *mila* ita(BC)

into the correct rule

ita(ABC) = ita(A) *cento* ita(BC)

We would therefore expect a longitudinal study to show that children who make such errors rapidly learn the correct forms.

Misplacing the Multiplicand. This category accounts for 40 errors, including 9 in which the wrong multiplicand was used (*cento* instead of *mila*). Most of these errors (27) occurred in stimulus Groups 3a and 4a (i.e. for arabic numerals of the form A00 and A000). In every case, the word meaning 100 or 1000 was placed at the beginning of the Italian numeral, as if the subject was applying the following rules:

ita(A00) = *cento* ita(A)
ita(A000) = *mille* ita(A)

Note that the singular form *mille* was always used in this pattern, rather than the plural form *mila*.

When interpreting this error it is important to remember that the words *cento* and *mille* occur at the beginning of Italian numerals in the ranges 100–199 and 1000–1999. For numbers below a hundred, all complex Italian numerals have the form BIG SMALL: for instance, in *venti tre* (twenty three) the word denoting the larger number 20 precedes the word denoting the smaller number 3. On encountering the word *cento* for numerals in the range 100–199, the child is likely to conclude that this new word, like *venti*, *trenta*, etc., occurs at the beginning of a BIG SMALL pattern. This hypothesis is contradicted when the child encounters the numeral 200, which introduces for the first time the pattern SMALL BIG. English-speaking children, by contrast, never encounter the words *hundred* or *thousand* at the beginning of a numeral, so we would expect them to be less prone to this error.

Grouping the Digits Wrongly. We obtained 21 errors in this category, all made by Guido; some of these were combined with the error of using the wrong multiplicand. After responding correctly to all stimuli in Groups 1–3b, Guido began to make systematic grouping errors illustrated by the following responses.

609	*sessanta mila nove*	sixty thousand nine
3612	*trenta sei mila dodici*	thirty six thousand twelve
5604	*cinquanta sei mila quattro*	fifty six thousand four

Of Guido's 24 responses to stimuli in Groups 3c, 4b, and 4c, 20 can be explained by two mal-rules:

ita(A0B) = ita(A0) *mila* ita(B)
ita(ABCD) = ita(AB) *mila* ita(CD)

Both these rules divide the input string according to the fragmentation policy described above. The first rule also leads to the further error of using the wrong multiplicand.

Stages in Acquiring the Transcoding Rules

From the errors that have been described, we can roughly trace the stages by which Italian children learn to transcode from arabic numerals to verbal numerals.

Stage I: Two-digit Numerals Only. Children at this stage know all the words for units, teens, and tens but have not yet learned the word for 100. Our results do not bear on this stage, since only one subject (Vittorio) did not use the word *cento*.

Stage II: Fronting "Cento". Children who have just learned to name 100 by *cento* assume that this word should always be used at the start of a numeral: They regularly transcode 200, 300, etc. by *cento due*, *cento tre* (hundred two, hundred three). At this stage they are perplexed by stimuli in Groups 3b and 3c, which they often transcode by the fragmentation policy, expressing ABC by conjoining AB and C. If they know the word for 1000 (*mille*), they treat it in the same way, transcoding 2000 as *mille due* (thousand two); stimuli in Groups 4b and 4c are usually fragmented, expressing ABCD by conjoining AB and CD.

The policy of coping with complex stimuli by fragmentation is also found in the converse task of transcoding from Italian to arabic numerals (Power & Dal Martello, 1990). In this study, subjects transcoded *tre cento sessanta cinque* (three hundred and sixty-five) by conjoining the arabic numerals for the two fragments *tre cento* and *sessanta cinque*, yielding the error 30065.

Stage III: Using "Cento" as a Multiplicand Word. The next step is to learn the verbal numerals for all the stimuli in Group 3a (200, 300, …, 900), which must be expressed by the new pattern SMALL BIG instead of BIG SMALL. Mastery of this pattern does not mean that the child can cope with Groups 3b and 3c. Two children (Franco, Lisa) continued to use fragmentation for Groups 3b and 3c in spite of responding correctly to 3a.

Stage IV: Experimenting with Digit Grouping and Multiplicand Word. During this stage, which yields the most interesting errors, the child tries various methods of naming the stimuli in Groups 3b–4c by splitting the digit string into parts and introducing a multiplicand. Arabic numerals in Group 4a (2000, 3000, …, 9000) are transcoded by analogy with Group 3a, using the pattern SMALL BIG, but there are frequent errors in the choice of multiplicand word. Stimuli of the form ABC may be mis-grouped as (AB)(C), and the wrong multiplicand may be used: For instance, Guido transcoded 803 as *ottanta mila tre* (eighty thousand and three). Stimuli of the form ABCD may be mis-grouped as (AB)(CD) (Guido transcoded 7921 as *settanta nove mila ventuno*, or seventy-nine thousand and twenty-one); or they may be correctly grouped as (A)(BCD) but with the wrong multiplicand (Emilio transcoded 6068 as *sei milioni e sessantotto*, or six million and sixty-eight).

Stage V: Correct Grouping, Correct Multiplicand Word. After various experiments, the child hits on the correct grouping of (A)(BC), using the multiplicand *cento*, for three-digit stimuli; and of (A)(BCD), using the multiplicand *mila*, for four-digit stimuli.

Through these stages, we can see how the rules for complex numerals might evolve. Consider for example the rule for stimuli in Group 3b. Before learning how to use the word *cento*, children apply a fragmentation policy, equivalent to the rule:

ita(ABC) = ita(AB) ita(C)

This rule persists through Stages I–III, but in Stage IV a multiplicand word is introduced:

ita(ABC) = ita(AB) *cento* ita(C)

By changing the grouping, this evolves into the correct rule:

ita(ABC) = ita(A) *cento* ita(BC)

It would be interesting to observe a group of children, over a period of several months, to see whether the rules do indeed evolve in such a way.

Conclusion

We have investigated how children learn to transcode from arabic to Italian numerals, with emphasis on numbers in the range 100–9999, assuming an asemantic transcoding procedure based on production rules.

Before learning the multiplicand words *cento* and *mila*, children transcode complex arabic numerals either by fragmenting them or, more rarely, by ignoring digits. These are general production strategies that are also found in other contexts (e.g. the converse transcoding task studied by Power & Dal Martello, 1990). Having learned *cento*, Italian children suffer at first from the misconception that it should always be used at the front of the numeral, following the pattern BIG SMALL. Once this hurdle is overcome, they experiment with various ways of subdividing the digit string, separating the parts by a multiplicand word, until they arrive at the correct solutions. This process takes some time because the most natural fragmentation strategy yields the wrong subdivision of the digit string.

Manuscript received 9 April 1996
Revised manuscript received 30 April 1997

REFERENCES

Cipolotti, L. (1995). Multiple routes for reading words, why not numbers? Evidence from a case of arabic numeral dyslexia. *Journal of Experimental Psychology: General*, *12* (3), 313–342.

Cipolotti, L., & Butterworth, B. (1995). Toward a multiroute model of number processing: Impaired transcoding with preserved calculation skills. *Journal of Experimental Psychology: General*, *124*, 375–389.

Cipolotti, L., Butterworth, B., & Warrington, E. (1994). From "one thousand nine hundred and forty five" to 1000945. *Neuropsychologia* , *32* (4), 503–509.

Cohen, L., Dehaene, S., & Verstichel, P. (1994). Number words and number non-words: A case of deep dyslexia extending to arabic numerals. *Brain*, *117*, 267–279.

Deloche, G. & Seron, X. (1982). From one to 1: An analysis of a transcoding process by means of neuropsychological data. *Cognition*, *12*, 119–149.

Deloche, G., & Seron, X. (1987). Numeral transcoding: A general production model. In G. Deloche & X. Seron (Eds.), *Mathematical disabilities: A cognitive neuropsychological perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition*, *4*, 171–196.

McCloskey, M., Sokol, S., & Goodman, R. (1986). Cognitive mechanisms in verbal-number production: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: General*, *115* (4), 307–330.

Power, R., & Dal Martello, M. (1990). The dictation of Italian numerals. *Language and Cognitive Processes*, *5*, 237–254.

Power, R., & Longuet-Higgins, H. (1978). Learning to count: A computational model of language acquisition. *Proceedings of the Royal Society of London*, *B*, *200*, 391–417.

Seron, X., & Deloche, G. (1983). Fom 4 to four: A supplement to "from three to 3". *Brain*, *106*, 735–744.

Seron, X., & Deloche, G. (1984). From 2 to two: An analysis of a transcoding process by means of neuropsychological evidence. *Journal of Psycholinguistic Research*, *13* (3), 215–236.

Seron, X., & Noel, M. (1995). Transcoding numbers from the arabic code to the verbal one or vice versa: How many routes? *Mathematical Cognition*, *1*, 215–243.

Seron, X., Van Lil, M., & Noel, M. (1995). La lecture de numeraux arabes chez des enfants en premiere et en deuxieme annees primaires: Recherche exploratoire. *Archives de Psychologie*, *63*, 269–300.

Welmers, W. (1973). *African language structures*. Berkeley, CA: University of California Press.