Optimizing the Learning Order of Chinese Characters Using a Novel Topological Sort Algorithm

James C. Loach*, Jinzhao Wang

INPAC and Dept. of Physics, Shanghai Jiao Tong University, Shanghai 200240, P. R. China.

* james.loach@sjtu.edu.cn

Abstract

We develop a novel algorithm for optimizing the order in which Chinese characters are learned, one that incorporates the benefits of learning them in order of usage frequency and in order of their hierarchal structural relationships. We show that our work outperforms previously published orderings and algorithms. Our algorithm is applicable to any scheduling task where nodes have intrinsic differences in importance and must be visited in topological order.

Introduction

One of the most fascinating aspects of the Chinese language presents one of the largest barriers to learning it, an irony not lost on generations of students. To the novice, Chinese characters can be inscrutable and their mastery is hard-won thing, rarely achieved until an advanced stage of study.

Becoming functionally literate in Chinese requires memorization of several thousand distinct characters [1], and the effort involved has profound consequences for the learning process. An early focus on learning characters can delay the acquisition of productive language skills, while learning them late can inhibit productive language learning strategies, such as extensive reading, and obscure the logic of the Chinese language. Either way, the consequences for students are a steep learning curve, high rates of attrition, and a certain preoccupation with techniques for learning and remembering characters [2, 3].

The task of learning thousands of distinct symbols is not, however, as difficult as it might appear. There is a logic to the structure of Chinese characters that relates them to one another, and to their meanings and pronunciations [4]. The vast majority of characters are compounds of two parts: a semantic component, indicating the meaning of the character, and a phonetic component, indicating the pronunciation. Components are usually characters in their own right and so, collectively, Chinese characters form a hierarchal network of related symbols. The network rests on a foundation of primitive characters, which typically originate as pictographs and cannot be meaningfully decomposed.

The structure of the character 照 (zhào, to illuminate) is illustrated in Fig. 1. This decomposition illustrates the way in which semantic relationships can be oblique and how phonetic relationships have been distorted during the evolution of the language. Note that only a single meaning is given for each character, even though most actually posses a range of broadly related meanings. A small minority of characters also have multiple pronunciations.

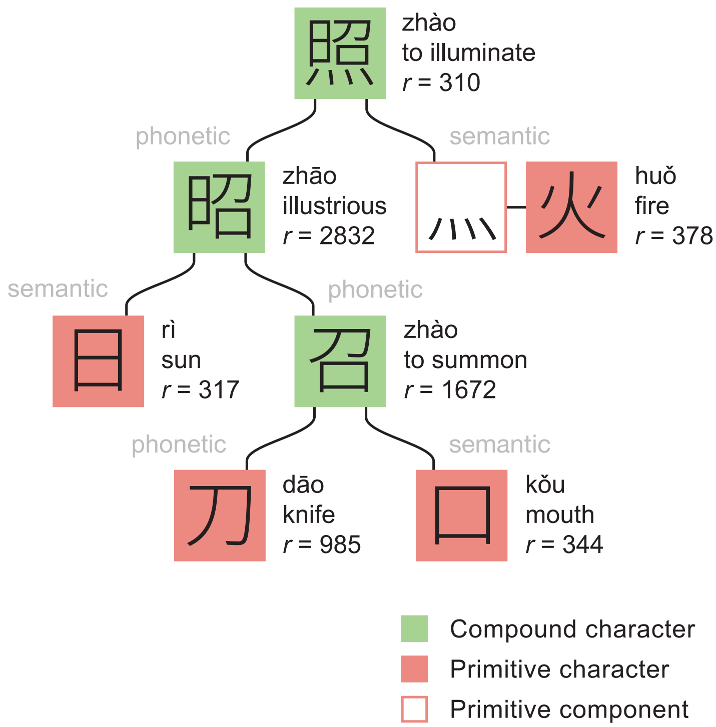

Figure 1. Structural decomposition of the character 照. Primitive characters appear as characters in their own right whereas primitive components do not. The primitive component 灬 is an abbreviated form of the primitive character 火. The parameter r is the SUBTLEX-CH usage frequency rank of the character. Pronunciations are given in pinyin romanization.

The semantic-phonetic structure of Chinese characters makes the learning process somewhat different for native Chinese speakers and second language learners. When Chinese learn to read and write they already know the spoken language and so find phonetic components useful for associating meanings with written forms [5]. For second language learners, who typically begin to learn characters before they know much of the language, this information is harder to use and the learning process is correspondingly more difficult.

But just as learning characters can be more challenging for second language learners it is also particularly useful. The Chinese language abounds in homophones, syllables that have identical pronunciations but different meanings - it has many fewer distinct syllables compared to English and many of these are distinguished only by tone. The consequence is a potential for ambiguity in the spoken language. However, this ambiguity does not translate into written Chinese because homophones are frequently represented by different characters. For example, the character 照 of Fig. 1 is pronounced identically to the unrelated characters 兆 (sign or portent), 罩 (cover) and 棹 (oar). Knowing characters can thus help the learner distinguish between homophones and assign distinct mental identities to the different meanings. This, in turn, can help with remembering words. For example, the verbs 照应 (zhàoyìng, to coordinate) and 照映 (zhàoyìng, to shine) are pronounced identically, but have differences in meaning that are suggested by the final characters 应 (to respond) and

映 (to reflect).

These considerations place the learner in a dilemma with no obvious resolution: knowing characters helps with learning words, knowing words (and therefore pronunciations) helps with learning characters. But something must be learned first.

In a typical Chinese course there is no attempt at a clear resolution. The emphasis of these courses is usually on fast acquisition of communicative ability, and so learning begins using pinyin romanization, a phonetic writing system in the latin alphabet. A mixture of pinyin and characters is used through to the upper intermediate level and, only at this stage, one of moderate conversational ability, is there a concerted effort to drop the romanization. Functional literacy is usually achieved only at an advanced stage.

However, there are approaches to character learning which do attempt a more formal resolution [6, 7]. They typically divide the learning process into two stages: at the beginning there is a rapid, *first-level* learning of the character network, followed by a longer phase of deepening understanding that happens alongside general acquisition of the language. First-level learning means acquiring the ability to write and assign simple meanings to the characters required for basic literacy. It is achieved by assigning meanings to all character components, whether semantic or phonetic, and assembling the character meanings using mnemonic techniques. For example, the character 召 might be assigned a first level meaning of 'to summon', and this could be understood in terms of the assumed-semantic components 刀 (knife) and 口 (mouth), which might vividly represent two methods by which someone might be persuaded to come. Mnemonic techniques are powerful [8], and used in this way allow rapid assimilation of large numbers of characters.

In the deepening phase, the goal is to expand and enrich the first-level network, through a deeper understanding of the relationships between characters (through their phonetic relationships) and of the characters themselves (by learning their pronunciations and subtleties of meaning). It also means embedding the character network in a representation of wider language, by forming connections between characters and multiple-character words. The meanings associated with phonetic components are allowed to fade from memory once their purpose has been served.

The act of assigning meanings to phonetic components might be considered unnatural because it is not etymologically correct. But the deviation from formal correctness is finite, controlled and applies only to the nature of the relationships between characters and not the relationships themselves - the topology of the character network remains accurate. A first-level understanding of 召 is knowing that it means 'to summon' and that it is composed of a 刀 and a 口. This gives the learner the ability to write it, recognize it and to make sense of the words where it appears. Deepening means appreciating its subtleties of meaning, learning its pronunciation (zhào) and knowing that the phonetic 刀 (dāo) can help you remember the pronunciation.

The problem addressed in this paper is how to efficiently learn the first-level network and, in particular, the order in which the characters should be learned. There are two character orderings that make intuitive sense: in order of usage frequency, from high to low, and in order of network hierarchy, starting with primitive characters and building up more complex characters using those that have already been learned as components. However, there is necessarily a tension between these approaches as usage frequency is only weakly correlated with complexity. This behavior can be seen in Fig. 1, where, for example, 照 appears around five times more often than its component 刀. It is also shown more systematically in Fig. 2. Learning characters in order of frequency would often mean learning characters before the components have been learned, while learning them in order of hierarchy would often mean learning rarer characters in advance of more common ones. The aim of this work is to find a character ordering that reconciles the competing demands of frequency and hierarchy, and minimizes the effort required of the learner.

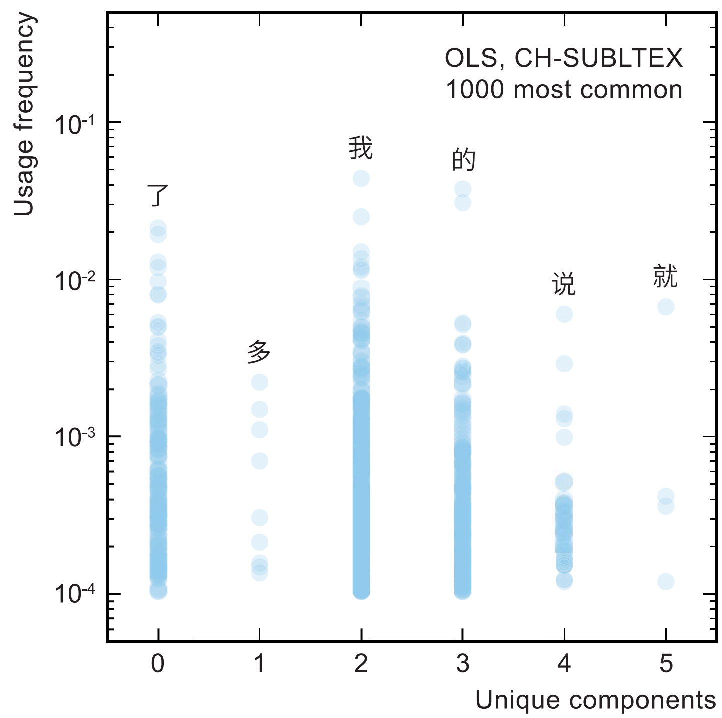

Figure 2. Usage frequency versus number of unique components for the 1000 most common Chinese characters. This plot shows the weak relationship between character usage frequency and complexity, the latter represented by the number of unique components used to construct the character. Usage frequency is normalized to 1.0 over the whole character frequency data set, which encompasses more characters than shown in this plot. The six characters illustrated are the most common in each column. Note that the number of unique components is not the same as visual complexity: the characters 我 and 说 have similar visual complexity (they are composed of similar numbers of strokes) but 我 is conceptually more simple, being, in the OLS character decomposition, composed of two relatively complex primitive components 手 and 戈, compared with the four from which 说 is composed.

This balance is achieved using the conceptual tools of network theory. We conceive of the network of Chinese characters as a directed analytic graph, where the nodes represent characters and the edges represent the structural relationships between them [9]. We devise a measure of node centrality that relates each character's usage frequency to the effort required to learn it, and then order the characters by this measure to provide a first approximation to the optimal learning order. We then sort this list into topological (hierarchal) order using an algorithm designed to minimally disturb the starting order. The final learning order follows the structural hierarchy but is also highly efficient, and we demonstrate quantitatively that it outperforms other published learning orders.

These prior published works fall into two categories: books on learning Chinese characters and a previous academic study by Yan et al. [10]. In books the characters are presented in an order based on the author's pedagogical principles and their particular representation of the character network. The works by Heisig and Richardson [6, 7] are typical and popular examples. However, algorithmically-optimized orders, such as this study and the one of Yan et al., offer the potential for much better performance than those compiled by hand. Yan et al. was the direct inspiration for this work and they, like us, use a network-based approach. In our work we use a different measure of character centrality and a novel ordering algorithm to find a learning order that is more efficient than theirs and also more pedagogical, in that it strictly follows a hierarchal ordering of characters.

Analysis

Overview.

Chinese characters can be represented as a network. Nodes represent characters, with their visual forms, pronunciations and meanings, and edges represent the structural relationships between characters and the nature of those relationships, whether semantic, phonetic or otherwise. Learning Chinese characters means memorizing this network.

The first-level network is a simplified version of the full network, where each node is assigned only the character form and a single representative meaning. Information on the nature of the edges is neglected and all edges are taken to represent semantic relationships. This network has an identical topology to the full network but with reduced information content. It is a network of structurally related visual forms and meanings, a type of network ideally suited to learning through mnemonic techniques. In this study it is the network being learned.

Our aim is to derive a character learning order which maximizes learning efficiency. Such an order maximizes cumulative usage frequency while minimizing the total effort required to learn it. To this end, we assign a usage frequency to each character along with an estimate for the effort required to learn it, its *learning cost*.

Learning costs are calculated using a model that assumes that characters are learned in topological order, that it always makes sense to learn the components of a character before learning the character itself. In Fig. 1 this means, for example, that it makes no sense to learn 召 before 刀 and 口, and would still not make sense even if 刀 and 口 had zero usage frequency and 召 was the most common character in the language.

We incorporate usage frequency and learning cost into a measure of character centrality. This measure indicates their relative importance of the character to the learner, prioritizing frequency and penalizing learning cost. Ordering characters by this centrality provides a first approximation to the final learning order. It is approximate because ordering by centrality does not imply ordering by topology, which must be imposed in a separate step.

Topology is imposed using an algorithm designed to topologically sort our centrality-ordered list in a way that minimally disturbs it. Higher-centrality characters are learned first only when allowed topologically.

The algorithm includes a feature that allows certain groups of characters to be fixed together into consecutive positions in the learning order, regardless of their differences in centrality. This allows the order to capture logical relationships between characters that are not encoded in the network.

The algorithm has potential applications beyond the learning of Chinese characters, and might be applied to any scheduling task where nodes have intrinsic differences in importance and must be visited in topological order.

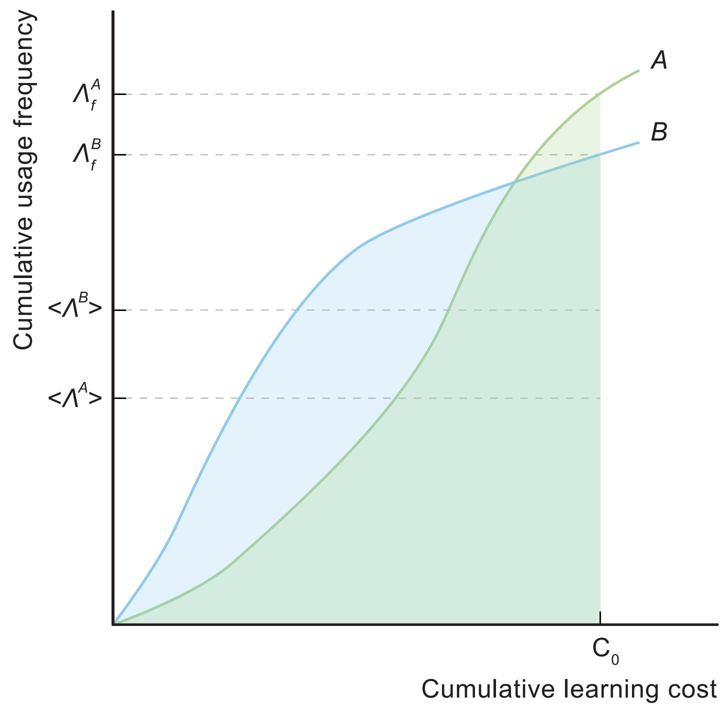

Figure 3. Measures of learning efficiency. The curves A and B represent two different learning curves. For each curve, the final learning efficiency Λ_f is the cumulative usage frequency for a specific cumulative learning cost C_0, and the integral learning efficiency $\langle\Lambda\rangle$ is the average cumulative usage frequency between the origin and C_0. Curve A has higher Λ_f but lower $\langle\Lambda\rangle$. Values for $\langle\Lambda\rangle$ are illustrative only.

Learning efficiency.

A typical learning scenario is characterized by a fixed available effort, with which the learner seeks to acquire the maximum cumulative usage frequency as rapidly as possible. The learning process can be visualized as a *learning curve* in a space defined by axes of cumulative usage frequency and cumulative learning cost. This is illustrated in Fig. 3. Efficient learning curves rise quickly and reach high end-points.

Learning curves for different learning orders can be compared visually, as in the figure, but it is convenient to parameterize them in a way that captures their most important features. We propose a two-parameter scheme.

The first parameter is the cumulative usage frequency at the end of the learning process, once the maximum cumulative learning cost C_0 is reached. We call this parameter the *final learning efficiency* Λ_f. High Λ_f is characteristic of efficient learning curves. Note that when comparing curves it is necessary to use the same usage frequencies for both curves, even though both curves may not cover identical sets of characters. In this work we normalize the entire usage frequency data set to 1.0, so that the goal of all learning curves is to approach as close as possible to a maximum Λ_f of 1.0.

The second parameter concerns how the maximum cumulative usage frequency is approached. Consider the two curves shown in Fig. 3. The curve A has higher Λ_f but, over much of the learning process, actually performs less well than curve B. Curve A prioritizes longer-term cumulative frequency at the cost of the shorter-term. The difference would be immaterial if the learning process had a short extent in time but this is not typically the case. Learning may take place over many months, during which the learner would likely be exposed to other parts of the language. In this case it is advantageous to have a learning curve that rises quickly, even at the cost of some longer-term cumulative usage frequency. We parameterize this using the average cumulative usage frequency, which we call the *integral learning efficiency* $\langle\Lambda\rangle$ and calculate using

$$\langle\Lambda\rangle = \frac{1}{C_0}\int_0^{C_0} f(C)\,\mathrm{d}C \tag{1}$$

where is the f is cumulative usage frequency and C is the cumulative learning cost.

Centrality.

We define the centrality η_i of character i to be

$$\eta_i = \frac{f_i}{C_i}, \tag{2}$$

where f_i is the usage frequency and C_i is the learning cost. This quantity is the ratio of the benefit and cost that each character represents to the learner. Learning characters in order of η will therefore tend to satisfy the prime concern of the learner, of maximizing cumulative usage frequency and minimizing effort. These learning curves will rise steeply and have high end points; or, in the language of the previous section, be characterized by high integral and final learning efficiencies.

Values for f_i can be extracted from corpora of written Chinese. Values for C_i are more difficult to assign objectively and so are estimated using a learning model. We begin by assigning a common learning cost α to all primitives so that, for example, the characters 口 and 豖 are considered equally difficult to learn. Furthermore we assume that the difficulty in learning compound characters is proportional to the number of combinations they contain, and call the cost per combination β. Thus, the cost of learning 的 would be $1 \times \beta$, because it is a compound of two components 白 and 勺, and the cost of 品 would be $2 \times \beta$, because it is a compound of three. In this work we take $\alpha = \beta = 1$. This would give all characters in Fig. 1 a learning cost of 1.

The final order is independent of the ratio α/β but it does depend on the assumptions that α is the same for all primitives and β is the same for all combinations. While the latter assumption is likely to be fairly robust, the one relating to α is more problematic, given that some primitives are clearly more difficult to learn than others. The true learning cost of a primitive would depend on, amongst other things, the learner's familiarity with the strokes of Chinese characters, their knowledge of similar primitives, and on the visual similarity between the primitive and the thing it represents. In the absence of experimental data, one simple improvement to the calculation of α might be to have it depend on the number of strokes used to write the character. This number is well-defined and provides a good measure of visual complexity.

Our learning cost model assumes that characters are learned in hierarchal order. When we calculate the cost of learning a compound character we do not include the cost of learning the components themselves, which have already been learned. Furthermore, our model implies that the total cost of learning any set of characters, provided the order is hierarchal, is identical. Orderings may differ only in the their integral learning efficiencies.

We take the character structure network to be fixed and do not consider how the decompositions might be altered in order to reduce the learning cost. While such a procedure is possible, it would require a more sophisticated learning model to be meaningful. We do study variant networks, but consider the choice between them to be outside the scope of this work.

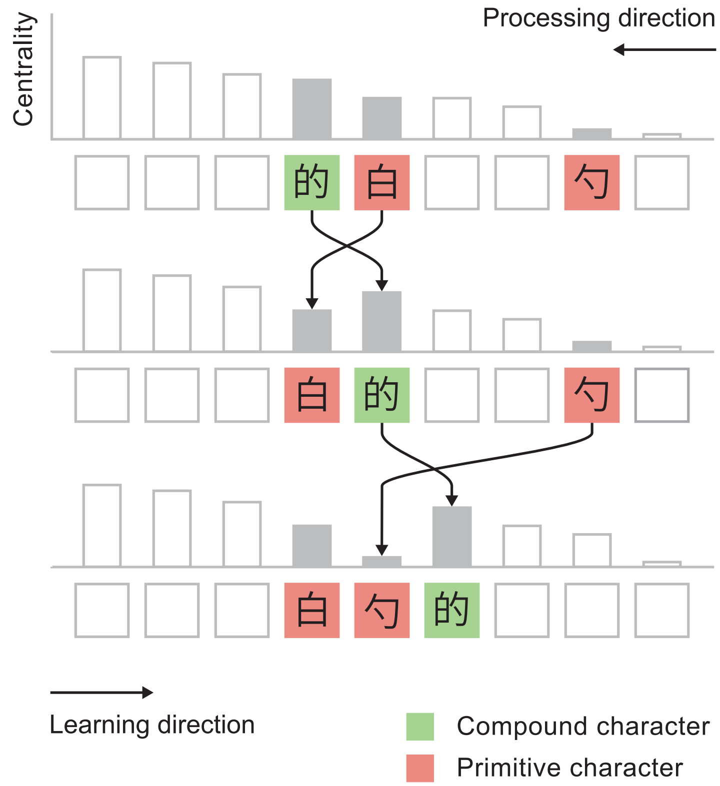

Figure 4. Illustration of the topological sorting algorithm. The ordered list is processed from low to high centrality (right to left in the figure). Once 的 is reached, its components are checked in turn. 白 is found to lie to the right of 的 and so is repositioned to its left. Likewise 勺 is found to the right of 的 and is similarly repositioned. 勺 is positioned to the right of 白 because it has lower centrality. The centralities used in this figure are for illustrative purposes only.

Topological sort.

Learning characters in order of centrality prioritizes characters that are useful and easy to learn but it does not ensure that characters are ordered according to the character hierarchy. For example, the simple and common 的 will appear a long way in advance of its components 白 and 勺, which appear much less frequently as characters. This is a problem because our measure of learning cost assumes that the components are already known. It also makes little sense: the character 的 cannot be learned before the forms of 白 and 勺 are in some sense committed to memory, and the extra effort to fully learn them as characters is minimal.

We resolve this issue with a sorting algorithm that modifies the centrality-order list to ensure that all characters appear before those in which they act as components. The algorithm is illustrated in Fig. 4 and may be described as follows:

1. Process characters from low to high centrality (right to left in the figure). This is opposite to the order in which characters are learned.

2. Decompose each character into a list of its primitives and all intermediate characters. For example, 照 should be decomposed into 昭, 召, 日, 刀, 口 and 灬.

3. Determine the position of each component in the centrality-ordered list. If the position is to the left of the character then no action is taken. If it is to the right of the character then it should be moved the character's left. Move the character as far left as it will go, while still remaining to the right of all characters with higher centrality.

This procedure ensures that characters are relocated only when necessary and always to the optimum position within the region allowed by the hierarchy. This results in a highly optimized ordering. However, it is not necessary the most efficient order possible and we can find special cases, such as the network in Fig. 5, where the algorithm generates a sub-optimal ordering. We have not found such instances in the real character network but cannot prove that they do not exist.

The algorithm is adaptive, in that it may be applied to a partially learned network. And it can also be used in other contexts, for any scheduling task where nodes have intrinsic differences in importance and must be visited in topological order.

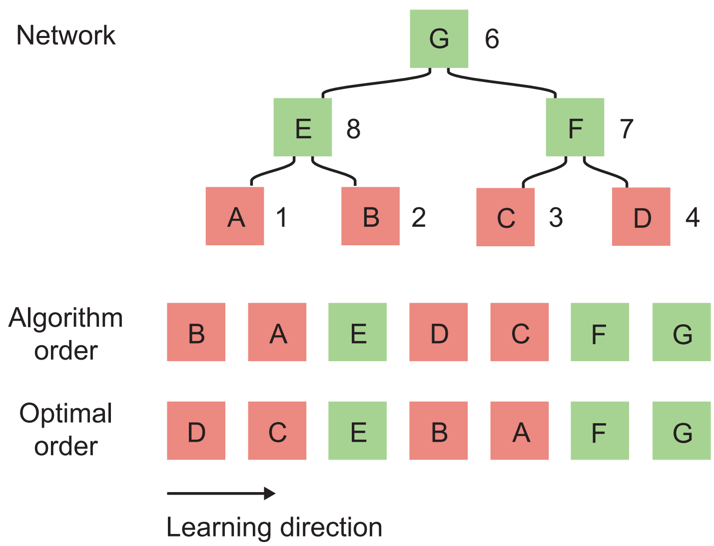

Figure 5. A network where our algorithm does not return the optimal character ordering. A hypothetical network where the integral learning efficiency of the order generated by the algorithm is lower than another possible order. Letters represent Chinese characters (for example, E is a compound character formed from primitives A and B) and the numbers are centralities. Both orders have identical final learning efficiencies.

Character groups.

In certain situations it may be pedagogically beneficial to fix the relative positions of certain characters in the learning order. This can be used to ensure that variant primitive forms are learned together (such as 火 and 灬, both meaning fire), as well as characters which have strong semantic relationships to one another (such as 上 and 下, meaning up and down). We call the former *variant primitives* and the latter *semantic groups*.

These groups can be incorporated into our algorithm by treating them as *meta-characters*. A meta-character has usage frequency and learning cost equal to the sum of the characters from which it's composed, and everywhere in the network replaces each of those characters. In this way, a meta-character 'fire' could represent 火 and 灬 and be identified as a component of characters such as 照 and 秋 (qiū, autumn). Once the ordering algorithm has been applied to this modified network, the meta-character is replaced in the final order by its component characters, ordered according to centrality.

In this work we implement six common variant primitives, primarily for illustrative purposes: 人 and 亻 (person); 水 and 氵 (water); 火 and 灬 (fire); 刀 and 刂 (knife); 手 and 扌 (hand); and 心 and 忄 (heart). We do not implement any semantic groups.

Source data.

We use two different representations of the simplified Chinese character network, one compiled with an emphasis on etymological correctness and one with an emphasis on the visual relationships between characters. The former is from a forthcoming dictionary by *Outlier Linguistic Solutions* [12] and the latter is taken from the books *Remembering Simplified Hanzi 1 and 2* by Heisig and Richardson [6, 7]. We refer to these networks as the *OLS* and *HR* networks, respectively. The networks have similar coverage: OLS covers 3495 characters and primitive components, and HR covers 3253, with 2990 in common between them. The majority of the decompositions are identical and the majority of the differences originate from decisions regarding encoding; many components do not have Unicode code points and others can reasonably be represented by more than one code point.

Usage frequency data is taken from the SUBTLEX-CH database [11], which is derived from Chinese movie and television subtitles. We chose this database because it is comprehensive and is representative of modern colloquial Chinese. In any practical application of our algorithm frequency data should be chosen with the specific goals of the learner in mind. The SUBTLEX database contains 5938 characters with frequencies calculated from a total corpus of 46841097 characters. All character frequencies used in this study are normalized to the whole SUBTLEX-CH database. Normalizing in this way, both the Outlier and Heisig networks have cumulative usage frequencies of 0.992.

It is instructive to consider scenarios where the learner aims to learn specific lists of characters rather than just acquiring cumulative usage frequency, especially when these lists are small compared to the size of the character network. We obtained character lists from the HSK Chinese Proficiency Test administered by the Chinese National Office for Teaching Chinese as a Foreign Language (NOTCFL) [13]. This exam covers six levels, the highest of which corresponds to upper-intermediate proficiency. The cumulative numbers of characters covered by each level start at 174 for the first level, increasing to 2663 by the end of the sixth. In this study we use cumulative character lists up to the 3rd, 4th, 5th and 6th levels, which we refer to as HSK13, HSK14, HSK15 and HSK16, respectively.

Results and discussion

Fig. 6 shows the first 85 characters from the learning order derived using the OLS network. Together, these characters have a cumulative usage frequency of 0.43. The full learning curve for the OLS network is shown in Fig. 7, and is compared with the algorithm of Yan et al. applied to the same data and to the fixed character order of Heisig and Richardson using the HR network. Learning efficiencies are presented in

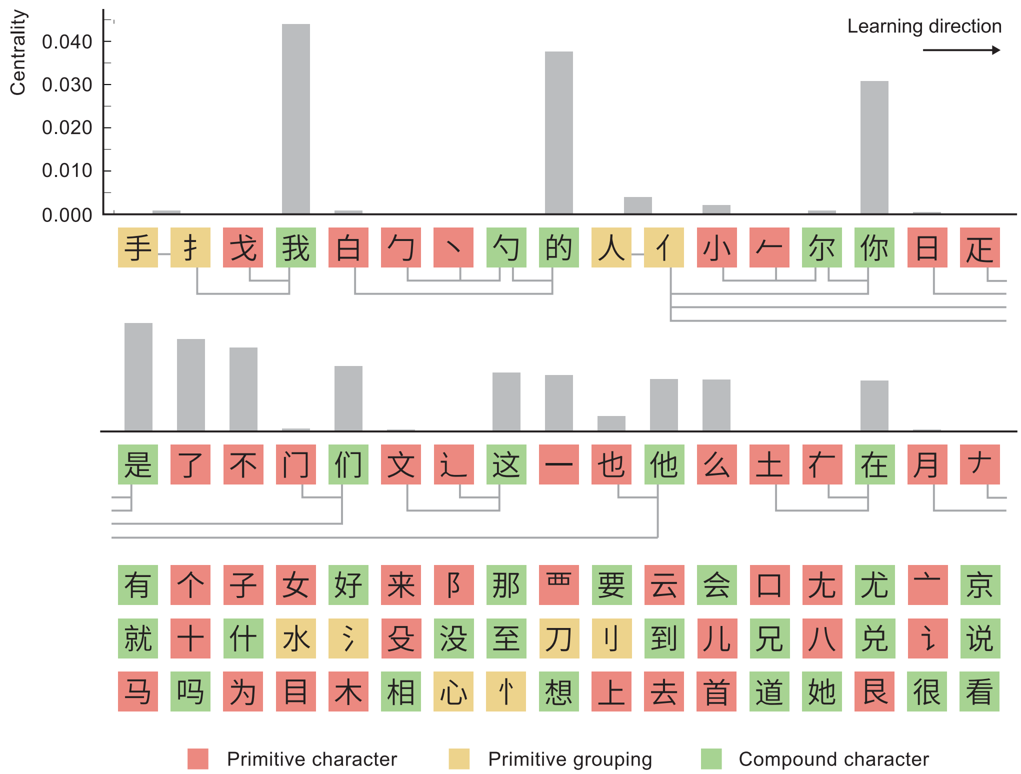

Figure 6. The first 85 characters of our optimized learning order. Taken together these characters have a cumulative usage frequency of 0.43.

Table 1. Our algorithm outperforms the algorithm of Yan et al. and the order of Heisig and Richardson on all measures. Fig. 8 shows the consequences for the learning curve when the target list of characters is only a subset of the full network. The use of target lists maximizes the efficiency for the list at a cost of overall cumulative usage frequency.

The shape of the Heisig and Richardson curve in Fig. 7 can be understood from the structure of their book. The first half of the curve, between the origin and the large discontinuity, covers their first volume, in which they introduce the bulk of the primitive components. These are grouped according to meaning and each one is followed by all the high-usage compound characters that can be made at that point. This explains the alternating pattern of sharp upwards jumps and gentler slopes. The second volume introduces the lower-frequency compounds that are not included in the first. The authors aim for a relatively high Λ_f by the end of the first volume but with no particular regard for the shape of the curve. Note that their curve in Fig. 7 was calculated using SUBTLEX-CH frequency data, which may differ from the frequency data which they used to select characters and order them.

The comparison with the algorithm of Yan et al. uses identical source data. However, in order to make the comparison between algorithms meaningful we had to make one alteration to their learning model relating to the normalization of the curves: the parameter equivalent to our β, which they take as the number of components rather than the number of combinations of components, was made identical with our definition of β. In these like-for-like conditions our algorithm gives better Λ_f and $\langle\Lambda\rangle$ efficiencies. It also results in a more logical ordering in that it guarantees that characters are learned after their components and also tends to cluster components directly before characters where they are used. These behaviors can be observed in Fig. 6.

In the paper of Yan et al. the problem of ordering characters is understood as the problem of finding a compromise between two requirements that are taken to be intrinsically in contradiction: of hierarchal order and of usage frequency order. This results in a compromise learning order that follows neither. The reason our algorithm performs better is that we conceive of the problem differently: we reject the fundamental contradiction and, instead, frame the problem as one of selecting the most efficient hierarchal order, the hierarchal order that corresponds most closely to the usage frequency order. Our decision to enforce the hierarchy eliminates the duplication of effort when characters are learned before their components but it does not, in itself, guarantee a higher overall efficiency. That it does lead to such an improvement is the main outcome of this work. We have shown that there need not be any contradiction between the pedagogical desirability of learning characters in hierarchal order and the goal of maximizing accumulated usage frequency. We have shown that the properties of the Chinese character network allow highly efficient hierarchal learning orders, and have developed an algorithm that allows these orders to be extracted.

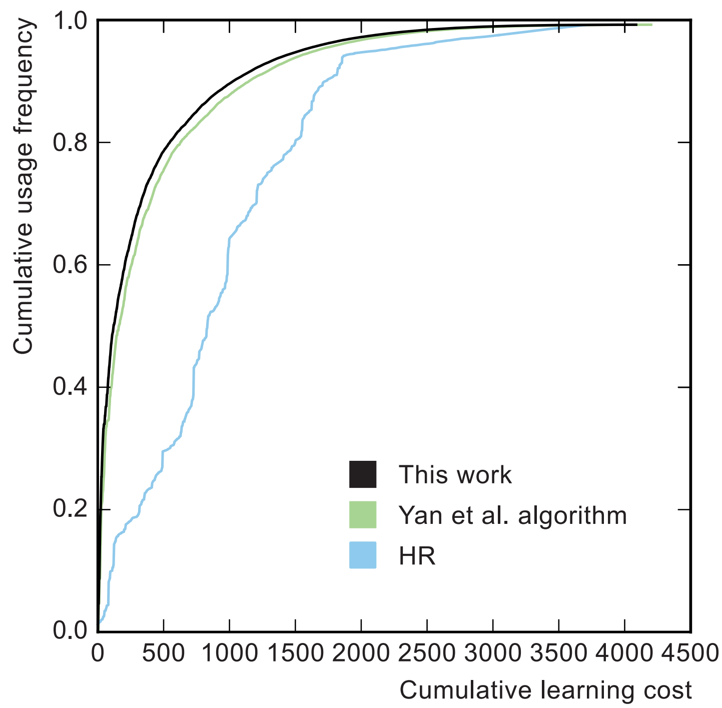

Figure 7. Learning curves. The black and green curves were created using the OLS character decompositions and the two different learning order algorithms. They both included the six sets of primitive variants described in the text, implemented as meta-characters. The Yan et al. algorithm was optimized up to a cumulative learning cost of 4000. The blue curve uses the HR network and a fixed character ordering. Learning efficiencies are presented in Table 1.

Acknowledgments

This work was supported by the Shanghai Key Lab for Particle Physics and Cosmology (SKLPPC), Grant No. 15DZ2272100. We acknowledge the generosity of Outlier

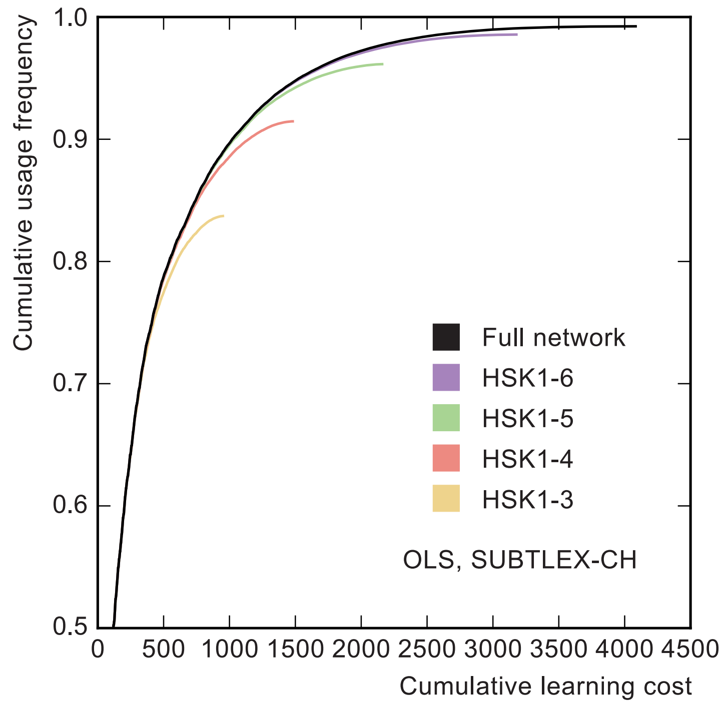

Figure 8. The effect of target character lists. The black curve is identical to the black curve of Fig. 7. The colored curves illustrate the effect of imposing target character lists that are subsets of the characters in network. The target list for each curve was taken from a particular set of HSK exams and augmented by any extra characters needed as sub-components. These curves represent efficient learning of the target list but at the cost of overall cumulative frequency.

Table 1. Learning curve parameters. The number of characters learned N, final learning efficiency Λ_f, and integral learning efficiency $\langle\Lambda\rangle$ for reference learning costs of $C_0 = 500$ and $C_0 = 1500$. The Yan et al. algorithm was optimized up to the relevant C_0.

Curve	**Network**	**Algorithm**	$C_0 = 500$			$C_0 = 1500$		
			N	Λ_f	$\langle\Lambda\rangle$	N	Λ_f	$\langle\Lambda\rangle$
This work	OLS	This work	491	0.788	0.595	1450	0.948	0.789
Yan et al.	OLS	Yan et al.	438	0.755	0.550	1309	0.936	0.761
HR	HR	N/A	457	0.296	0.166	1363	0.803	0.439

Linguistic Solutions in providing us part of their dictionary prior to publication.

References

1. Kane D. The Chinese Language: It's History and Current Usage. Tuttle Publishing; 2006.

2. Lam HC. A Critical Analysis of the Various Ways of Teaching Chinese Characters. Electronic Journal of Foreign Language Teaching. 2011; 8: 57-70.

3. Tse SK, Cheung WM. Chinese and the Learning of Chinese. In: Marton F, Tse SK, Cheung WM, editors. On the Learning of Chinese. Sense Publishers; 2010. pp. 1-7.

4. Lam HC. Orthographic Awareness. In: Marton F, Tse SK, Cheung WM, editors. On the Learning of Chinese. Sense Publishers; 2010. pp. 53-73.

5. Tse SK, Marton F, Ki WW, Loh EKY. Learning Characters. In: Marton F, Tse SK, Cheung WM, editors. On the Learning of Chinese. Sense Publishers; 2010. pp 75-102.

6. Heisig JW, Richardson TW. Remembering Simplified Hanzi: Book 1, How Not to Forget the Meaning and Writing of Chinese Characters. 1st ed. University of Hawaii Press; 2008.

7. Heisig JW, Richardson TW. Remembering Simplified Hanzi: Book 2, How Not to Forget the Meaning and Writing of Chinese Characters. 1st ed. University of Hawaii Press; 2012.

8. Hulstijin J. Mnemonic methods in foreign language Vocabulary Learning: Theoretical Considerations and Pedagogical Implications. In: Coady J, Huckin T, editors. Second Language Vocabulary Acquisition. Cambridge University Press; 1997. pp 203-224.

9. Li J, Zhou J. Chinese character structure analysis based on complex networks. Physica A. 2007; 380. doi:10.1016/j.physa.2007.02.059

10. Yan X, Fan Y, Di Z, Havlin S, Wu J. Efficient Learning Strategy of Chinese Characters Based on Network Approach. PLoS ONE. 2013;8: e69745. doi:10.1371/journal.pone.0069745

11. Cai Q, Brysbaert M. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. PLoS ONE. 2010;5: e10729. doi: 10.1371/journal.pone.0010729

12. Henson A, Renfroe J, Schmidt C. Outlier Dictionary. Outlier Linguistics Solutions USA, LLC; 2015.

13. HSK Chinese Proficiency Test. Available: http://www.chinesetest.cn. Accessed: 16 January 2016.