Register and Representation of Shanghai Chinese

Shanghainese (Shanghai Chinese) is a variety of Wu Chinese spoken in Shanghai. It has long served as an important case for illustrating the interaction between tone and segmental features. Despite being relatively well studied, Shanghainese still presents a number of long-standing, unresolved debates concerning the behavior and representation of its tones. One of the central debates concerns the representation of register. **?**. Murmur is a phonetic manifestation of register: in addition to the usual pitch lowering, it is accompanied by breathy voice. Previous research differs on where register should be located within tonal geometry. Two competing views exist: one treats register as a feature that dominates at the word level, while the other treats it as a feature associated with the voicing node of a segment.

This paper is motivated by that question and has two goals. First, we provide a more rigorous mathematical framework to model these two approaches, allowing for a precise comparison of their predictive power and expressiveness. Second, we employ computational methods to learn the tonotactics of Shanghainese under both representational assumptions, comparing the learning results to evaluate their relative outcomes.

1 Shanghai Tonology

Shanghai Chinese has five tones (T1–T5). Previous research has shown a variety of transcription systems for these tones, whether using the five-point scale notation (Chao, 1930) or more phonologically oriented representations indicating tonal height (**?**). These systems largely converge on the overall contour shapes but differ in the precise pitch value of tonal onsets and offsets.

Tone	CUHK (2018)	Xu and Tang (1988)	Duanmu (1999)	?	Zee and Maddieson (1979)
T1	53	53	/HL/	/H/	HL
T2	34	24	/LH/	/L/	/MM/
T3	23	13	/LH/	/L/	/ML/
T4	5	<u>55</u>	/LH/	/L/	/H/
T5	12	<u>13</u>	/LH/	/L/	/LM/

Table 1: Transcriptions of Shanghai tones across sources

The Shanghai tonal system is regarded as a *complex* system. Unlike Mandarin, in which tonal contrasts are primarily distinguished by pitch, tonal recognition in Shanghai Chinese also relies on *register* and vowel *length* (Zhu and Wang, 2015). As shown in Figure 2, there are two registers [UPPER] (discussed in detail in the next section), and two major syllable types: slack syllables and checked syllables (those closed by a glottal stop). The following section will focus on previous studies of register and length contrasts in Shanghai tonology.

Register	Tone	Slack syllable			Checked syllable		
		TV(ŋ)	**DV(ŋ)**	**SV(ŋ)**	**TVʔ**	**DVʔ**	**SVʔ**
+upper	**T1**	pa "father"	×	ma "mother"	×	×	×
	T2	pa "dam"	×	me "pretty"	×	×	×
	T4	×	×	×	paʔ "eight"	×	aʔ "duck"
-upper	**T3**	×	ba "climb"	ma "horse"	×	×	×
	T5	×	×	×	×	baʔ "white"	maʔ "pulse"

Table 2: Co-occurrence restrictions on tones with consonants (Chen and Gussenhoven, 2015).

1.1 Register

Register is the primary distinction in Shanghai tones. Although it is often associated with overall pitch shift cross-linguistically, in Shanghainese, a lower register is also accompanied by breathy phonation. **?** classifies Shanghai syllables into murmured and clear (or plain) types, with murmured syllables corresponding to those that co-occur with the [−upper] register. Murmur can thus be understood as the phonetic realization of Shanghai register (Zhu, 1999). The terminology for this feature varies across researchers, including [slack] (Duanmu, 1999), [+voice] (Ren, ???), and [+slack vocal cords] (Keating, as cited in Yip 1993). For consistency, this paper will use the term register, with the understanding that it refers not only to pitch lowering but also to breathy phonation.

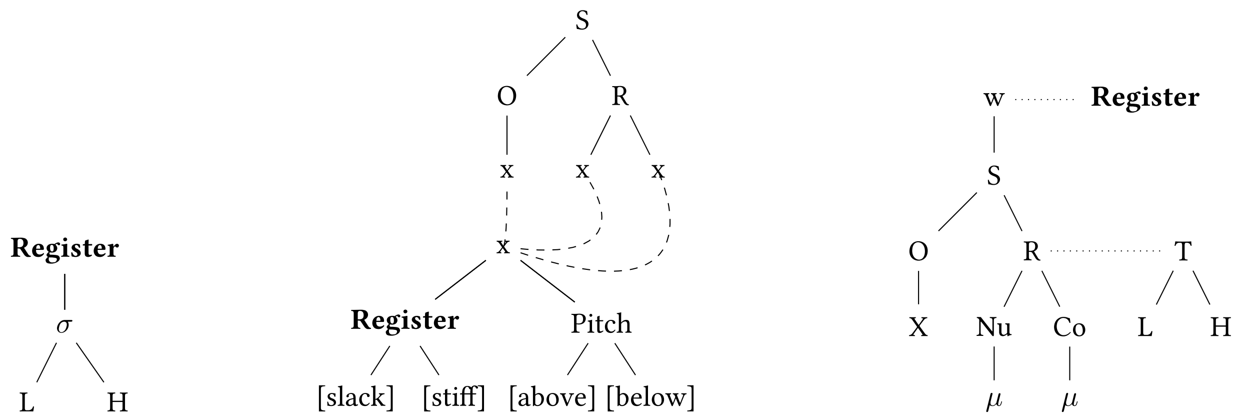
Figure 2 shows that all syllables with voiced obstruents onsets fall neatly under the [−upper] register. This is why register is often regarded as a feature of voicing. For syllables beginning with sonorants, however, both registers are possible and contrast with each other. Two points should be noted in the phonetic realization of register. First, in monosyllabic words, the [−upper] register, even when restricted to voiced obstruents, causes devoicing of these segments (e.g., [d] → [d̥]), nearly neutralizing the VOT distinction between voiced and voiceless unaspirated stops. This distinction is retained when the syllable occurs in a non-initial position. Second, in polysyllabic words, only the register of the initial syllable is realized on the surface, and will also dominate the whole phrase and change the voicing of the following consonants.

give some examples

1.2 Representation of Register

Since register is a major distinction in Shanghai tonology, many models of Shanghai tones have been proposed within feature geometry frameworks (Clements, 1985; Sagey, 1986). The main debate concerns where register should be located. The proposals are listed below, and their corresponding feature geometry representations are sketched beneath.

1. A tonal feature (Yip, 1980), or perhaps shared by laryngeal (Yip, 1993).
2. A feature of the *onset*, conditioned by [+slack] (Duanmu, based on Bao (1999)).
3. A feature of the *initial consonant* (Ren, 1992) (to read).
4. A property of the *phonological word* (Zhu, 1999).



Yip (1980, 1993) propose that the register should be a feature under the tonal node, and also dominate the pitch, as indicated in Figure ??. Note in Yip's model, the register dominate the pitch nodes and apply to each syllable. In polysyllabic forms, only the register of the inital syllable are retained and the rest are delated. The shortcoming is that it does not reflect why [+upper] should condition onset voicing.

Different from Yip (1980, 1993), Duanmu (1988) proposed that pitch and register are represented on two separate planes: one encodes the stiff/slack distinction in voice quality, while the other contains tonal features. The motivation is that register is consistently reflected in voice quality but not always in pitch. A major difference between Duanmu's proposal and others is that the node represents an X-slot (either a consonant or a vowel), and each X-slot is specified for both register and pitch. Consequently, whereas a tone can be represented as a single feature bundle in Yip's system, two sets of features are required in Duanmu's model. For example, [55] is represented as [+u, +H] in Yip's model but as [+st, +H][+st, +H] in Duanmu's.

Another candidate proposed by Zhu (1999) places register at the top of the phonological word, while the pitch node is linked to the rime. This differs from Yip's model in that, whereas Yip connects register to each syllable, Zhu's model assigns a single register node dominating the entire word. This accounts for the fact that in polysyllabic words only one register value is retained. Unlike the other two models, Zhu's analysis does not associate register with voicing at all.

Another issue in the feature geometry of Shanghainese concerns the representation of medial and coda positions. Beyond the placement of register, scholars also debate how the syllable should be internally structured. Duanmu proposes that all Chinese dialects share the same underlying syllable structure. Rather than mora theory, he employs X-slots, positing three slots under the syllable node: one for the onset and two for the rime. Zhu, however, diverges from this proposal, offering a different analysis of syllable structure in Shanghainese.

2 Proposal

There are four goals of this paper. First, I compare two competing theories from Duanmu (1988) and Zhu (1999) from a learning perspective. These two models are chosen because their underlying syllabic structures are both comparable and explicitly formulated, and because register and pitch nodes are embedded at different levels of structure in each model. Moreover, the models are distinct both representationally and mathematically. In contrast, Oakden (2020) argues that the long-standing competition between Bao's and Yip's models are notationally equivalent and translatable into each other without loss of contrast. This distinction will be discussed further in Section XXX.

Each model is represented in model-theoretic terms so that relations such as association, succession, and dominance are rigorously encoded. Learning is carried out using the Bottom-Up Factorial Inference Algorithm (BUFIA; Chandlee et al. (2019)), which operates over model-theoretic inputs based on the two representations.

Our learning goals have two components. First, we examine tonotactics: the co-occurrence patterns between tones and features (register or voicing). We run BUFIA on both representational models, testing them on monosyllabic and disyllabic forms, to evaluate which model more effectively captures these patterns, particularly the agreement between voicing and tone.

Register/voicing assimilation: the rules governing disyllabic compounds. The central question is which representational hypothesis best captures the relevant generalizations in both tonotactics and register assimilation. Conversely, if a representation fails to encode the appropriate dependencies, learning over such a representation will either over generate, under generate, or yield predictions that are empirically unattested.

2.1 Duanmu

Duanmu's model employs the x-slot as the basic unit, proposing that most Chinese dialects can conform to this template. In his analysis, glides are instead treated as secondary features of the onset (which takes one slot); the rime of a citation monosyllable occupies two slots when the vowel is long, or one slot when the syllable is closed by a glottal or nasal coda. Each slot can be specified for VOICING/REGISTER and PITCH values. To avoid over-generate, Duanmu proposes the following universal constraints: (1) only rime may be assigned pitch; (2) each syllable bears a single register specification.

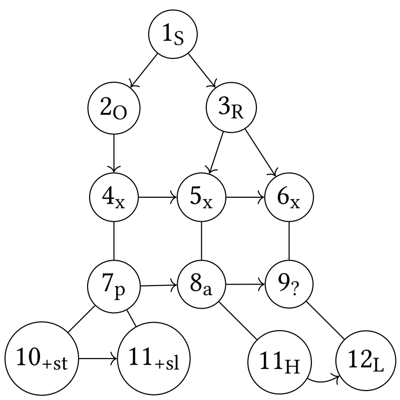
This model can be formalized in a model-theoretic framework, in which a linguistic object of finite size is represented as a mathematical structure. Formally, a model consists of two components: a domain $\mathcal{D}$, which specifies the set of elements that constitute the object (e.g., tones, moras, syllables), and a set of relations over that domain, denoted as $\mathcal{R} = \{R_1, R_2, \ldots, R_n\}$. These relations define the *signature* of the model.

Taking a specific word such as /paʔ/, we can instantiate it within the model to obtain an explicitly defined structure $\mathcal{M}^{\text{paʔ}}$. Since the model contains symbols such as σ, onset, rime, x, and various IPA symbols, we treat these "atoms" as unary relations. In contrast, relations that represent hierarchical or associative connections—those that involve two or more symbols—are modeled as binary relations. Within this framework, we define three primary binary relations: (1) dominance (▼), representing hierarchical relations such as syllable structure dominance; (2) association (α), representing feature or tonal association with segments; and (3) successor ($\triangleleft$), representing the temporal ordering among elements within each tier.relation $\triangleleft$). The signature of $\mathcal{M}^{\text{paʔ}}$ can be written as

$$\mathcal{M} = \langle D; \{\ R_i \mid i \in \{\text{a, e, i, p, b, t}, \dots\} \cup \{\text{H, L, st, sl}, \sigma, \text{onset, rime, slot}\}\ \} \cup \{\ \blacktriangledown, \triangleleft, \alpha\ \}\rangle$$

We can further use those atoms to include the features:

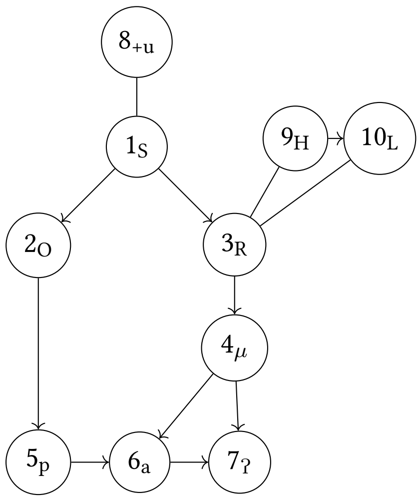
$$\begin{aligned}
\mathsf{cons} &\overset{\text{def}}{=} \{p, b, t, d, k, g\} \\
\mathsf{vocalic} &\overset{\text{def}}{=} \{a, i, e, u\} \\
\mathsf{nasal} &\overset{\text{def}}{=} \{m, n, N\} \\
\mathsf{glottal} &\overset{\text{def}}{=} \{\text{ʔ,ɦ}\} \\
\mathsf{voiced} &\overset{\text{def}}{=} \{\text{b,d,g}\}
\end{aligned}$$



D	$\overset{\text{def}}{=}$	$\{1, 2, 3 \dots 13\}$
$\triangleleft(x, y)$	$\overset{\text{def}}{=}$	$\{\langle 4, 5\rangle, \langle 5, 6\rangle, \langle 7, 8\rangle, \langle 8, 9\rangle, \langle 11, 12\rangle\}$
$\alpha(x, y)$	$\overset{\text{def}}{=}$	$\{\langle 4, 7\rangle, \langle 5, 8\rangle, \langle 6, 9\rangle, \langle 7, 10\rangle, \langle 8, 11\rangle\ \langle 9, 12\rangle\}$
$\blacktriangledown(x, y)$	$\overset{\text{def}}{=}$	$\{\langle 1, 2\rangle, \langle 1, 3\rangle, \langle 2, 4\rangle, \langle 3, 5\rangle, \langle 8, 11\rangle, \langle 9, 12\rangle\}$
$\sigma(x)$	$\overset{\text{def}}{=}$	$\{1\}$
$\mathsf{rime}(x)$	$\overset{\text{def}}{=}$	$\{2\}$
$\mathsf{onset}(x)$	$\overset{\text{def}}{=}$	$\{3\}$
$\mathsf{slot}(x)$	$\overset{\text{def}}{=}$	$\{4, 5, 6\}$
$\mathsf{p}(x)$	$\overset{\text{def}}{=}$	$\{7\}$
$\mathsf{a}(x)$	$\overset{\text{def}}{=}$	$\{8\}$
$\text{ʔ}(x)$	$\overset{\text{def}}{=}$	$\{9\}$
$\mathsf{st}(x)$	$\overset{\text{def}}{=}$	$\{10\}$
$\mathsf{sl}(x)$	$\overset{\text{def}}{=}$	$\{11\}$
$\mathsf{H}(x)$	$\overset{\text{def}}{=}$	$\{11\}$
$\mathsf{L}(x)$	$\overset{\text{def}}{=}$	$\{12\}$

The same word paʔ in the representational proposed by Zhu will have a different model with a different signature

$$\mathcal{M} = \langle D; \{\ R_i \mid i \in \{\text{a, e, i, p, b, t}, \dots\} \cup \{\text{H, L, upper}, \sigma, \text{onset, rime}, \mu\}\ \} \cup \{\ \blacktriangledown, \triangleleft, \alpha\ \}\rangle$$



D	$\stackrel{\text{def}}{=}$	$\{1, 2, 3 \ldots 7\}$
$\lhd(x, y)$	$\stackrel{\text{def}}{=}$	$\{\langle 2, 3\rangle, \langle 5, 6\rangle, \langle 6, 7\rangle, \langle 9, 10\rangle\}$
$\alpha(x, y)$	$\stackrel{\text{def}}{=}$	$\{\langle 1, 8\rangle, \langle 3, 9\rangle, \langle 3, 10\rangle\}$
$\blacktriangledown(x, y)$	$\stackrel{\text{def}}{=}$	$\{\langle 1, 2\rangle, \langle 1, 3\rangle, \langle 2, 5\rangle, \langle 3, 4\rangle, \langle 4, 6\rangle, \langle 4, 7\rangle\}$
$\sigma(x)$	$\stackrel{\text{def}}{=}$	$\{1\}$
$\mathsf{rime}(x)$	$\stackrel{\text{def}}{=}$	$\{2\}$
$\mathsf{onset}(x)$	$\stackrel{\text{def}}{=}$	$\{3\}$
$\mu(x)$	$\stackrel{\text{def}}{=}$	$\{4\}$
$\mathsf{p}(x)$	$\stackrel{\text{def}}{=}$	$\{5\}$
$\mathsf{a}(x)$	$\stackrel{\text{def}}{=}$	$\{6\}$
$\mathsf{ʔ}(x)$	$\stackrel{\text{def}}{=}$	$\{7\}$
$\mathsf{upper}(x)$	$\stackrel{\text{def}}{=}$	$\{8\}$
$\mathsf{H}(x)$	$\stackrel{\text{def}}{=}$	$\{9\}$
$\mathsf{L}(x)$	$\stackrel{\text{def}}{=}$	$\{10\}$

$$
\begin{aligned}
\phi_{\mathsf{license}}(x) &\overset{\text{def}}{=} \neg\Big(\mathsf{slot}(x) \wedge \exists y\, [\, \blacktriangledown(\mathsf{onset}(y), x) \vee \blacktriangledown(x, \mathsf{glottal}(y))\,]\Big) \\
C &= 1, \\
\sigma^1(x) &\overset{\text{def}}{=} \sigma(x) \\
\mathsf{onset}^1(x) &\overset{\text{def}}{=} \mathsf{onset}(x) \\
\mathsf{rime}^1(x) &\overset{\text{def}}{=} \mathsf{rime}(x) \\
\mathsf{H}^1(x) &\overset{\text{def}}{=} \mathsf{H}(x) \\
\mathsf{L}^1(x) &\overset{\text{def}}{=} \mathsf{L}(x) \\
\mu^1(x) &\overset{\text{def}}{=} \mathsf{slot}(x) \\
\mathsf{cons}^1(x) &\overset{\text{def}}{=} \mathsf{cons}(x) \\
\mathsf{vowel}^1(x) &\overset{\text{def}}{=} \mathsf{vowel}(x) \\
\mathsf{upper}(x^1) &\overset{\text{def}}{=} \mathsf{st}(x) \vee \neg\mathsf{sl}(x) \\
\lhd^{(1,1)}(x, y) &\overset{\text{def}}{=} \lhd(x, y), \\
\blacktriangledown^{(1,1)}(x, y) &\overset{\text{def}}{=} \blacktriangledown(x, y) \\
&\vee \exists x, y, z(\lhd(y, z) \wedge \mathsf{cons}(y) \wedge \mathsf{vocal}(z) \wedge \mathsf{onset}(x)) \\
&\vee \exists x, y, z(\lhd(z, y) \wedge \mathsf{glottal}(y) \wedge \mathsf{vocal}(z) \wedge \mathsf{slot}(x)) \\
\alpha^{(1,1)}(x, y) &\overset{\text{def}}{=} \mathsf{upper}(x) \vee \sigma(y) \\
&\vee \mathsf{rime}(x) \vee (\mathsf{H}(y) \wedge \mathsf{L}(y))
\end{aligned}
$$

$$
\begin{aligned}
\phi_{\mathsf{license}}(x) &\stackrel{\text{def}}{=} \text{True}\\
C &\stackrel{\text{def}}{=} 2,\\
\sigma^1(x) &\stackrel{\text{def}}{=} \sigma(x)\\
\mathsf{onset}^1(x) &\stackrel{\text{def}}{=} \mathsf{onset}(x)\\
\mathsf{rime}^1(x) &\stackrel{\text{def}}{=} \mathsf{rime}(x)\\
\mathsf{H}^1(x) &\stackrel{\text{def}}{=} \mathsf{H}(x)\\
\mathsf{L}^1(x) &\stackrel{\text{def}}{=} \mathsf{L}(x)\\
\mathsf{slot}^1(x) &\stackrel{\text{def}}{=} \mathsf{slot}(x)\\
\mathsf{slot}^2(x) &\stackrel{\text{def}}{=} \mathsf{cons}(x) \wedge \exists y(\lhd (x,y) \wedge \mathsf{vocal}(y))\\
&\vee \mathsf{glottal}(x) \wedge \exists y(\lhd (y,x) \wedge \mathsf{vocal}(y))\\
\mathsf{cons}^1(x) &\stackrel{\text{def}}{=} \mathsf{cons}(x),\\
\mathsf{vowel}^1(x) &\stackrel{\text{def}}{=} \mathsf{vowel}(x)\\
\mathsf{st}(x^1) &\stackrel{\text{def}}{=} \mathsf{upper}(x) \vee \neg \mathsf{sl}(x)\\
\lhd^{(1,1)}(x,y) &\stackrel{\text{def}}{=} \lhd(x,y),\\
\blacktriangledown^{(1,1)}(x,y) &\stackrel{\text{def}}{=} \blacktriangledown(x,y)\\
&\vee \neg \mathsf{onset}(x) \wedge \mathsf{glottal}(y))\\
\blacktriangledown^{(1,2)}(x,y) &\stackrel{\text{def}}{=} \exists x,y,z(\lhd(y,z) \wedge \mathsf{cons}(y) \wedge \mathsf{vocal}(z) \wedge \mathsf{onset}(x))\\
&\vee \exists x,y,z(\lhd(z,y) \wedge \mathsf{glottal}(y) \wedge \mathsf{vocal}(z) \wedge \mathsf{rime}(x))\\
\blacktriangledown^{(2,1)}(x,y) &\stackrel{\text{def}}{=} \exists x,y,z(\lhd(y,z) \wedge \mathsf{cons}(y) \wedge \mathsf{vocal}(z) \wedge x = y)\\
&\vee \exists x,y,z(\lhd(z,y) \wedge \mathsf{glottal}(y) \wedge \mathsf{vocal}(z) \wedge x = y)\\
\alpha^{(1,1)}(x,y) &\stackrel{\text{def}}{=} \exists x,y,z(\mathsf{cons}(x) \wedge \mathsf{vow}(z) \wedge \mathsf{st}(y) \wedge \lhd(x,z))\\
&\vee \exists x,y,z(\mathsf{vowel}(x) \wedge \mathsf{H}(y) \wedge \mathsf{L}(y))
\end{aligned}
$$

References

Bao, Z. (1999). The structure of tone. Oxford University Press.

Chandlee, J., Eyraud, R., Heinz, J., Jardine, A., and Rawski, J. (2019). Learning with partially ordered representations. In Proceedings of the 16th Meeting on the Mathematics of Language, pages 91–101.

Chao, Y.-R. (1930). A system of tone letters. Le maître phonétique.

Chen, Y. and Gussenhoven, C. (2015). Shanghai chinese. Journal of the International Phonetic Association, 45(3):321–337.

Clements, G. N. (1985). The geometry of phonological features. In van der Hulst, H. and Smith, N., editors, Phonology Yearbook 2, pages 225–252. Cambridge University Press, Cambridge.

CUHK (2018). Multi-function chinese character database. https://humanum.arts.cuhk.edu.hk/Lexis/lexi-mf/.

Duanmu, S. (1988). A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages. PhD thesis, Massachusetts Institute of Technology.

Duanmu, S. (1999). Metrical structure and tone: evidence from mandarin and shanghai. Journal of East Asian Linguistics, 8(1):1–38.

Oakden, C. (2020). Notational equivalence in tonal geometry. Phonology, 37(2):257–296.

Ren, H. (1992). On the Domain of Tone in Shanghai Chinese. PhD thesis, Ohio State University.

Sagey, E. C. (1986). The Representation of Features and Relations in Nonlinear Phonology. Phd dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Xu, B. and Tang, Z. (1988). A description of the dialect of the shanghai city.

Yip, M. (1980). The tonal phonology of Chinese. PhD thesis, Massachusetts Institute of Technology.

Yip, M. (1993). Tonal register in east asian languages. In Snider, K. and van der Hulst, H., editors, The Phonology of Tone: The Representation of Tonal Register, pages 245–268. Mouton de Gruyter, Berlin.

Zee, E. and Maddieson, I. (1979). Tones and tone sandhi in shanghai: Phonetic evidence and phonological analysis. UCLA working papers in phonetics, 45(1):93–129.

Zhu, X. (1999). Shanghai Tonetics. LINCOM studies in Asian linguistics. LINCOM EUROPA.

Zhu, X. and Wang, C. (2015). Tone. In Wang, W. S.-Y. and Sun, C., editors, The Oxford Handbook of Chinese Linguistics, chapter 37, pages 503–516. Oxford University Press, Oxford.