

Variations of focus prominence in three tone languages

Katrina Kechun Li, Francis Nolan, Brechtje Post

University of Cambridge
kl502@cam.ac.uk, fjn1@cam.ac.uk, bmbp2@cam.ac.uk

ABSTRACT

Prosodic focus involves a diverse range of phonological and phonetic devices, and cross-linguistic variation is often subject to the constraints of the prosodic profile of a language. In tone languages, the functional load of tones limits the scope for focus realisation in terms of discrete categories. The current study investigates whether tone languages can still realise focus in distinctive prosodic patterns by examining three closely related languages: Chengdu dialect (a variety of southwest Mandarin), Changsha dialect (a variety of Xiang) and Guangzhou Cantonese. The analysis of pitch and duration reveals systematic differences in terms of both phonology and phonetic realisation between them. It emerges that the three languages demonstrate various degrees of tonal reduction at pre- and post-focal positions. We relate the variations in sentence-level focal prominence to the languages' lexical-level tonal prosody, and discuss the results in the light of metrical properties of tone languages.

Keywords: prosodic focus, prosodic typology, metrical structure, tone languages, Chinese languages

1. INTRODUCTION

The conveying of information structure, especially focus, has been an important phenomenon in informing typological studies on prosody. Despite the diverse phonological and phonetic devices involved in prosodic focus, the prosodic means available to a language are subject to constraints within its prosodic profile [1]. For example, in languages with lexical stress (e.g. English), focal prominence (or its absence) is typically realised as the application (or non-application) of pitch accents to lexically stressed syllables. This is reflected in the autosegmental-metrical theory (AM) by the association of pitch accents with metrically strong syllables [2]. In contrast, languages lacking lexical stress (e.g. Korean) do not have such an anchor for pitch accents and thus rely more on prosodic phrasing. Nevertheless, there is no straightforward relationship between focus encoding and the lexical prosody of a language, as the phonetic realisation also interacts with other language-specific constraints such as syntactic and semantic features [3]–[5].

In tone languages, prosodic encodings of focus are subject to the constraint of preserving tonal categories, which limits the scope for realising focus in discrete categories. Previous studies have focused on the gradient phonetic effects, primarily pitch modification, which is often accompanied by duration and intensity changes [6]–[8]. However, recent studies also suggest that variations of focus marking in terms of discrete categories might exist even within closely related tonal languages. At the on-focal position, for example, Cantonese and Southern Min have been found to be more limited in f_0 modification than standard Mandarin and only consistent in showing a durational change [9][10]. At the post-focal position, tone languages also differ in having post-focal compression or not [11]–[13].

This study extends previous comparative studies by investigating the realisation of prosodic focus in three Chinese languages. Furthermore, we relate the variations in focal prominence to other aspects of their prosodic systems, particularly the reduction in lexical prosody (cf. Standard Mandarin neutral tone). Finally, we discuss the implications of our findings for understanding the underlying metrical organisation of tone languages.

2. METHODOLOGY

Three dialects from distinct Chinese language families were chosen: Chengdu dialect (a variety of southwest Mandarin), Changsha dialect (a variety of Xiang) and Guangzhou Cantonese.

2.1. Materials

Target sentences with SVO structure were constructed in each language. In Chengdu and Changsha, the subject (S) was a name prefix plus a commonly used name, and the monosyllabic verb (V) carried the same tone as the name. The object (O) was also designed to be a disyllabic word with the first syllable carrying the same tone as the preceding name and verb, and the second syllable being a suffix (examples below; the key syllables that were designed to carry the same tone are in bold). The Cantonese dataset, collected at an earlier time and not designed for comparison with other languages, did not control for tone on the object nor the type of onsets. However, it will not affect our ability to

address the main research question, which is how the tonal contour of the same sentence changes in different focus conditions. Two tonal conditions on the key syllables which demonstrated comparable pitch contours across the three dialects are reported here: a maximally dynamic contour reaching the ceiling of the tonal space (‘dynamic’ group: Chengdu T3[53], Changsha T4[45], Cantonese T2[35]) and a relatively static tonal contour (‘static’ group: Chengdu T1[45], Changsha T3[42], Cantonese T1[55]). The numbers are tonal representations in Chao number for citation syllables, but Changsha T3[42] undergoes sandhi and changes to [44].

(1) Chengdu:

eiao53	jm53	mai53	lǎ0	pa53	ji53	tsi0
Name	Name	Verb	Aspect	Classifier	Noun	Suffix
Prefix			marker			
Xiao-yin		bought		a	chair	

(2) Changsha:

eiao42	jm45	mai45	ta21	tʂa34	iou45	tsi33
Name	Name	Verb	Aspect	Classifier	Noun	Suffix
Prefix			marker			
Xiao-ying		sold		a	grapefruit	

(3) Cantonese:

a33	jm35	hau35	tʃɔ35	tʃɛt5-fap1	kei35	fɛn55
Name	Name	Verb	Aspect	Number	Numb	Noun
Prefix			marker		-er	
A-ying		achieved		seventy	-ish	points

The target sentences were elicited under various focus conditions, including the three reported here (Table 1), which were all induced by wh-questions, but targeted either subject, object or neither (broad focus). The Cantonese dataset did not contain an object focus condition.

Table 1: Focus conditions

Condition	Prompt	Answer
Broad focus	What happened?	Xiao-yin brought a basket.
Subject focus	Who brought a basket?	Xiao-yin brought a basket.
Object focus	What did Xiao-yin bring?	Xiao-yin brought a basket .

2.2. Recording Procedure

Native speakers of Chengdu (10, f=5), Changsha (8, f=4) and Cantonese (13, f=6) were recorded in a quiet room locally. Participants were presented with an audio recording of the prompt question (recorded by a native speaker) together with the written mini-dialogue, in order to resemble the environment of natural dialogues. Focus and tonal conditions were semi-randomized, with the broad focus condition for each sentence always appearing first. Each sentence was recorded three times (twice in Cantonese).

2.3. Annotation and Analysis

The recordings were segmented using *The Penn Phonetics Lab Forced Aligner* (P2FA) [14]. Although developed for Mandarin Chinese, adapting the dictionaries to fit other Chinese languages yields reasonably good performance given the similar syllable structures. The automatically aligned boundaries for onsets and rhymes were manually corrected. We then extracted duration of the syllable and f0 from ten equidistant time points of the rhyme. A set of *Praat* [15] scripts was developed to facilitate the process (scripts can be retrieved from [16]).

In the analysis, f0 was converted to semitones with each participant’s average f0 as their base [17]. We built generalised additive mixed models (GAMM) to analyse the effects of focus, tone (dynamic/static), and language on f0 contours, using R [18] and the *mgcv* package [19]. We also included factor smooth to model the non-linear random effect of the tone and focus conditions for each speaker. The final model was corrected for autocorrelation and fitted with the scaled *t* distribution [20]. Duration was analysed using linear mixed-effects models (LMM), including focus, tone, language, position (Prefix, Name, etc) and their interactions as fixed effects and a random intercept for each speaker.

3. RESULTS

Figure 1 (left) shows the average f0 contours of the dynamic group on the five key syllables. The visualisations can be inspected in two ways. From a paradigmatic perspective, we can compare the f0 contours ‘vertically’ at the same position, which reveals the effect of different focus conditions. While this may be influenced by different baselines (e.g. a higher overall f0 under broad focus due to all new information), a syntagmatic perspective can reveal how focus changes relative prominence within a focus condition by comparing the pitch level of different positions ‘horizontally’. Viewing the f0 contours of Chengdu and Changsha paradigmatically, at the position ‘Object(Syll)’, the f0 contour of subject focus condition (blue solid line) is significantly reduced compared to the object focus (orange dashed line) and broad focus condition (grey dotted line), indicating that post-focal material is realised with reduced pitch range. In contrast, the f0 contour in Cantonese barely changes. Similarly, at the position ‘Name’, the f0 contour of object focus condition also shows some degree of compression compared to that of subject focus and broad focus in Chengdu and Changsha, indicating a ‘pre-focal’ compression. At both positions, Chengdu has a higher

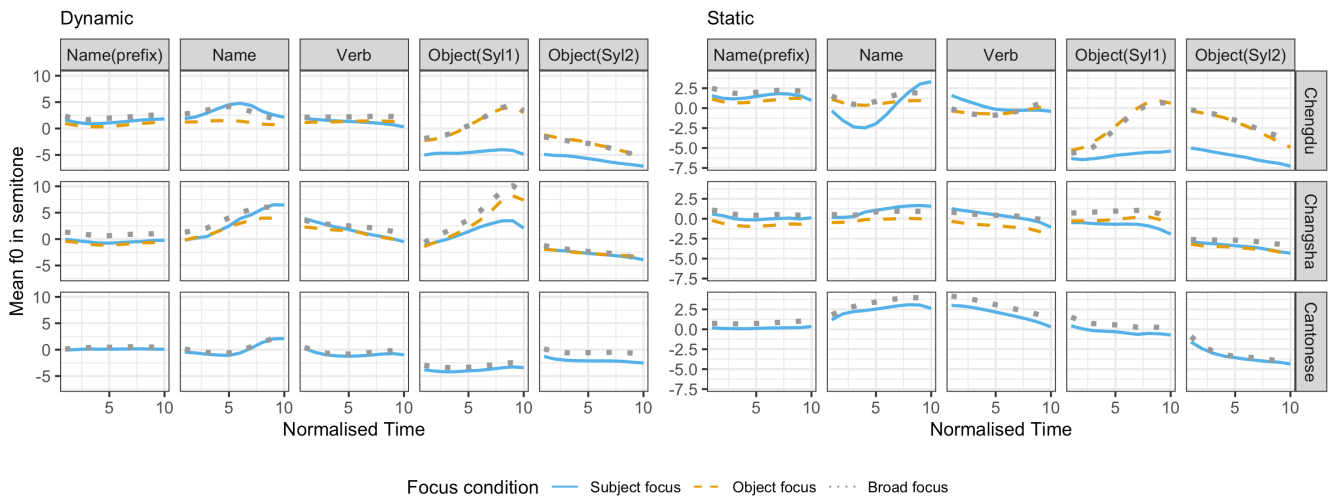


Figure 1: The f_0 contours under different focus conditions (left: dynamic group; right: static group).

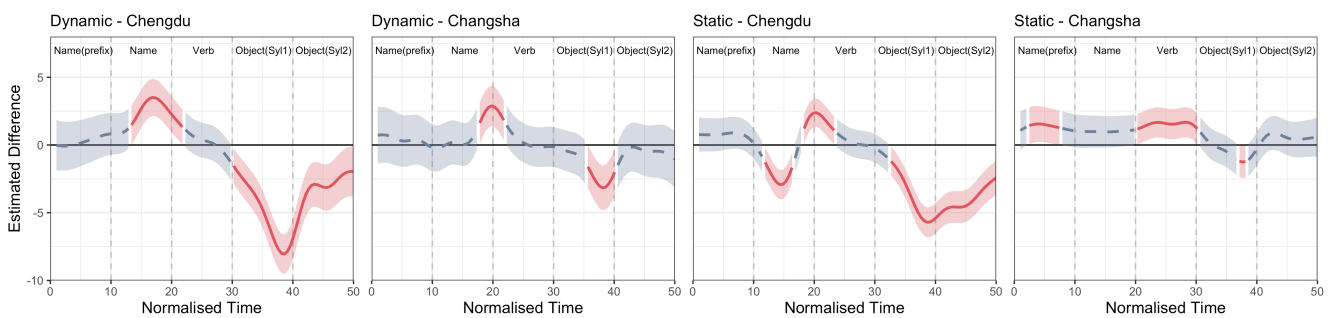


Figure 2: Predicted difference curves (the subject focus minus the object focus condition) in Chengdu and Changsha for the dynamic group (left) and static group (right). The shading displays the 95% CI of the predicted mean difference. The red solid line indicates when the value is significantly different from zero.

degree of compression such that the f_0 contour almost becomes flattened, accompanied also by a strong register-lowering effect. The different degrees of compression in Chengdu and Changsha are confirmed using the GAMM model, as shown by the predicted f_0 difference between the subject focus and the object focus condition in Figure 2 (left). The significant difference (red solid line) is larger in scale and covers a longer time range in Chengdu.

The f_0 findings for the static group tell a similar story as shown in Figure 1 (right; notice the different scale on the y-axis compared to the left). Given that pitch range is less likely to change for level tones, the most noticeable effect from a paradigmatic view is that the slight rising trend on key syllables in Chengdu becomes much more significant under focus. Although there is no such tendency in Changsha, the effect of focus is still evident from a syntagmatic view. The f_0 contour of subject focus (blue solid) condition is higher at ‘Name’ than at ‘Object(Syl1)’ whereas this tendency is reversed under object focus (orange dashed). For Cantonese, however, the f_0 contours remain parallel under different focus conditions, reflecting only a change in baseline. The predicted difference curve generated by the GAMM model (Figure 2 right) confirms that Chengdu involves more f_0 changes in marking focal prominence.

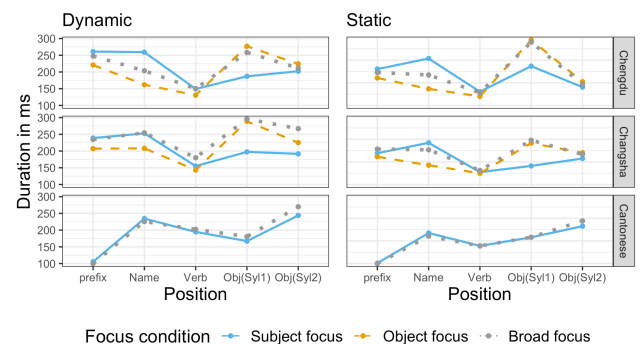


Figure 3: The duration of the dynamic group (left) and static group (right) under different focus conditions.

The duration patterns (Figure 3) mirror the f_0 patterns on the syllables. In both Chengdu and Changsha, the blue solid lines of both dynamic and static groups are significantly higher than the orange dashed line at ‘Name’, and lower at ‘Object(Syl1)’, indicating a boosting prominence under focus and a reduced prominence elsewhere. The post-hoc pairwise comparisons of different focus conditions on two key syllables (Table 2) confirm the existence of pre- and post-focal durational compression in both languages, and the differences are larger in Chengdu, especially between the subject and object focus, as shown in the last line of Table 2. In contrast, the duration changes are not significant in Cantonese.

Table 2: Post-hoc pairwise comparisons of duration at ‘Name’ and ‘Object(Syll)’ under different focus conditions in Chengdu and Changsha. The number indicates model estimates with p values in brackets (estimates with $p < 0.0001$ are marked in bold).

Focus Condition	Chengdu		Changsha	
	Name	Object	Name	Object
Broad-Subject	-64.06 ($<.0001$)	88.53 ($<.0001$)	-14.58 (0.3967)	105.65 ($<.0001$)
Broad-Object	51.60 (0.0057)	-14.28 (0.3023)	57.17 ($<.0001$)	9.69 (0.6640)
Subject-Object	115.65 ($<.0001$)	-102.80 ($<.0001$)	71.74 ($<.0001$)	-95.97 ($<.0001$)

4. DISCUSSION

In summary, our findings demonstrate systematic variations in sentence-level compression when marking focal prominence in terms of both f_0 contour and duration across the three languages. We showed that both pre-focal and post-focal compressions are possible, despite post-focal compression receiving more attention in previous literature [11]. While Changsha shows a moderate degree of compression with the tonal shapes well maintained, Chengdu exhibits a much larger degree of compression at both pre- and post-focal positions such that the tonal contours are flattened. In contrast, no compression is found in Cantonese. Although the pre-focus condition is not examined in Cantonese, it is likely that this pre-focal compression is also absent since it is often less manifest than the post-focal compression.

These different degrees of sentence-level compression reflect the stability of tonal realisations, which can be related to the phenomena of lexical-level tonal reductions. Specifically, tones can undergo phonetic reduction at unstressed positions in Changsha words, but the tonal neutralisation is incomplete and the original tonal category is partially identifiable [21]–[23]. In Chengdu, in contrast, the tonal neutralisation leads to a complete loss of the original tonal identities and the phonetic realisations entirely depend on the surrounding tonal contexts, similar to Mandarin neutral tones. In Cantonese, however, there is no such lexical tonal reduction.

This apparent correspondence of sentential and lexical level tonal reductions suggests the relevance of underlying metrical structure. Metrical structure captures the relative prominence relationship within a word (e.g. lexical stress) as well as within a phrase or sentence (e.g. relative metrical strength [24, p. 263]). Cross-linguistically, metrical structure can be ‘phonologically activated’ [25] to different degrees, driving positional (non-)prominence effects on segments or tones. Specifically, languages with a stronger metrical motivation, such as Chengdu, have more freedom to manipulate the tonal contour to

express pragmatic functions, as well as showing more stress-like properties on the lexical level. Lexical stress in Chengdu is almost an undisputable property, unlike its debated status in Mandarin, which is evidenced not only by the neutral tone but also the consistent duration lengthening at stressed positions in disyllabic words [26]. Changsha has a weaker metrical motivation compared to Chengdu, followed by Cantonese. For languages like Cantonese where the metrical structure does not allow for drastic changes in relative prominence, there are no phonological phenomena that can be attributed to lexical stress. Alternative means such as syntactic particles are employed for pragmatic functions [27].

Although it is not straightforward to find stable acoustic correlates of metrical structure in tone languages due to the functional loads of tones, the metrical structure at word level has been discussed in other Chinese languages with respect to tonal neutralisation as well as the tone-spreading kind of tone sandhi [28][29]. These languages whose focal prominence has been examined seem to conform to our prediction that their compression at sentence level is in accordance with tonal reduction at word level. The implication is that the dimension of reduction might be particularly useful for characterising prosodic types in tone languages. More studies are needed to further testify to the connection.

5. CONCLUSIONS

The current study provides a comparative analysis of focal prominence by examining the changes that occur in comparable tonal shapes under different focus conditions across languages. The analysis of f_0 along with duration reveals that Cantonese, Changsha, and Chengdu show (in this order) increasing levels of compression, not only at post- but also at pre-focal positions. Moreover, sentence-level compression is found to correlate with lexical tonal reduction in these languages, suggesting the relevance of the underlying metrical structure. Languages with a higher degree of activation of the metrical structure have greater flexibility in manipulating the tonal contour at both word and sentence level. At word level, metrical structure can be inferred through phenomena like tonal neutralisation and tone sandhi. At sentence level, focal prominence is particularly informative, as it tends to induce the most obvious alternation of strong and weak. Interpretation of the sentence-level prosodic pattern nevertheless shows the need of combining a paradigmatic and syntagmatic comparison. Finally, this study highlights the importance of prominence reduction in collaboration with prominence enhancement in characterising typological differences in tone languages.

6. ACKNOWLEDGEMENTS

This work was supported by the CSC Cambridge Scholarship and a grant from the University of Cambridge Faculty of Modern and Medieval Languages and Linguistics Fieldwork Fund.

7. REFERENCES

- [1] F. Kügler and S. Calhoun, 'Prosodic encoding of information structure: A typological perspective', in *The Oxford Handbook of Language Prosody*, Oxford University Press, 2020. doi: 10.1093/oxfordhb/9780198832232.013.30.
- [2] J. B. Pierrehumbert, 'The phonology and phonetics of English intonation', Thesis, Massachusetts Institute of Technology, 1980.
- [3] R. S. Burdin, S. Phillips-Bourass, R. Turnbull, M. Yasavul, C. G. Clopper, and J. Tonhauser, 'Variation in the prosody of focus in head- and head/edge-prominence languages', *Lingua*, vol. 165, pp. 254–276, 2015, doi: 10.1016/j.lingua.2014.10.001.
- [4] S. Calhoun, 'The interaction of prosody and syntax in Samoan focus marking', *Lingua*, vol. 165, pp. 205–229, 2015, doi: 10.1016/j.lingua.2014.11.007.
- [5] F. Hamlaoui, M. Žygis, J. Engelmann, and M. Wagner, 'Acoustic correlates of focus marking in Czech and Polish', *Lang. Speech*, vol. 62, no. 2, pp. 358–377, 2019, doi: 10.1177/0023830918773536.
- [6] M. Breen, E. Fedorenko, M. Wagner, and E. Gibson, 'Acoustic correlates of information structure', *Lang. Cogn. Process.*, vol. 25, no. 7–9, pp. 1044–1098, Sep. 2010, doi: 10.1080/01690965.2010.504378.
- [7] Y. Xu, 'Effects of tone and focus on the formation and alignment of f0 contours', *J. Phon.*, vol. 27, no. 1, pp. 55–105, 1999, doi: 10.1006/jpho.1999.0086.
- [8] Y. Chen and C. Gussenhoven, 'Emphasis and tonal implementation in Standard Chinese', *J. Phon.*, vol. 36, no. 4, pp. 724–746, 2008, doi: 10.1016/j.wocn.2008.06.003.
- [9] H. S. H. Fung and P. P. K. Mok, 'Temporal coordination between focus prosody and pointing gestures in Cantonese', *J. Phon.*, vol. 71, pp. 113–125, 2018, doi: 10.1016/j.wocn.2018.07.006.
- [10] H.-H. Pan, 'Focus and Taiwanese unchecked tones', in *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, C. Lee, M. Gordon, and D. Büring, Eds. Dordrecht: Springer Netherlands, 2007, pp. 195–213.
- [11] Y. Xu, S. Chen, and B. Wang, 'Prosodic focus with and without post-focus compression: A typological divide within the same language family?', *Linguist. Rev.*, vol. 29, no. 1, pp. 131–147, 2012, doi: 10.1515/trlr-2012-0006.
- [12] B. Wang, L. Wang, and T. Qadir, 'Prosodic realization of focus in six languages/dialects in China', in *ICPhS XVII*, Hong Kong, 2011, p. 4.
- [13] W. L. Wu and Y. Xu, 'Prosodic focus in Hong Kong Cantonese without post-focus compression', in *Proceedings of Speech Prosody 2010*, Chicago, IL, USA, 2010, pp. 1–4.
- [14] J. Yuan and M. Liberman, 'Speaker identification on the SCOTUS corpus', *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 3878–3878, 2008, doi: 10.1121/1.2935783.
- [15] P. Boersma and D. Weenink, 'Praat: doing phonetics by computer'. 2019. Accessed: May 26, 2019. [Online]. Available: <http://www.praat.org/>
- [16] K. Li, 'InTone', 2023, https://github.com/kc-li/InTone_Visualisation.
- [17] J. Zhang, 'A comparison of tone normalization methods for language variation research', in *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, Hong Kong, 2018.
- [18] R Core Team, 'R: A language and environment for statistical computing', R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: <https://www.R-project.org/>
- [19] S. N. Wood, 'Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models', *J. R. Stat. Soc. B*, vol. 73, no. 1, pp. 3–36, 2011.
- [20] M. Wieling, 'Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English', *J. Phon.*, vol. 70, pp. 86–116, 2018, doi: 10.1016/j.wocn.2018.03.002.
- [21] C. Guo, 'On the tonal neutralization of bi-syllabic prosodic words of Changsha dialect', *Studies in Prosodic Grammar*, vol. 7, pp. 131–149, 2021.
- [22] Q. Zhong, 'Light tones in Changsha dialect', *Fangyan*, vol. 3, pp. 255–264, 2003.
- [23] G. Wei, 'On the neutral tone and the atonicity', *Fangyan*, no. 01, pp. 20–29, 2000.
- [24] D. R. Ladd, *Intonational Phonology*, 2nd ed. Cambridge: Cambridge University Press, 2008.
- [25] L. M. Hyman, 'Do all languages have word accent? or: what's so great about being universal?', *UC Berkeley Phonol. Lab Annu. Rep.*, 2010
- [26] C. Liu, K. Li, and F. Nolan, 'On the speech rhythm and stress pattern of tone languages', *Journal of Sichuan University (Philosophy and Social Science Edition)*, no. 3, pp. 151–163, 2022.
- [27] P.-P. Lee, *Focus manifestation in Mandarin Chinese and Cantonese: a comparative perspective*. Abingdon; New York: Routledge/Taylor & Francis Group, 2019.
- [28] C. Guo and F. Chen, 'Phonetic realizations of metrical structure in tone languages: Evidence from Chinese dialects', *Front. Psychol.*, vol. 13, 2022,
- [29] S. Duanmu, 'Metrical structure and tone: Evidence from Mandarin and Shanghai', *J. East Asian Linguist.*, vol. 8, no. 1, pp. 1–38, 1999, doi: 10.1023/A:1008353028173.