Typological tendencies in verse and their cognitive grounding

De Castro Arrazola, V.

Author: De Castro Arrazola, V.
Title: Typological tendencies in verse and their cognitive grounding
Issue Date: 2018-05-03

Typological tendencies in verse and their cognitive grounding

Varuṇ deCastro-Arrazola

deCastro-Arrazola, V. 2018. *Typological tendencies in verse and their cognitive grounding*. Utrecht: LOT.

ISBN: 978-94-6093-284-7
NUR: 616

Cover illustration: Irati Gorostidi Agirretxe
Typesetting software: XƎLaTeX
Layout based on: LaTeX class langscibook.cls developed by Timm Lichte, Stefan Müller, Sebastian Nordhoff & Felix Kopecky for the open-access linguistics publisher Language Science Press (langsci-press.org).

Published by:
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111
e-mail: lot@uu.nl
http://www.lotschool.nl

Typological tendencies in verse and their cognitive grounding

Proefschrift
ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 3 mei 2018
klokke 13:45 uur

door
Varuṇaśarman de Castro Arrazola
geboren te Canberra
in 1988

Promotores
Prof.dr. Marc van Oostendorp (Radboud Universiteit Nijmegen)
Prof.dr. Johan Rooryck (Universiteit Leiden)

Promotiecommissie
Prof.dr. Janet Grijzenhout (Universiteit Leiden)
Prof.dr. Paula Fikkert (Radboud Universiteit Nijmegen)
Prof.dr. Nigel Fabb (University of Strathclyde)

The research for this book was carried out as part of the Horizon project 317-70-010 *Knowledge and culture*, funded by the Dutch Organisation for Scientific Research (NWO).

Indrari
ikerketa berdin
bizi baitugu

Contents

Acknowledgements

I'm all for reproducibility, and a thesis without acknowledgements would be a collection of results with no source code. If I need to redo any of this, I should keep in mind how much the people around me have helped, listened and inspired. I'm grateful to all of them.

First, but not least, Marc and Johan have guided me virtuously, and their office doors have always been open. It may seem a small recognition, but an open door and a willing ear should be precious to any supervisee.

Second, I have always had warm, supportive, industrious colleagues. Naturally, they are linked to particular locations, peculiar buildings, where we shared offices, coffee, plank, lunch and conversation. Naturally, they also transcend these buildings.

Xico, Sophie, Anna, Kathrin, Nina, Lotte, Peter, Berit at the old Coca-Cola factory. Those were days of beginner's excitement and thrill, maybe uncertainty. And those were days when, for the first time, with no prejudice whatsoever, phonologists accepted me as a phonologist, and Germans accepted me as a speaker of Dutch.

Etske, Reza, Olha, Diedrik, Folgert, Ruby, Jolien, and my loyal, caring, dear paranymphs, Mirella and Edoardo, at the old women's prison. This was the toughest period, when the chapters had to materialise and when existential questions assailed at their fiercest. Yet the steady human environment at work (with the occasional writing-in-a-train expedition, with the occasional dimibo) helped in many ways.

The rest was done in two far-away offices. At the very top of van Eyckhof's tower with Sjoerdieke, Arthur and Teresa. If research works in cycles of expansion and contraction, my time there was one of expansion, with discussion cocktails suitably summarised as *Gödel, Escher, Bach.* The other Horizon-fellows also played a key role in the cocktail: Joey, Paula, Caitlin, Lisa. Around the corner, a number of LUCLers made me feel at home, particularly Leticia, Martin, Mari Carmen, and Maarten, who generously introduced me to the Iraqw.

16,679 kilometres away from Leiden, Nick (via Hedvig) kindly let me use his office at the legendary Coombs building. No office-mates there, but I was lucky enough to meet Owen, who would introduce me to real-world, tangible parallelism (from Rote Island). And I was lucky enough to encounter Mark, who would provide condensed advice in condensed time, and welcome me within the lively Himalayan, Bumthang crowd.

Third, sporadic conferences and the like helped framing and re-framing the research questions in different contexts. Simon was readily available to discuss iterated learning. Petr and Steve endured my inexperience with database design. Darrell rightly insisted on explicit definitions for pattern mining. All the bright scholars I met in Paris, Tallinn and Helsinki are also noteworthy.

Fourth, before and after office hours, the dissertation kept on crystallizing through many explicit and implicit conversations. At Okapi, with Floris and Gaby, where it all once began. Nearby, with a collection of people from the MPI, particularly Ewelina, Francisco, Seán, Harald. Around food, films and woods, with Sara, Joan, Silas and family. Around bikes, runs and the precise amount of pasta, with Marco and Silvia.

Fifth, I cannot read or write properly while listening to music; as a rule. Still, some has been helpful at times, to focus or invigorate. At least: musicians from Antandroy-land, from Bitti and Orosei, Pérès, Burning Spear, Jayaprakash, and Lazkao Txiki while cycling through Vondelpark.

Sixth, but not least, my family was there. Despite the distance, they asked, they read, listened, challenged, and provided the firm anchor everyone needs to follow a path. And Amaia, beloved, has walked all along.

1 Introduction

This dissertation is about verse, some of its recurrent features, and cognitive aspects which can explain their prevalence. On the one hand, I present computational tools to describe systematically regularities in verse corpora; on the other hand, I conduct experiments to show how aspects of human cognition can shape the verse traditions of the world.

1.1 On verse

By *verse* I refer to a range of verbal phenomena, most typically songs and poems, but also nursery rhymes, religious chants or demonstration slogans. Compared to everyday speech, all these forms show additional layers of structure, like a regular alternation of accented syllables, a fixed melody, or a systematic number of syllables per utterance. This kind of verse-specific structures can be conceived in terms of a template to which words are set; examples of templates are the tunes of strophic songs, a poetic form like a limerick, or four-beat structures typical of slogans.

Besides introducing rhythmic or melodic regularities to the discourse, verse templates introduce their own constituent structure, such as stanzas, couplets, lines or feet. Crucially, verse constituents cannot be defined just by relying on syntactic, phonological or pragmatic criteria. Often, the presence of one of these extra-speech constituency levels, namely, the line, is considered a necessary and sufficient condition to define verse: "one can accept a minimal definition of poetry as discourse organized in lines" (Hymes 1977:454); "verse is text which is divided into lines" (Fabb 2014:29).

The distinction between everyday conversation and verse is also described in terms of a continuum, where aspects of language such as prominence or pitch are subject to increasing restrictions or regularities (List 1963). This view accommodates emic taxonomies where, beyond forms we would name *song* or *speech*, there exist intermediate forms such as chanting and recitation (e.g. Seeger 2004:45 for such an analysis among the Suyá). Even in the lack of a clear-cut definition, and without a comprehensive survey of every human society, there is a consensus that verse is a universal phenomenon (Brown & Jordania 2011).

Notwithstanding its unifying features, the fact that verse is both widespread and diverse has led to it being studied from an array of angles within fields such as metrics, linguistics, musicology, literary studies or psychology. This dissertation takes an explicitly inclusive approach by drawing from these different types of sources.

1.2 Explaining verse

As with other cultural universals such as language (Christiansen & Kirby 2003) and cooking (Wrangham et al. 1999), verse traditions include, on the one hand, an extreme diversity typical of culturally evolved practices, and, on the other hand, a shared set of principles, which, minimally, serve us to identify the phenomenon as universal. The alleged universality of verse poses questions related to its origins in phylogeny, and its function or evolutionary advantage within the biological and cultural context. These issues in relation to music more generally have been of interest to science for a long time (Darwin 1871), and are receiving renewed attention, mainly due to the development of music cognition research and to comparative studies of music-like phenomena across cultures and species (Wallin, Merker & Brown 2001; Honing et al. 2015; Fitch 2015; Hoeschele et al. 2015; Richter & Ostovar 2016; Patel & Demorest 2013).

The current thesis takes for granted the universality of verse, and tackles the issue of why, among its conspicuous diversity, certain features of verse remain fairly constant across traditions. This belongs to the more general question of causality in verse: why does a given instance of verse show a feature x? Leading ethnomusicologist Bruno Nettl, in his overview of the discipline, states that the central question of ethnomusicology concerns "what it is that determines the nature of the musical style (…) of a human society or ethnic group" (Nettl 2010:340).

Researchers have tackled the problem from a number of perspectives, often dividing the potential determinants according to the nature–nurture dichotomy: "the function of tones in relation to each other cannot be explained adequately as part of a closed system without reference to the structures of the sociocultural system of which the musical system is a part, and to the biological system to which all music makers belong" (Blacking 1974:30). Underlying, there is the idea that songs (or other cultural practices) are constrained by (1) highly variable, culture-specific or even idiosyncratic principles or values, and (2) fixed, innate properties which keep constant across populations.

Even if popular, the concept of *innateness* has proven highly problematic given the complex interactions between genetic endowment and environment during the development of individuals (Mameli & Bateson 2006; 2011). Hence, bearing in mind the readiness of humans to construct and modify the environment they live in (i.e. niche construction, Odling-Smee, Laland & Feldman 2003), the distinction between fixed and variable determinants is better regarded as a fuzzy one. In this respect, we can regard the complete set of variables potentially explaining the verse features of a tradition as the *verse niche* surrounding the poets and songwriters (comparable to the linguistic niche paradigm proposed by Lupyan & Dale 2015). The verse niche extends from the microscopic anatomical details of an individual larynx or auditory cortex, to aesthetic and economic values of the closest community or broader society, and to wide-ranging properties of the ecosystem, such as the climate. The dissertation only tackles cognitive determinants of verse, but I will first discuss other types of determinants in order to establish the appropriate context.

As an illustration of the potential effect of the broader environment, consider the observation by Blacking (1965) about the relation between terrain and musical tempo in two Bantu languages: "the tempo of Venda music is related to the steady walking pace of the people, thrust upon them by the mountainous environment", whereas the neighbouring Tsonga people "live in the flatlands, walk faster, and have music of more rapid tempo" (quoted in Nettl 2010:345). According to Blacking (1967:25), properties of the terrain make certain types of dances and their accompanying music more suitable than others. This link between the walking affordance promoted by the surrounding and the preferred tempo has been generalised as a bias active in all human music:

> Since adults take about 10,000 steps per day (Tudor-Locke & Myers 2001) the presentation of the sound/sensory stimuli to the fetus is ubiquitous under normal circumstances. Music is often played at tempos similar to walking (Changizi 2011) and it is reasonable to propose that one of the bases for this connection results from the concurrent stimuli of the tactile and sonic sensations of maternal footfalls informing the developing brain of the fetus. (Teie 2016:3)

A note of caution applies to this and the following potential determinants of musical or linguistic traits: given their correlational nature (e.g. walking tempo correlates with musical tempo), the hypotheses call for thorough data on phylogenetically diverse populations, and/or convergent evidence from e.g. experiments.

Otherwise, spurious correlations with little or no explanatory value may accumulate given the increasing availability of relatively large but unbalanced digital datasets (Roberts & Winters 2013).

Climate constitutes another popular source of determinants from the so-called-natural environment. Higher humidity and temperature have been related to a higher incidence of a number of interconnected phonological features, such as vocalic sounds (Everett 2017), sonority (Fought et al. 2004), CV syllables (Munroe & Silander 1999), lower phonological complexity (Maddieson et al. 2011), and tonal languages (Everett, Blasi & Roberts 2015). As in the case of the terrain biasing tempo via walking speed, climate is argued to have an indirect effect on the linguistic system. For instance, higher temperatures may increase the physical distance between speakers, and higher humidity may facilitate certain aspects of phonation; these, in turn, can promote some phonological features and not others.

The effect of the social niche on a verse system can be considered more direct, although proving its robustness is comparably challenging. For instance, even though statistical correlations have been found between sexuality-related values and phonetic/phonological aspects of language and music, the causal mechanism is not well understood. Societies with higher levels of extra-marital sex exhibit languages with greater phonological sonority (Ember & Ember 2007), a correlation comparable to the one between warm climate and sonority mentioned before. Similarly, the singing style of societies imposing severe sexual control on female pre-marital sex is associated with narrower and more nasal phonation (Lomax 1976:23). Nonetheless, the explanatory value of these observations remains limited, as reflected by the author's conjectural tone: "it is as if one function of song was to voice the level of sexual tension in culture" (Lomax 1976:262).

Cultural features tend to be inherited in bundles, which means that, if an early population (e.g. Proto-Bantu) shows a correlation between two features, it is likely that its descendant populations (in the case of Bantu, hundreds) will also show the correlation; not because of a causal link between the two, but as a result of inheritance. In order to increase the explanatory value of correlational studies, it becomes critical to tackle this issue (i.e. Galton's problem, Simonton 1975). For instance, Lupyan & Dale (2010:1) show that "languages spoken by large groups have simpler inflectional morphology than languages spoken by smaller groups as measured on a variety of factors such as case systems and complexity of conjugations." The correlation is robust within a number of independent linguistic families, controlling thus for the inheritance problem. Besides, they defend an explicit causal mechanism, namely that languages spoken by larger groups tend to have a higher proportion of speakers who learnt the language as adults, and that complex morphology has been shown to be easily acquirable by children, but challenging for adult learners.

Among the properties of the social niche most directly relatable (regarding causality) to features of verse, language plays a central role. More specifically, it is common among metricists to consider that "the possible versification systems for a language" can be derived "from its phonology" (Hanson & Kiparsky 1996:288). Following this line of thought, Golston & Riad (1997; 2000) analyse the metrics of of Arabic and Greek as substantially shaped by their phonological properties. Indeed, prosodic features such as stress are frequently mentioned as determining the shape of songs and poems: "the Czechs and Hungarians have languages in which utterances begin with stressed syllables, and their folk songs typically begin with stressed beats" (Nettl 2010:346). The effect has been shown even within the work of composers of instrumental music, where (1) there is no overt use of words, and (2) there is a particular emphasis on creativity, with its tendency to counteract structures inherited from tradition (Patel & Daniele 2003).

Finally, one can consider that the smallest niche with a potential to influence properties of verse is contained within each person's body, e.g. at the anatomical and cognitive levels. Phoneticians have produced a vast literature on how our anatomical features influence the sounds of speech, and, more relevant here, how these influences can explain typological tendencies in the world's phonological systems (for an overview, see Ohala 2010, Gordon 2016, and references therein). Linguists tend to accept that the average anatomical features of our species as a whole can account for many phonological generalisations in typology; nevertheless, they are often sceptical about anatomic variation as a source of phonological differences between populations:

> One idea is that languages change because their speakers have particular physical characteristics: short tongues, or thick lips, or gappy teeth, or something else in this vein. This is of course nonsense: the size of your tongue has no more to do with the way you speak than has the size of your feet. But that doesn't stop people making this kind of argument. (Trask 2010:19)

The preceding argument is made within the context of diachronic accounts of language change. Notwithstanding an understandable suspicion for simplistic anatomical explanations which can be incorporated too easily into racist discourses, large-enough datasets may provide evidence for subtle anatomical effects over long periods of time:

> very small, quantitative differences between populations in the anatomy of the vocal tract (here, the alveolar ridge) can influence the articulatory and acoustic properties of a class of sounds (here, post-dental/alveolar clicks) facilitating their inclusion in, and their further elaboration as part of, the phonological inventory, resulting in cross-linguistic variation. (Dediu, Janssen & Moisik 2017:16)

The interactive nature of most forms of speech (and verse to a lesser degree) leads to the conceptualisation of individuals' biases as a trade-off between two competing forces: "minimization of articulatory effort and maximization of perceptual distinctness" (Gordon 2016:17). Most forms of verse also exhibit this double dimension of production and perception, although the pressure for efficient, pragmatic transmission of the signal is not always as fundamental as in everyday conversation. To be sure, engaging in verse can be notoriously challenging in many traditions, e.g. by explicitly making the phonological parsing obscure or imposing complex formal restrictions on the choice of words (for an overview, see "The difficulty of poetry", Fabb 2015:63–71).

The fact that verse is made cognitively demanding for aesthetic or social reasons diminishes the effect of effort-minimising constraints (e.g. Zipf 1949), but all these explicit difficulties need to be framed within our cognitive capabilities nonetheless. Hence, the prevalence of particular verse features can be explained also by the cognitive niche, as it has been proposed for other aspects of culture more generally (Sperber & Hirschfeld 2004; Hauser 2009). Within the broad spectrum of cognition, it is particularly relevant to focus on traits which appear reliably in human development (Mameli & Bateson 2006). The framework of Core Knowledge Systems (Spelke & Kinzler 2007) describe some such traits, which arguably guide our basic representations of e.g. objects, number or space, and which "may provide some of the foundations for uniquely human cognitive achievements, including the acquisition of language and other symbol systems" (Spelke & Kinzler 2007:92). In Chapter 2, for instance, I argue that limits on our numeric cognition (Feigenson, Dehaene & Spelke 2004; Mandler & Shebo 1982) constrain the size of verse constituents.

All the different factors mentioned above interact in a variety of ways, as evidenced generally by the fact that humans adapt to, *and* modify, their environment (Laland, Odling-Smee & Myles 2010). In order to address these factors and their interactions systematically, we can explain their causal effect at different temporal levels. Vygotsky (1962) and, more recently, Smith, Brighton & Kirby (2003) propose a triparte division between effects occurring in the process of (1) biological evolution, (2) cultural evolution, and (3) ontogeny. Even smaller temporal levels can be distinguished, such as effects active at the level of milliseconds during online production and processing of words (Rączaszek-Leonardi 2010). As Enfield (2014) argues, these and similar proposals provide “conceptually distinct but interconnected causal frames”, which can be useful to understand a phenomenon such as verse. This dissertation focuses on a temporally low-level frame, that is, on causes related to cognitive processing.

1.3 Outline of the dissertation

The dissertation is divided into three parts, each containing a corpus-based chapter followed by an experimental chapter. Each of the parts tackles a different aspect of verse deemed to be typologically widespread. The corpus-based chapter of each part focuses on how we can derive structural tendencies by analysing verse productions which have already been documented (in collections of poems, songs, or descriptions of metrical systems). The experimental chapters try to replicate these observations using controlled verse-like stimuli. In this way, I elucidate the extent to which aspects of general cognition can explain the patterns described in the corpora.

The first part deals with the cornerstone of verse: the structure of the templates. I review (in Chapter 2) the strategies used in different languages to demarcate constituents such as stanzas and lines. Based on these chunking strategies, I examine the numeric demands imposed by templates, reflected in the number of objects contained within each constituent. I conclude that, when the number of identical adjacent objects within a constituent exceeds four, an additional subgrouping is created. This limit may be related to subitizing and working memory constraints. Further (Chapter 3), I demonstrate that this kind of chunking can emerge spontaneously when subjects are asked to reproduce long, random sequences of syllables. Using an iterated learning paradigm, we observe that the small chunking tendencies showed by each person get amplified in the process of transmission, thus offering a model for the cultural emergence of verse-like structures.

The second part tackles the issue of how templates are realised into actual poems or songs; more precisely, I investigate patterns of deviations from a template. First, I offer statistical support for a widely held proposal that verse lines deviate more often at the beginnings rather than towards the end (Chapter 4). Second, I show experimentally that this asymmetry also holds for the perception of verse-like sound sequences: subjects take longer to detect deviant tones towards the beginning of the sequences (Chapter 5). These results can be explained by theories of dynamic attention and prediction, according to which later events benefit of the regularity of the preceding ones to build more accurate representations, and hence make it less likely for poets to produce deviations, and for subjects to miss them.

The third and final part also deals with the realisation of templates, but in the specific case of words being set to musical tunes (i.e. textsetting). A number of previous studies have described how phonological features of a language such as stress or tone are aligned with musical features such as metrical prominence or melodic contours. I propose a computational method to address this alignment issue in a systematic way, as a means to ease, standardise and encourage comparisons across genres and languages (Chapter 6). The chapter takes a large corpus of Dutch folk songs as a case study, showing that opposite stress and prominence contours are strongly avoided, particularly within content words and when no melismas are involved. An important limitation of most approaches to textsetting is that they rely on introspection or song corpora. I address this issue by developing a simple perception experiment (Chapter 7) which supplies reliable and detailed textsetting judgements from a given language community, i.e. Dutch, providing converging evidence for the corpus-based hypotheses from the preceding chapter.

References

Blacking, J. 1965. The role of music in the culture of the Venda of the Northern Transvaal. *Studies in Ethnomusicology* 2. 20–53.

Blacking, J. 1967. *Venda children's songs: a study in ethnomusicological analysis*. University of Chicago Press.

Blacking, J. 1974. *How musical is man?* Seattle: University of Washington Press.

Brown, S. & J. Jordania. 2011. Universals in the world's musics. *Psychology of Music* 41(2). 229–248.

Changizi, M. 2011. *Harnessed: how language and music mimicked nature and transformed ape to man*. BenBella Books.

Christiansen, M. H. & S. Kirby (eds.). 2003. *Language evolution*. Oxford University Press.

Darwin, C. 1871. *The descent of man, and selection in relation to sex*. London: John Murray.

Dediu, D., R. Janssen & S. R. Moisik. 2017. Language is not isolated from its wider environment: vocal tract influences on the evolution of speech and language. *Language & Communication* 54. 9–20.

Ember, C. R. & M. Ember. 2007. Climate, econiche, and sexuality: influences on sonority in language. *American Anthropologist* 109(1). 180–185.

Enfield, N. J. 2014. *Natural causes of language: frames, biases, and cultural transmission*. Language Science Press.

Everett, C. 2017. Languages in drier climates use fewer vowels. *Frontiers in Psychology* 8.

Everett, C., D. E. Blasi & S. G. Roberts. 2015. Climate, vocal folds, and tonal languages: connecting the physiological and geographic dots. *Proceedings of the National Academy of Sciences* 112(5). 1322–1327.

Fabb, N. 2014. The verse-line as a whole unit in working memory, ease of processing, and the aesthetic effects of form. *Royal Institute of Philosophy Supplement* 75. 29–50.

Fabb, N. 2015. *What is poetry? Language and memory in the poems of the world*. Cambridge University Press.

Feigenson, L., S. Dehaene & E. Spelke. 2004. Core systems of number. *Trends in cognitive sciences* 8(7). 307–314.

Fitch, W. T. 2015. Four principles of bio-musicology. *Phil. Trans. R. Soc. B* 370(1664). 20140091.

Fought, J. G., R. L. Munroe, C. R. Fought & E. M. Good. 2004. Sonority and climate in a world sample of languages: findings and prospects. *Cross-Cultural Research* 38(1). 27–51.

Golston, C. & T. Riad. 1997. The phonology of Classical Arabic meter. *Linguistics* 35(1). 111–132.

Golston, C. & T. Riad. 2000. The phonology of Classical Greek meter. *Linguistics* 38(1). 99–167.

Gordon, M. K. 2016. *Phonological typology*. Oxford University Press.

Hanson, K & P Kiparsky. 1996. A parametric theory of poetic meter. *Language* 72(2). 287–335.

Hauser, M. D. 2009. The possibility of impossible cultures. *Nature* 460(7252). 190–196.

Hoeschele, M., H. Merchant, Y. Kikuchi, Y. Hattori & C. ten Cate. 2015. Searching for the origins of musicality across species. *Phil. Trans. R. Soc. B* 370(1664). 20140094.

Honing, H., C. ten Cate, I. Peretz & S. E. Trehub. 2015. Without it no music: cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370(1664). 20140088–20140088.

Hymes, D. 1977. Discovering oral performance and measured verse in American Indian narrative. *New Literary History* 8(3). 431–457.

Laland, K. N., J. Odling-Smee & S. Myles. 2010. How culture shaped the human genome: bringing genetics and the human sciences together. *Nature reviews. Genetics* 11(2). 137–148.

List, G. 1963. The boundaries of speech and song. *Ethnomusicology* 7(1). 1–16.

Lomax, A. 1976. *Cantometrics: an approach to the anthropology of music: audiocassettes and a handbook*. Berkeley: University of California Extension Media Center.

Lupyan, G. & R. Dale. 2010. Language structure is partly determined by social structure. *PloS one* 5(1). e8559.

Lupyan, G. & R. Dale. 2015. The role of adaptation in understanding linguistic diversity. In R. De Busser & R. J. LaPolla (eds.), *Language structure and environment: social, cultural, and natural factors*, vol. 6, 289–316. John Benjamins Publishing Company.

Maddieson, I., T. Bhattacharya, D. E. Smith & W. Croft. 2011. Geographical distribution of phonological complexity. *Linguistic Typology* 15(2). 267–279.

Mameli, M. & P. Bateson. 2006. Innateness and the sciences. *Biology and Philosophy* 21(2). 155–188.

Mameli, M. & P. Bateson. 2011. An evaluation of the concept of innateness. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 366(1563). 436–443.

Mandler, G. & B. J. Shebo. 1982. Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General* 111(1). 1.

Munroe, R. L. & M. Silander. 1999. Climate and the consonant-vowel (cv) syllable: a replication within language families. *Cross-Cultural Research* 33(1). 43–62.

Nettl, B. 2010. *The study of ethnomusicology: thirty-one issues and concepts*. University of Illinois Press.

Odling-Smee, J. F., K. N. Laland & M. W. Feldman. 2003. *Niche construction. The neglected process in evolution*. Princeton University Press.

Ohala, J. J. 2010. The relation between phonetics and phonology. In W. J. Hardcastle, J. Laver & F. E. Gibbon (eds.), *The handbook of phonetic sciences, second edition*, 653–677. Wiley-Blackwell.

Patel, A.-D. & J.-R. Daniele. 2003. An empirical comparison of rhythm in language and music. *Cognition* 87. B35–B45.

Patel, A. D. & S. M. Demorest. 2013. Comparative music cognition: cross-species and cross-cultural studies. In D. Deutsch (ed.), *The psychology of music*, 647–681. Elsevier.

Rączaszek-Leonardi, J. 2010. Multiple time-scales of language dynamics: an example from psycholinguistics. *Ecological Psychology* 22(4). 269–285.

Richter, J. & R. Ostovar. 2016. "It don't mean a thing if it ain't got that swing"– an alternative concept for understanding the evolution of dance and music in human beings. *Frontiers in human neuroscience* 10.

Roberts, S & J Winters. 2013. Linguistic diversity and traffic accidents: lessons from statistical studies of cultural traits. *PLoS ONE* 8(8). 1–13.

Seeger, A. 2004. *Why Suyá sing. A musical anthropology of an Amazonian people*. Cambridge University Press.

Simonton, D. K. 1975. Galton's problem, autocorrelation, and diffusion coefficients. *Behavior Science Research* 10(4). 239–248.

Smith, K., H. Brighton & S. Kirby. 2003. Complex systems in language evolution: the cultural emergence of compositional structure. *Advances in Complex Systems* 6(04). 537–558.

Spelke, E.-S. & K.-D. Kinzler. 2007. Core knowledge. *Developmental Science* 10(1). 89–96.

Sperber, D. & L. A. Hirschfeld. 2004. The cognitive foundations of cultural stability and diversity. *Trends in cognitive sciences* 8(1). 40–46.

Teie, D. 2016. A comparative analysis of the universal elements of music and the fetal environment. *Frontiers in Psychology* 7. 1–8.

Trask, R. L. 2010. *Why do languages change?* Cambridge University Press.

Tudor-Locke, C. E. & A. M. Myers. 2001. Methodological considerations for researchers and practitioners using pedometers to measure physical (ambulatory) activity. *Research quarterly for exercise and sport* 72(1). 1–12.

Vygotsky, L. S. 1962. *Thought and language*. Cambridge, MA: MIT press.

Wallin, N.-L., B Merker & S Brown (eds.). 2001. *The origins of music*. Massachusetts Institute of Technology.

Wrangham, R., J. H. Jones, G. Laden, D. Pilbeam & N. Conklin-Brittain. 1999. The raw and the stolen: cooking and the ecology of human origins. *Current Anthropology* 40(5). 567–594.

Zipf, G. K. 1949. *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge, MA: Addison-Wesley Publishing.

Part I

How verse templates are built

2 Numeric control in verse constituent structure

2.1 Introduction

This chapter investigates verse constituent structure from a typological perspective. Songs and poems typically follow some abstract template like a stanza form composed of two couplets, where each line within the couplet is divided into four musical beats. Regularities such as the repetition of a melody, the placement of beats, stressed syllables and pauses, or the rhyming patterns make this kind of segmentation transparent. Here, I provide an overview of segmentation cues used in verse templates from different traditions, and discuss how they can inform the problem of numeric control during the creation of verse. By numeric control I refer to the fact that verse constituents often require a discrete number of sub-sections, such as four-line stanzas or seven-syllable lines. This case study illustrates the kind of research questions which can benefit from thorough typological data on verse traditions. Besides, numeric control relates more generally to aspects of human cognition such as working memory constraints, which can be helpful in explaining the observed typological data. Conversely, limits in the variation of verse forms can be informative for theories of numeric cognition.

2.1.1 Segmentation in music analysis and cognition

The study of segmentation plays a prominent role in the fields of music analysis and cognition. Most studies focus on instrumental music, and on how *listeners* make sense of the acoustic input, i.e. which cues they use to chunk continuous event sequences into discrete, meaningful sections. This chapter, instead, addresses the issue of segmentation from the perspective of the creator of verse. Although not identical, the cues used in perception are likely to overlap with those meaningful during production.

In their influential book on the analysis of Western tonal music, Lerdahl & Jackendoff (1983:13) hold that "grouping can be viewed as the most basic component of musical understanding." By analysing a range of pieces of tonal music and their own intuitions on how these are parsed, the authors put forward a collection of well-formedness and preference rules for the segmentation and grouping of music stimuli. For instance, part of the second Grouping Preference Rule states that, in "a sequence of four notes $n_1 n_2 n_3 n_4$ (...) the transition $n_2 n_3$ may be heard as a group boundary if the interval of time between the attack points of n_2 and n_3 is greater than that between the attack points of n_1 and n_2 and that between the attack points of n_3 and n_4" (Lerdahl & Jackendoff 1983:45). Some rules are music-specific, while others resemble more generally Gestalt-type principles of perception, such as the cited rule on the proximity of adjacent events. The *implication-realization model* by Narmour (1990) offers a similar kind of approach to an ideal listener's grouping of musical inputs, but concentrating specifically on the pitch-structure.

In *The cognition of basic musical structures*, Temperley (2001) develops a comparable formalisation of grouping based on preference rules, and he implements and tests them computationally. After this kind of bridge work between theoretical and computational approaches, a wealth of studies have emerged within the field of music information retrieval (Burgoyne, Fujinaga & Downie 2015). These have offered the possibility of grounding empirically introspection-based proposals, and further refining them. Brown, Gifford & Davidson (2012), for instance, implement the principles of melodic closure proposed by Narmour (1990) and test the algorithms against a corpus of more than 3000 folk tunes. Nieto & Bello (2016) evaluate the performance of eight previously-published computational algorithms designed to analyse the constituent structure of music audio recordings. A lengthier overview of how the music information retrieval community tackles the problem of segmentation is provided by Rodríguez López (2016), with a thorough taxonomy of the main segmentation cues treated in the literature.

Besides computation, the main empirical grounding for segmentation cues comes from music psychology studies. Deutsch (2013, and other contributions within the volume) discusses state-of-the-art experiments on music segmentation, and how they relate to theoretical proposals such as those by Lerdahl & Jackendoff (1983) and Narmour (1990). Besides providing experimental evidence for well-established grouping cues such as pitch proximity, perception studies have also enabled the demonstration of otherwise difficult to proof principles, e.g. that listeners heavily rely on transition probabilities between adjacent notes to abstract motifs which are readily identifiable in future input (Saffran et al. 1999).

The study of verse constituent structure can draw important insights from this kind of research on music segmentation, particularly because most verse traditions are sung. Nevertheless, a study of verse segmentation needs to take into account its linguistic features as well. As a matter of fact, the main available source of evidence for grouping and segmentation in verse are structural analyses of the verbal material in poetic corpora, from which researchers infer a set of abstract templates (also known as *poetic metres*). These are meant to capture the regularities found in the texts, and reflect the emic structures employed in a productive way by the original creators. Studies on the verse of European languages tend to focus on a genre (e.g. the French alexandrine, Dominicy 1992), but other studies may deal with individual songs (e.g. a Havasupai *Origin Song*, Hinton 1990), or with the (virtually) complete verse repertoire of a language (e.g. the songs of the Pintupi, Moyle 1979).

Besides language-specific studies, there have been a number of attempts at summarising the range of possible principles used in verse traditions across the world (Lotz 1960; Fabb 1997; Aroui 2009). These provide insightful taxonomies to understand the basis on which verse constituents *can* be defined. That is comparable to the traditional goal of linguistic typology, i.e. "to determine the limits of possible human languages" (Bickel 2007:239). More recently, however, "instead of asking 'what's possible?', more and more typologists ask 'what's where why?'" (Bickel 2007:239). A comprehensive typology of verse constituent structure, thus, needs to examine balanced samples of verse traditions, and address issues of geographical and genealogical skewing both from a synchronic and a diachronic perspective. Within this context, the present study focuses exclusively on synchronic explanations to patterns which appear to follow a global skew.

2.1.2 Typological approaches to verse

Compared to other aspects of language, such as the syntax or phonology of everyday speech, versification systems have not received as much worldwide typological investigation. There are several reasonable explanations for this. First, verse arguably plays a marginal role in all societies compared to speech. Second, verse productions are more prone to particular kinds of creativity which reflect the inventiveness and idiosyncrasies of an individual singer, composer or poet, rather than a core set of conventional structures shared across the community. Compare, for instance, the phonological inventory or number system of a language, with its inventory of melodies or metrical patterns; the latter tend to be open-ended and more unstable. Third, partly due to the previous two issues, there has been less theorization in the field of metrics, which may have hindered the development of comparative concepts necessary in a typological enterprise (Haspelmath 2010).

Nonetheless, some attempts have been made to capture patterns of stability and variation across verse traditions. The ethnomusicologist Alan Lomax led a seminal enterprise in the 1960s to study song systems from a comparative perspective. The ambitious Cantometrics project (Lomax & Grauer 1968; Lomax 1976) surveyed a sample of more than 4000 songs from ca. 400 cultures. Each song was analysed by completing a questionnaire covering 37 features, related to aspects such as the performers (e.g. presence of soloists, chorus, instruments), rhythm (e.g. simple, complex, irregular metre), melodic shape and range (e.g. arched, terraced), or phonation features (e.g. falsetto, nasalized).

Although this kind of comparative studies had been largely abandoned (Patel & Demorest 2013), the original Cantometrics approach has received renewed attention in the twenty-first century. The CantoCore project (Savage et al. 2012; 2015b) builds upon the same questionnaire-based methodology, improves the inter-rater reliability, and omits performance and instrument-related features to focus on 26 aspects related to the structure of vocal songs. Some of these features are potentially useful for a typological study of constituent structure, e.g. feature #2 (number of beats in a bar), or feature #21 (phrase repetition). However, their usefulness for the purposes of this chapter is limited by two methodological issues. First, for each question, only the maximal value found in the song is chosen. Second, the responses are expected to fall within a discrete number of categories: e.g. feature #2 (number of beats in a bar) takes three possible values: duple, triple, and complex (i.e. "the number of beats can only be divided by prime numbers greater than 3" Savage et al. 2012:95). On the one hand, there is a general granularity issue, i.e. by accepting a single, maximal value per song, we are capturing the upper limits and ignoring within-song heterogeneity. On the other hand, employing pre-defined, closed sets of feature values interferes with typological aims:

> Traditional typological databases first define a set of crosslinguistic types in a functionally defined domain (...) and then assign each language in the sample to a type. One problem of this procedure is that the initial definitions make it unlikely that hitherto unknown types can ever be discovered: if a type does not fit the original scheme, it is usually treated as a transitional phenomenon, although outside the a priori typology it may as well be a primitive type of its own. (Bickel & Nichols 2002)

Although not ideal for the research question at hand, datasets using the CantoCore methodology prove to capture an adequate amount of inter-population variation. This has enabled the use of verse data to address questions of population migration (Brown et al. 2014; Savage et al. 2015a). Further along this method-

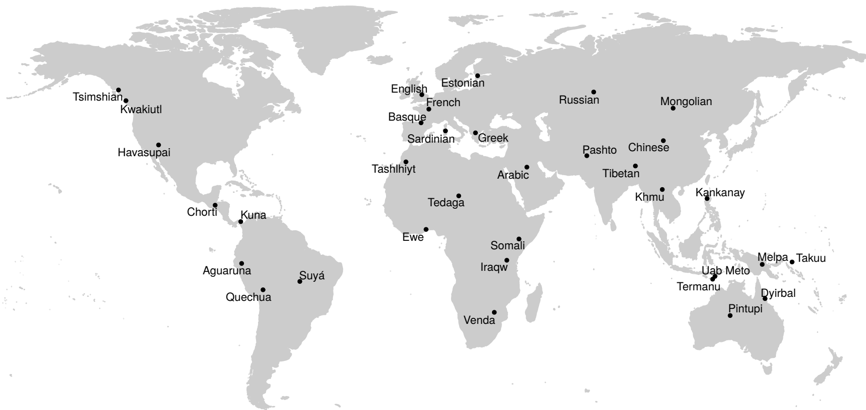


Figure 2.1: Languages discussed in the chapter.

ological line, Le Bomin, Lecointre & Heyer (2016) exploit 322 features (mostly binary, encoding presence or absence) to demonstrate the vertical transmission of musical features among 58 populations from Gabon.

In addition to musicological surveys, linguists and literary scholars have long compiled repertoires of metrical templates, focusing on European traditions. These typically overlook musical aspects, but, unsurprisingly, offer more rigorous analyses of phonological features such as stress and vowel length. Recently, efforts are being made to integrate datasets from different languages, opening the possibility for comparative work (González-Blanco García, Manailescu & Ros 2016). A promising pilot database within the linguistic approach to verse is Versetyp (Versace & Fabb 2011). It follows a typological-questionnaire methodology comparable to the Cantometrics, although the fields are less restricted, hence being more prone to capturing variation.

The data I discuss in this chapter, combined with the lessons from these typological projects, may prove useful for a future collection of constituent structure data. Such an enterprise can be minimally divided into two modules. First, a prerequisite would be to compile a database of available published resources on the verse traditions of the world's languages (this can be modelled e.g. after the Glottolog repository of linguistic references, Nordhoff & Hammarström 2011). Second, a coherent way of representing verse templates is needed, in order to capture the complete constituent structure and the feature specifications associated to each constituent (e.g. Humphrey et al. 2014 for a convenient JSON-based approach).

The remainder of the chapter is divided into two sections, closed by a general discussion. In Section 2.2 I provide a catalogue of cues or strategies used in different traditions to demarcate verse constituents. The data is based on sources about the versification system of several languages from all major landmasses of the world. The approximate location of the main languages discussed in the text are shown in Figure 2.1, and a complete list with additional details is available in Table 2.1 of the Supplementary Information.

The sampling of languages is not by any means comprehensive, yet it attempts to maximise the genealogical coverage given the available bibliographical sources. In the process of selecting languages, I check whether verse templates are reported, or, in the lack of templates, whether analyses of individual songs are provided. The description of templates is usually preferred because, arguably, it reflects more accurately the cognitive structure and principles used to create the songs. In Section 2.3, I discuss the range of numeric control requirements observed in the analysed templates. Finally (Section 2.4), I propose that the way in which verse constituents are segmented, and their numeric demands, reflect limits on working memory, and how chunking strategies can alleviate them.

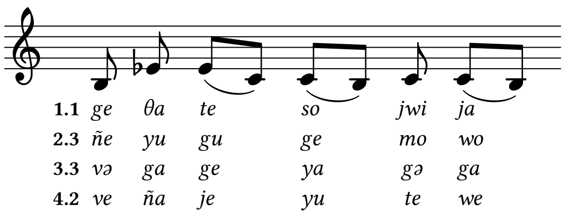
2.2 Segmentation cues in verse

The segmentation cues observed in the data can be grouped into three categories: (1) parallelism, (2) contrast, (3) boundary marking. Probably the most widespread example of parallelism is melodic parallelism, that is, the repeated use of a melodic contour with different words attached to it. Boundary marking can overlap with parallelism, as in the case of line-final rhyme, where some degree of phonological parallelism is required for the rhyming segments to fulfil their bounding function. The use of contrast can also be used in a parallel fashion, e.g. when each stanza alternates a solo line and a group line. Nonetheless, these two kinds of cues can also work independently of parallelism, by relying on locally detectable features, such as the lengthening of the line-final syllable.

2.2.1 Parallelism

Parallelism establishes an identity relation between two or more sections of text. As such, parallelism is inherent to the notion of verse template, as it always relies on the reiteration of some structural pattern. Jakobson (1966a:399) opens a major article on poetic parallelism with the following quote by the nineteenth-century poet Gerard Manley Hopkins: “The artificial part of poetry, perhaps we shall be right to say all artifice, reduces itself to the principle of parallelism.” This cen-

Example 2.1: Melodic parallelism in a Havasupai *Sweathouse Origin Song* (Hinton 1984:236). All four lines have different texts but are set to the same tune. Numbers correspond to the line identifiers in the original publication.



trality of parallelism as a defining feature of verse is also highlighted by Anthony Seeger (2004:45) when he places seven categories of Suyá [Nuclear-Macro-Je; Je][1] verbal production on a continuum according to their level of fixity or parallelism, starting with free, everyday speech, and ending with entirely fixed songs based on extensive repetition and parallelism.

2.2.1.1 Melodic parallelism

Melodic parallelism is extremely common. Whenever a tune is used repeatedly with different words attached to it we can talk of melodic parallelism. As with all kinds of parallelism, we can be more easily aware of it if the parallel sections are adjacent. Strophic songs are a widespread example, and it is straightforward to locate in song collections because most often the musical notation is given only for one of the strophes, and for the remainder just the words are provided. This kind of presentation of strophic songs with melodic parallelism can be observed in the Havasupai [Cochimi-Yuman; Yuman] *Sweathouse Origin Song* (Hinton 1984; 1990), in the Northern Kankanay [Austronesian; Nuclear Austronesian] songs transcribed by Maceda (1958), or in the study of Tashlhiyt [Afro-Asiatic; Berber] versification by Dell & Elmedlaoui (2008). Example 2.1 shows four sample lines from the Havasupai *Sweathouse Origin Song*, all set to the same tune.

Another context where melodic parallelism is very clearly manifest is improvised verse. In Basque [Isolate] (and in many other languages, e.g. Yaqub 2007 for Palestinian Arabic [Afro-Asiatic; Semitic], Zedda 2009 for Sardinian [Indo-European; Italic]), a singer chooses a melody from a conventional repertoire and uses it as a placeholder to produce new strophes (Egaña 2007). This can also be done in the form of a duel, where the melody is used in alternation by two differ-

[1] In the first mention of a language, I include its genealogical information between brackets, extracted from Glottolog (Nordhoff & Hammarström 2011).

ent singers. The complete text can be easily divided tracking each repetition of the tune.

Melodic parallelism can also mark the internal structure of a strophe by establishing identity relations between lines. Verse scholars often make explicit these kinds of parallel structure by describing the formal composition of a song using codes such as *aabb*, where each of the letters corresponds to the melodic contour of a line. Some of the many languages for which lines with parallel melodies have been reported are Melpa [Nuclear Trans New Guinea; Central East New Guinea Highlands] (Niles 2011), Suyá (Seeger 2004), or Venda [Atlantic-Congo; Volta-Congo] (Blacking 1969). From a production perspective, creating a stanza with the melodic structure *aabb* can be conceived as creating two groups of two lines with parallel melodies instead of four disconnected lines.

An important issue concerning parallelism in general and melodic one in particular is the fact that the link between the terms is one of similarity; that is, identity with some difference. In the Melpa *amb kenan* courting song, the central section of each stanza is divided into two halves, where "the first half of the melody is partially tonally transposed a step lower in the second part" (Niles 2011:281). The melodic parallelism between the two sections provides a cue to group them together and differentiate them from the preceding or following material; however, the fact that they are similar but not *identical* makes it possible to set apart the two sections as different objects (a_1a_2), and not as a reiteration of the same object ($a \times 2$).

The Melpa kind of parallelism-with-a-difference can be considered to be encoded in the template, since it remains constant throughout the song. In other cases, though, differences in melodic detail of otherwise parallel sections can be attributed to the instantiation of the template; hence, these differences cannot play a structural role in the planning of a new text. For instance, every line of the Venda *domba* female initiation song uses the same basic melody, but with variations: "the melody of the last five syllables of each word-phrase remains the same throughout, but that of the first four (or three, five or six) syllables varies according to changing patterns in the speech-tones of the words" (Blacking 1969:256). The lack of complete sameness between the lines, thus, is not employed as a grouping or segmentation cue, but rather derives from the process of setting new words to the same melody.

2.2.1.2 Rhythmic parallelism

In many cases, the melodic structure of songs and poems is either unspecified or unreported. Verse classified as spoken poetry (of the kind well-known in European languages), for instance, does not employ melodic parallelism. Chanted verse such as that present in nursery rhymes often do not require fixed pitch sequences either, but follow intonational contours similar to those used in speech. The length of a melody very frequently determines the length of verse lines; in the absence of melodies, rhythmic parallelism can serve this purpose. By rhythmic parallelism I refer to the formal identity in the number of temporal units, be it syllables, linguistic accents, or musical beats, among others.

Syllabic poetry constitutes a simple case of this kind of parallelism. Each line of French [Indo-European; Italic] classical alexandrine contains twelve full syllables, plus an optional schwa (Dominicy 1992:161). This equal number of placeholders per section provides parallel templates where the words can be attached to. In Classical Tibetan [Sino-Tibetan; Bodic], one of the most common forms is the heptasyllabic quatrain, where each line consists of seven syllabic placeholders (Poucha 1950:196). The rhythmic parallelism in classical Tibetan is more detailed; beyond mere placeholders, there is evidence that lines tend to share the placement of stress. Tibetan shows fixed word-initial stress, and most lines are made up of two bisyllabic words followed by a trisyllabic one, giving the general structure (σ́σ)(σ́σ)(σ́σσ) (Vekerdi 1952:223). Besides the syllabic parallelism, there is evidence of stress parallelism too.

This brings us to a generalisation about rhythmic parallelism. The rhythmic identity between sections relies on an equal number of placeholders; these, however, can be defined in several ways. A placeholder can correspond to a simple phonological constituent like the syllable, it can be further specified as a stressed or heavy syllable, or it can correspond to a non-phonological category such as the musical beat. The key principle is that the size of the parallel sections is controlled by a template where a number of placeholders are required. A melody can also control the size of the constituents; nonetheless, we categorise melodies differently (i.e. as melodic parallelism) for two reasons. First, the identity of a melody can be recognized and used as template in spite of variations in the number of notes or placeholders, as long as certain interval and contour properties are preserved (Deutsch 2013b). Second, each placeholder in a melodic template can take many more values (i.e. notes of a scale) than rhythmic placeholders, where mostly unary (syllabic placeholder) or binary (heavy or light syllable) features are employed.

Example 2.2: First three stanzas of the song *Fast car* by Tracy Chapman, with number of syllables per line. Each line is set to two four-beat musical bars, but the number of syllables per line varies throughout.

5 *You got a fast car*
9 *I want a ticket to anywhere*
6 *Maybe we make a deal*
10 *Maybe together we can get somewhere*
6 *Anyplace is better*
10 *Starting from zero got nothing to lose*
6 *Maybe we'll make something*
9 *Me, myself I got nothing to prove*

5 *You got a fast car*
10 *I got a plan to get us out of here*
10 *I been working at the convenience store*
12 *Managed to save just a little bit of money*
6 *Won't have to drive too far*
11 *Just 'cross the border and into the city*
7 *You and I can both get jobs*
12 *And finally see what it means to be living*

9 *You see my old man's got a problem*
11 *He live with the bottle that's the way it is*
10 *He says his body's too old for working*
9 *His body's too young to look like his*
8 *My mama went off and left him*
10 *She wanted more from life than he could give*
11 *I said somebody's got to take care of him*
9 *So I quit school and that's what I did*

Controlling the length of sections with an equal number of beats is a very common type of rhythmic parallelism. Take, for instance, the song *Fast car* by Tracy Chapman. The first three stanzas are shown in Example 2.2, each line preceded by the number of syllables it contains. All three stanzas follow the same template, defined, among other things, by being composed of eight lines of the same duration: one musical bar with four beats. This temporal placeholder is used to set new lines of text, which can be composed of a varying number of syllables. Lines contain between 5 and 12 syllables, and no single line shows the same number of syllables across the three stanzas.

Nursery rhymes in many languages employ beat-number parallelism without syllable-number parallelism (Burling 1966). Very frequently, lines have a length of four beats but an unequal number of syllables, such as *Hickory dickory dock*, with lines of 6, 7 or 8 syllables set to the constant four-beat structure. Still, it is also common to employ both beat-number and syllable-number parallelism, such as in *Eeny meeny miny moe*. The tunes used in the traditions of Basque improvised verse (Egaña 2007) and Tashlhiyt singing (Dell & Elmedlaoui 2008) also follow both types of parallelism.

Finally, there are verse traditions where evidence for a musical beat is lacking, where the number of syllables per line is not constant, but where we can infer the use of rhythmic parallelism. Medieval poetry in Germanic languages such as Dutch, English or German, the number of stressed syllables per line is kept constant, with freedom with respect to the occurrence of unstressed syllables (Gasparov 1996). To be sure, this is very similar to beat-parallelism (cf. Example 2.2 by Tracy Chapman), but with no evidence of isochrony between stresses, due to shortage of documentation on how these texts were performed.

2.2.1.3 Semantic and syntactic parallelism

In the field of metrics, the term *parallelism* usually refers to the juxtaposition of semantically paired constituents: "perhaps the most familiar and recognized variety of parallelism is semantic parallelism, especially at the level of the metrical line in relation to adjacent verses" (Frog 2014:12). We can track this association between *parallelism* and *semantic identity* back to the 18th century, when Robert Lowth presented an analysis of ancient Hebrew poetry as based primarily on semantically linked pairs of lines (Jakobson 1966b:399). This work initiated a line of research which has since analysed a wide variety of verse traditions in terms of semantic parallelism (see Fox 2014a for a thorough discussion, on which I rely for this section).

Geographic areas where this kind of verse composition seems most pervasive (or, at least, has attracted more scholarly attention) include Central America (e.g. Mayan), central Eurasia (e.g. Turkic, Mongolic), and Melanesia (e.g. Austronesian). Still, "the extent of the literature and the frequency with which canonical parallelism has been cited in different oral traditions suggest it is a phenomenon of near universal significance" (Fox 2014a:29).

Even if this kind of parallelism is reported for many languages, its spread is not as extensive as other putative universals such as the use of isochronous metres or melodic phrases (Brown & Jordania 2011). For instance, regarding Indo-European, one of the largest language families of the world, Jakobson (1966b:405) argues that Russian [Indo-European; Balto-Slavic] folk poetry is the "only living oral tradition (...) which uses grammatical parallelism as its basic mode of concatenating successive verses." It is worth noting that Jakobson refers here to *canonical* parallelism: "poetic patterns where certain similarities between successive verbal sequences are compulsory or enjoy a high preference" (Jakobson 1966b:399). As with other structuring strategies, semantic parallelism can also be used in a less strict way "as a device for creating cohesion within sequences" (Frog 2014:15).

A well-researched verse tradition where this kind of parallelism represents

Example 2.3: Parallel couplets in Termanu (Rotinese, Austronesian), taken from a funeral chant by Stefanus Adulanu (Fox 2014b:104).

	Termanu text	**English translation**
1	*Besak-ka ana lino ba'e*	Now she rests on a branch
	Ma sa'e ndanak.	And perches on a limb.
2	*De ana kukuta Dela Kolik*	She continues to munch Dela Kolik
	Ma mumumu Seko Bunak.	And continues to suck Seko Bunak.
3	*De na,a na-mada man*	She eats to dry her tongue
	Ma ninu na-meti apen-na.	And drinks to slake her thirst.
4	*De henu tein-na boe*	To fill her stomach
	Ma sofe nutun-na boe.	And to satisfy her gizzard.

the main composition strategy is that of Rote Island, Indonesia (Fox 2014b). In Example 2.3 I reproduce eight lines from a funeral chant in the language of Termanu [Austronesian; Nuclear Austronesian], organised into parallel couplets (Fox 2014b:104). The couplets are not only parallel semantically, but they also follow equivalent syntactic structures; this is evident, for instance, in the translation of the third couplet, where the structure of each line can be represented schematically as: [verb phrase] *to* [verb] *her* [noun]. Indeed, both kinds of parallelism appear to overlap very frequently, and, particularly, verse traditions for which syntactic parallelism is described, semantic parallelism is also always mentioned (e.g. Sherzer 1982 for Kuna [Chibchan; Core Chibchan], Lundström 1984 for Khmu [Austroasiatic; Khmuic], and Hull 2003 for Chorti [Mayan; Core Mayan]). According to Fox (2014a:30): "syntactic and semantic parallelism are often distinguished although they are just as often intimately related. In his use of the term *grammatical parallelism*, Jakobson attempts to encompass both."

Rotenese parallel couplets are based on dyads, i.e. pairs of semantically-related words which tend to be used together, such as 'branch'–'limb', 'munch'–'suck', or 'stomach'–'gizzard' (Example 2.3). Words may have more than one potential pair, effectively creating a semantic network from which to choose from to compose couplets. For instance, the word for 'stomach' is linked to 'gizzard', but also to 'breast', which can pair with 'womb'. These semantic networks are highly constrained, and using a non-conventional pairing is "immediately detectable and—among Rotenese—quickly challenged" (Fox 2014a:39). Yet more restrictedly, in the neighbouring island of Timor, Uab Meto [Austronesian; Nuclear Austronesian] speakers use a similar kind of parallelism, where "not only are the words which can form doublets fixed, but the order in which each member of a doublet occurs is also fixed" (Edwards 2016:329).

2.2.1.4 Repetition

All examples of parallelism reviewed so far involve an identity link between some, but not all, features of the parallel sections. Melodic and rhythmic parallelism may keep an invariant pitch or temporal structure while introducing new statements; grammatical parallelism may reiterate a single statement with a variation in wording. In this sense, the ultimate degree of parallelism is attained via exact repetition of both form and content.

The use of systematic repetition as a productive device appears infrequently compared to the other strategies. It is reported, for instance, in Coast Tsimshian [Tsimshian] verse, such as the *Great blue heron* children's song, where the first half (four musical bars) consists of an exact repetition of a two-bar line, and the second half introduces a different line (Mulder 1994:106). The relative scarcity of allusions to exact repetition may derive from the fact that repetitions are often considered an aspect of performance, and not inherent to the template underlying the text. Hence, some transcriptions omit the repetitions altogether to focus on the variable aspects of the performance (Hymes 1981), which makes it particularly challenging to evaluate systematic patterns of exact repetition and their geographic distribution.

Rhyme also creates parallel identity relations where a set of sounds are repeated among two or more sections of verse. Nevertheless, rhyme differs from the repetition of words in that (at least some of) the parallel sounds involved in rhyme typically belong to different morphemes (Aroui 2005:180). Besides being a parallel strategy, rhyming also typifies boundary marking strategies, and will be further discussed in Section 2.2.3.

2.2.2 Contrast

Parallelism often involves repetition with a difference, which implies an overlap with the principle of contrast. In Northern Ewe [Atlantic-Congo; Volta-Congo], for instance, some songs systematically alternate lines sung by a solo and by a group: "a leader 'calls' a line at a time, and the rest of the chorus responds by repeating it" (Agawu 1995:14). This kind of verse structuring with a systematic contrast between solo and group (also known as *call-and-response*) has worldwide presence (Lomax 1976), and is noted as being particularly pervasive across Africa (Stone 2010:307).

Nevertheless, it is not always the case that contrastive sections are also linked by an identity relation like in the Northern Ewe example. In several types of Iraqw [Afro-Asiatic; Cushitic] singing, such as the song *bumbunáy*, solo lines

Example 2.4: First stanza of a song performed by Pacyaya (Maceda 1958:46). The right edge of each line is marked by rhyme.

Northern Kankanay text	**English translation**
ay ínnáo ínnáo ínnáo	"Oh! goodness gracious me!"
kankanáʔ en nan dóngʔ ao	Says the little frog,
makétse nan ipogáo	"Cruel are the people;
uméyak isnan máttáo	I go to the grassy place
ay dadaʔ ét manóngyao	And they burn me out.
ay ínnáo ínnáo ínnáo	Oh! goodness gracious me!"

also alternate with a group response, but the response consists of a fixed text while the words produced by the solo singer vary (own fieldwork).

The structure of the *bumbunáy* into pairs of lines, hence, relies on two contrasting principles: the solo vs group contrast, and the fixed vs changing text contrast. This second kind of contrastive cue is common in many traditions where the role of the group is to produce a refrain. In some cases, the refrain provides the solo singer with some time to think about the following improvised words, e.g. in Aguaruna [Jivaroan] drinking songs (Overall 2007:15). In other cases, the solo text may be pre-composed, so that the refrain works as an organizational device, or, perhaps, provides time to *recall* the following words from memory. The Tashlhiyt *Ndalb irbbi* song, for instance, alternates long sections of solo singing with a fixed group refrain (Dell & Elmedlaoui 2008:7).

Given the impossibility of having a whole group sing a just-improvised text in synchrony, it makes sense that groups produce the fixed section of a song. Nevertheless, an individual singer can also use the fixed-changing contrast without the need of a chorus. In the Melpa *Amb kenan* song described by Niles (2011), each line can be divided into three sections where the central one introduces new words while the other two are kept fixed throughout. Similarly, in the Sardinian *Anninnia di Bosa* lullaby, stanzas are composed of three lines, the last of which shows an invariant text.

A third contrastive principle often correlates with the preceding ones: the contrast between semantically meaningful and meaningless material. In songs such as the Tashlhiyt *Ndalb irbbi*, the refrain does contain a meaningful text; in other cases, such as in Aguaruna drinking songs or the Melpa *Amb kenan*, however, the sections with invariable text are made up of vocables (i.e. words with no semantic content). Suyá *akia* shout songs also provide a clear illustration, where each line of meaningful lyrics is followed by a sequence containing exclusively repetitions of the syllable *te* (Seeger 2004:41).

Example 2.5: First two stanzas of a lyrical song performed by Pajai (Kara 1970:197). The left edge of each line is marked by alliteration.

Mongolian text	**English translation**
čigig ilči-ni tengčeged	if there is enough humidity and warmth,
čečeg quwar-ni delgerejü bayin-a	flowers blossom.
čing sedkilten nököd qorsiyad	if loyal partners get together,
čenggel jirgal orgiju bayin-a	joy and happiness flourish.
usu naran-ni tengčeged	if there is enough water and sunlight,
urgumal bükü soyogalaju bayin-a	every plant grows.
uqagan sanay-a saragulsigad	if the intelligence and determination become clear,
urugsi dabsin gilayiju bayin-a	they shine as they move forward.

2.2.3 Boundary marking

The best known example of a boundary cue is probably rhyme. Consider the Northern Kankanay song under Example 2.4 about a little frog (Maceda 1958:46). This is the first of 22 stanzas, all following the same template. Each of the six lines in the stanza ends with the same sequence of phonemes: *ao*. The division of stanzas into lines is noticeable through several pieces of converging evidence, e.g. all lines are rhythmically parallel since they contain seven syllables and four musical beats. On top of that, the right edge of every line is marked by a rhyming sequence. It is important to note that this kind of boundary marking only works when repeated, so as to create an identity relation between lines. Rhyme, hence, belongs to the boundary marking category as well as to the parallelism category.

The mirror image of rhyme is alliteration, i.e. a sound identity at the left edge of constituents. Alliteration, like rhyme, does not *necessarily* correlate with a line boundary. In Somali [Afro-Asiatic; Cushitic] *geeraar* stanzas, for instance, each line includes a word alliterating in a specified phoneme, but these words do not need to be line-initial (Banti & Giannattasio 1996:99). Nevertheless, some traditions do use alliteration as a line-boundary cue. Example 2.5 reproduces the first two stanzas of a Jarut Mongolian [Mongolic; Eastern Mongolic] lyrical song performed by Pajai (Kara 1970:197).[2] Every line in the first stanza starts with *č-*, while the second stanza marks lines with an alliterating *u-*. Even if rhyme and alliteration correspond to two symmetric cases of boundary-marking parallelism, typological data suggest that rhyme is more frequent than alliteration (Fabb 1999).

Next, consider the case of the Iraqw *slufay* genre, where every line sung by

[2] I have based the English translation on the French version published by Kara (1970).

the solo singer is closed by a collective *haya* 'okay' (Beck & Mous 2014:361). Likewise, in Khmu *hrlɨɨ* solo singing, the end of each strophe is marked by the word *sáh* 'I say' (Lundström 1984:34). This too constitutes a line-boundary cue involving sound parallelism. Unlike rhyme and alliteration, however, the Iraqw and Khmu markers involve an exact repetition of a complete word. If we compare it to rhyming pairs such as *good-neighbourhood* or *high-sky* (from *Kaya* by Bob Marley), we observe that rhyme does not rely on repeating a morpheme, but rather draws an identity relation between phonemes from different morphemes.

In the Iraqw example, the boundary-marking word is kept constant (i.e. parallel) across the whole song; in Kuna singing, on the contrary, a number of different words can play this structural role: "lines are marked grammatically by means of an elaborate set of initial and final words, particles, and affixes" (Sherzer 1982:373). Similarly, in some Ku Waru and Melpa sung tales, lines are closed by a vowel taken from a restricted set: *e*, *o*, or *a* (Rumsey 2010; Niles 2011). In contrast to the Kuna and Iraqw examples, though, these vowels are vocables, i.e. they lack semantic content. This way of defining verse constituents is common also in American traditions, such as Coast Tsimshian (Mulder 1994), or Kwakiutl [Wakashan; Northern Wakashan]: "what have often been called *nonsense syllables* are shown in Kwakiutl to be structural abstracts, covarying with the form of stanzas" (Hymes 1981:10).

Finally, probably the most extensively used boundary-marking cue is the pause; that is, driving adjacent sections apart by introducing some amount of emptiness in between. This is realised in a variety of ways, such as lengthening the final syllable, inserting a silent gap, breathing, or leaving an empty musical beat. This cue can be readily related to the Gestalt principle of proximity, as argued in several cognitive and computational approaches to music segmentation (e.g. Cambouropoulos 1997 and other papers in that volume). Besides, it also bears a close resemblance to final lengthening and pausing as constituency cues in everyday speech (Beckman & Edwards 1990; Fletcher 2010).

In most of the examples cited where a musical transcription is available, we can observe that the notes with the longest duration are systematically placed at the end of constituents, e.g. the Sardinian *Anninnia* lullaby (Sassu & Sole 1972), where line-final notes are at least twice as long as the longest non-final note. In other cases, the authors explicitly mention that singers pause at the end of constituents, e.g. Takū [Austronesian; Nuclear Austronesian] (Moyle 2007:197), Khmu (Lundström 1984:34), Tedaga [Saharan; Western Saharan] (Brandily 1976:171), or Kuna: "there is a notable pause between lines" (Sherzer 1982:377).

Besides variable or non-categorical pauses (often related to breathing), many songs show a categorical pause in the form of an empty musical position. Consider the English [Indo-European; Germanic] nursery rhyme *Eeny meeny miny moe* (Example 3.1 on page 55), where each beat contains two syllables except for the last one in each line, with a single syllable. The gap becomes even longer in cases such as *Hickory dickory dock*, where all lines except the third have a completely empty final beat. Hayes & MacEachern (1998) discusses further English data, and Burling (1966) shows that this kind of gaps at the edges of lines is a common feature of nursery rhymes in several unrelated languages. In this context, whether the last syllable in a line is sustained (i.e. as a long note) or not does not change the fact that the total time span between the last syllabic onset of a line and the first onset of the following line is larger than the neighbouring ones (Temperley 2001:68).

2.3 Numeric control in verse templates

A Shakespearean sonnet contains fourteen lines, and a typical blues verse twelve bars. The poet or musician needs to keep track of those quantities somehow (not necessarily consciously), or else the resulting composition may not adhere to the conventional form. This kind of numeric control can be inferred from regularities in verse productions: after reading, say, fifty fourteen-line sonnets by Shakespeare, we can expect the fifty-first to contain fourteen lines too.

In some of the reviewed traditions, certain styles of singing consist in improvising new verse by following a specified template (e.g. Basque, Tashlhiyt, Northern Kankanay). In these cases, we assume that the performer must take care of the regularities imposed by the template, including numeric ones. In many other styles, though, the texts are fixed and performers do not need to keep track of these numeric regularities (e.g. Pintupi, Coast Tsimshian). In any case, when discussing numeric control, we assume that these fixed texts had to be created at some point, and that the templatic regularities we observe had to be taken care of at the time of composition.

In this section I provide an overview of the requirements of numeric control imposed by different verse templates. In the case of the sonnet, for instance, one has to be aware that the fourteenth and final line has been reached. Nevertheless, Shakespearean sonnets show further structure beyond a concatenation of lines. Rhyming evidence, for instance, shows that the last two lines form an independent section, while the preceding twelve lines are grouped into three quatrains. Hence, the numeric control can be carried out at more than one level: one can

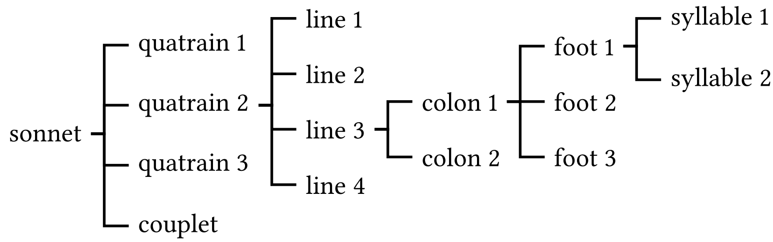


Figure 2.2: Schematic constituent structure of a Shakespearean sonnet. All lines have the same internal structure; the couplet has two instead of four lines.

track a series of three quatrains followed by a couplet, so that reaching the final line does not require keeping track of fourteen constituents, but concludes by reaching the second line of the fourth stanza (i.e. the couplet).

In order to evaluate the numeric requirements imposed by a template, I examine the number of children constituents contained within a given verse constituent, from the bigger ones, such as complete song series or stanzas, to the smallest, such as syllabic feet. In this way, I describe the *minimal* numeric control required to realise a template. In the case of the Shakespearean sonnet, the constituent structure of the template may be characterised using the diagram in Figure 2.2 (Kiparsky 1977; Hanson 2006).

Focusing on the lowest-level constituents (i.e. the syllables) and without considering the hierarchical structure, we would derive the *maximal* numeric control: that is, the number of syllables needs to be tracked up to 140 (i.e. 14 lines $\times$ 10 syllables each).[3] The opposite analysis makes full use of the chunking we have evidence for and provides the *minimal* numeric control, which would be 2, 3, or 4 depending on the constituent level. For instance, in order to create a well-formed quatrain, a quantity of four lines needs to be tracked, while a foot only requires a quantity of two syllables.

In Section 2.2, I have reviewed a number of segmentation cues in verse. Arguably, these constitute structural evidence for putative mental chunks used during the creation of verse. Poets and singers may group constituents based on such cues, and hence alleviate the numeric control imposed by verse templates, e.g. bringing it down from 140 to 4 in the case of the sonnet. Provided that a verse constituent contains a regular number of children constituents, the are at least two possible proxies for the numeric requirements imposed by the template.

[3] Lines can also contain 11 syllables if ending in a feminine rhyme; I overlook this variation for argumentation purposes.

First, the total ***n**umber of **c**hildren **c**onstituents* (*ncc* score) amounts to the highest possible numeric requirement within a constituent. Second, considering that adjacent constituents can be both set apart via contrasting cues or connected through parallelism, a reduced numeric requirement is embodied by the ***m**aximum number of **i**dentical **a**djacent **c**onstituents* (*miac* score). If we describe the stanza structure of a sonnet as *aaab* (i.e. three quatrains and a couplet), the ncc score equals four, and the miac score is given by the series of three quatrains.

I propose that, in general, the miac better reflects the numeric requirements imposed by a template because it reduces the tracking of numerosity to a minimum. In the presence of explicit chunking or contrastive adjacent constituents, one can make use of alternative strategies to adhere to the template, such as relying on transition-based information. As a simple example, consider a call-and-response structure (such as those presented in Section 2.2.2) where each stanza is composed of two lines (*ab*), one sung by the solo singer (*a*), the other by the group (*b*). One can rely on the fact that *a* is always followed by *b*, without needing to resort to counting. In the cases where adjacent constituents are defined by exactly the same set of features (e.g. number of syllables, melody, rhyme), however, numerosity has to be controlled directly.

2.3.1 Above the line

Very often lines are grouped into higher order constituents such as couplets or stanzas. In the Havasupai *Origin song*, for instance, a melody composed of three lines is repeated 39 times (Hinton 1990). In most strophic songs like this one, there is no evidence that the actual number of stanzas is specified in the template. A corpus of songs all containing 39 stanzas would provide evidence for the non-arbitrariness of the number. Hence, we can state that the miac score in these cases is unbounded, and thus uncontrolled. This also applies to the various traditions based on parallel couplets, such as Kuna or Termanu.

Next, there are templates where supra-line constituents must come in pairs such as *ab*, where both *a* and *b* contain several lines. In these cases, the numerosity of supra-line constituents is constrained, but minimally so. For instance, the Melpa *kang rom* sung tale is based on the repetition of a melody containing eight lines, clearly grouped into two quatrains with partial melodic parallelism (Niles 2011). In this case, the number of children constituents equals two, but the miac score is reduced to one, because the melody of the first and second quatrain differs.

The numeric requirement increases in templates with a similar binary supra-line division, but with an *aa* type of structure, where the two sections have identical feature specifications. This is common in sung stanzas with an *ab–ab* couplet structure, such as the Quechua coplas from Cochabamba [Quechuan; Quechua II] (Solomon 1994:384). In a spoken verse context, the lü-shi regulated verse in Literary Chinese [Sino-Tibetan; Sinitic] uses a pentasyllabic eight-line template divided into two equal quatrains of the form *abcd–abcd* (Chen 1979:373). In this case, the numeric requirements are summarised as: ncc = 2, miac = 2.

The Shakespearean sonnet (Figure 2.2) imposes yet greater demands, with ncc = 4, miac = 3. There is a total of four stanzas, with a sequence of three quatrains of identical formal specification. Verse templates realised primarily in written form, such as the sonnet, may increase supra-line numeric requirements by relying on visual support.

2.3.2 Line level

Some traditions compose verse as a sequence of lines which are not further grouped in a systematic way, in a comparable way to the strophe or couplet sequences of arbitrary size mentioned for Havasupai or Kuna. Blacking (1967:91) describes the *Ŋwana wa Vho-Mavhungu* Venda song for girls, which contains eleven lines. An Iraqw *slufay* performance by Haawú Tarmo contains more than two hundred lines (Beck & Mous 2014). In both cases, the exact number of lines does not appear to be regulated, so we can conclude that their miac is unbounded (∞).

The Iraqw and Venda examples are cases of chanted or recited verse, where there is not a clear melodic structure with discrete notes. In verse which is *sung* (to a tune), though, it is common to use templates where different lines have different melodic specifications. The stanzas of the Havasupai *Origin song* show an *aba* tune, with three lines per stanza (ncc = 3), but no adjacent lines with the same melody within the stanza (miac = 1). Each of the sub-stanzas of the Melpa *kang rom* contains four lines (ncc = 4), all with a uniform rhythmical and vocable structure (which would indicate miac = 4), but each with a different melody (miac = 1).

Many traditions are based on couplets, i.e. two-line templates. These can create two scenarios: (1) the two lines have the same formal properties (miac = 2), and (2) the lines show some contrast (miac = 1). Aguaruna drinking song couplets offer a simple illustration of the second scenario: the first line is improvised by a soloist, while the second line is a fixed refrain sung by the group. The Tashlhiyt *Ndalb irbbi* solo stanzas show a similar case where lines follow an *ab* pattern, and thus there is no numeric requirement. Here, both lines are sung by the soloist, but they contrast in the metrical pattern, and the words in the second line are fixed.

Traditions based on semantically parallel couplets may offer an example of the *aa* pattern, where the miac equals two. This can be seen in the systematic binarism of the Termanu sample shown in Example 2.3. Nonetheless, it is unclear the extent to which the two sections of a parallel couplet conform to the criterion of identical adjacent constituents, because the second section may be partially determined by the first one, hence creating an asymmetry. This is manifest in the neighbouring Uab Meto tradition: the semantically parallel terms follow a conventional order, and they often contrast in their phonological shape, by presenting parallel verbs in the unmetathesised form, followed by the metathesised form (Edwards 2016:332). Hence, couplets with completely identical lines may be less frequent than initially apparent.

Greater numerical requirements are easy to find in written traditions, for which certain features related to pitch or intonation are usually absent. The prototypical stanza of Classical Tibetan is composed of four heptasyllabic lines, formally identical and without evidence for subgrouping (Poucha 1950; Vekerdi 1952). In this kind of quatrain, both the ncc and the miac scores equal four. The lack of recorded performance features, however, may conceal other specifications which make lines dissimilar. Consider the Basque *zortziko txikia* template, consisting of four monorhyme formally identical lines, i.e. *aaaa* (Garzia, Sarasua & Egaña 2001). Given that it is a living tradition for which performance data is available, we can observe that the template is realised using a tune such as *Haizeak bidali du*, with an *aabc* line structure (Dorronsoro 1995:339). The miac score is reduced from four to two.

This melody (Example 2.6, Supplementary Information) raises a common issue with published descriptions of verse templates using synthetic codes such as *aabc*. The letters describe identity and contrast relations between constituents, but the same code is also used in cases of similarity or identity *with a difference*. In the present example, the first two lines share their beginnings, but end differently. The beginning of the *a* lines is not shared by the following lines, so a parallelism is drawn between the first two lines, setting them apart. Hence, it may be more appropriate to analyse such cases as having a supra-line constituent: i.e. a couplet which includes the first two lines following an *ab* pattern. Hence, the incorporation of melodic structure to a stanza very easily reduces the numeric requirements of the template (as reflected by the reduction of the miac score from four to one).

The examples of longer forms, such as the sonnet with fourteen lines, usually group lines into sections, and/or differentiate adjacent lines somehow, resulting in low miac scores overall. The Russian Onegin stanza contains fourteen lines, and, although it lacks a unique canonical grouping of the lines, contrary to what we find in Shakespearean sonnets, its rhyme scheme allows at most two identical adjacent lines (i.e. miac = 2) (Scherr 2006).

2.3.3 Below the line

There are a several sub-line constituents which are prone to numeric control: e.g. hemistichs, feet, beats, syllables. In some traditions, lines do not seem to exhibit any numeric control below the line, such as the verse based on semantically parallel couplets. Syllable-count tends to be the primary regularity looked for; however, other regularities are sometimes obscured because only the linguistic structure of the lines, but not the musical one, is described or taken into account (e.g. compare the analysis of Mongolian parallel verse by Poppe 1958 vs Kara 1970).

The simplest example of sub-line structure is a binary division into half-lines of the type *ab* or *aa*. All four lines of the Basque template *Haizeak bidali du* contain two half-lines (aka *hemistichs*) of seven and six syllables respectively (Dorronsoro 1995:339). The French classical decasyllable contains two hemistichs of four and six syllables each (Gouvard 1999). In these two cases, the sub-line sections share formal similarities with the lines, e.g. in requiring their boundaries to coincide with word boundaries. In this kind of asymmetric binary division, the miac equals 1, with no real numeric demand.

The lines of the *Ŋwana wa Vho-Mavhungu* Venda song, on the other hand, are divided into two symmetric sections of three syllables each (Blacking 1967:91). The subdivision of the line is made explicit by a handclap aligned with the first of the three syllables. Formally similar to the decasyllabic line, but with a symmetric division, the French classical alexandrine line contains two hemistichs of six syllables each (Dominicy 1992:161). Regarding their length in number of syllabic positions, the last two examples display both an ncc and a miac score of two. Nevertheless, adjacent constituents which seem identical may show finer-grain differences, and have effectively asymmetric representations for the poet. For instance, the second section of the Venda lines are closed by a fixed bisyllabic word (*khithi*), whereas the first section shows no such restriction. Regarding the classical alexandrine hemistichs, an additional schwa-syllable is used optionally at the end of the second hemistich, though not at the end of the first (Dominicy 1992:162).

Lines with a greater number of subdivisions occur in both sung and spoken verse. Consider the popular Pashto [Indo-European; Indo-Iranian] *misrəy* template; the first line contains nine syllables, divided 1+4+4, and the second line thirteen, divided 1+4+4+4 (MacKenzie 1958:323). The Ancient Greek [Indo-European; Graeco-Phrygian] trimeter line also shows a miac of 3, with just one syllable less: 4+4+4 (Prince 1989:61). This ternary structure is similar to the long colon of an English iambic pentameter line, which contains three binary feet (Figure 2.2). In the Northern Kankanay song discussed in Example 2.4, lines contain seven syllables, divided into four musical beats of equal length (Maceda 1958:52). The first three constituents contain two syllables, but the fourth one contains a single syllable followed by a gap (cf. Section 2.2.3), yielding an *aaab* pattern, with ncc = 4 but keeping the miac = 3. Unlike the Pashto and Greek sub-line sections, the Northern Kankanay ones are also described with distinctive melodic structures. Thus, it is more accurate to represent its line substructure as *abcd*, and miac = 1.

Pintupi [Pama-Nyungan; Desert Nyungic] *tingarri* lines are also subdivided into four sections, where rhythmic parallelism is frequent (Moyle 1979:83). Among the different patterns employed, an *aaaa* structure where all subsections follow the same rhythmical configuration shows the highest frequency. *Tingarri* lines are sung, and the four sub-line constituents show different melodies, even if they are rhythmically parallel. However, there is stronger evidence for an *aaaa* pattern instead of an *abcd* one, unlike in the Norhtern Kankanay case. The words in *tingarri* lines (and lines in Pintupi verse more generally) exhibit a fixed rhythmic configuration, but each unique line of text gets aligned to a variety of melodic contours during the performance of a song. Hence, the rhythmic features of a line are probably part of the underlying template potentially used to generate new song words, which are then aligned to a melodic contour (for further details on the independence of the melody in other Central Australian traditions, see Turpin 2007). In the Northern Kankanay example, on the contrary, there is no evidence to disconnect the rhythmic and melodic features.

The sub-line constituents reviewed thus far are all subdivided into further constituents, primarily syllables. In the English and Northern Kankanay examples, sub-line constituents are divided into groups of two syllables (i.e. *feet*), comparable to spoken verse forms such as the Estonian [Uralic; Finnic] trochaic tetrameter (Lotman & Lotman 2011). This type of supra-syllabic constituents all seem to be asymmetric, with one of the syllables being somehow more prominent. In the English case, the second syllable is the prominent one (i.e. it usually allocates a stressed syllable), and the opposite is true for the Estonian tetrameter, as well as for the Northern Kankanay example (i.e. the musical beat falls on the first syllable). In both cases, the smallest sub-line constituents show an *ab* asymmetric pattern, with a miac score of one.

Larger groups of syllables show similar asymmetries. Each of the half-lines in the Venda example is composed of three syllables, with the first being prominent by being aligned with a hand-clap, suggesting an *abb* pattern. The Pashto four-syllabic sections show a 1020 prominence contour (0 indicating unstressed syllables, and 2 and 1 primary and secondary stress respectively).

In languages where the primary metrical strategy is said to be syllable counting (Fabb 2015:85–91), we find yet longer sub-line sections containing a systematic number of syllables. The first section of Dyirbal [Pama-Nyungan] *Gama* lines contain five syllables (Dixon & Koch 1996:52); each hemistich of a French alexandrine contains six syllables (Dominicy 1992:161); each line of a Classical Tibetan stanza contains seven syllables (Poucha 1950:196). Common to these three languages is the fact that they exhibit a non-contrastive, fixed word-stress. However, there is evidence that the syllables-holders within these constituents are not completely identical, thus ruling out potential miac scores of five, six, or seven.

Dyirbal has fixed initial stress, and 85% of the *Gama* lines begin with a bisyllabic word followed by a trisyllabic word, which can be analysed as an *ab* asymmetric binary division, or at least in terms of a 10100 stress pattern. This sub-line section is always followed by a bisyllabic word, then by a tetrasyllabic word. Tibetan also shows fixed initial stress (Beyer 1992:408), and its heptasyllabic constituents exhibit a subdivision similar to the Dyirbal one. Most lines are composed of three words (2+2+3), with a possible four-word variation (2+2+2+1; Vekerdi 1952:223).

In these two examples, the statistical regularities of the word make-up of the line provides evidence against a long sequence of syllable-holders with identical cognitive representation. Referring to French verse, Biggs (1996:165) argues that “the traditional account of the line as purely syllabic is incomplete”, further proposing that the first hemistich of the alexandrine line is composed of three iambic feet, and the second hemistich by two anapestic feet (Biggs 1996:178). The robustness of these results may be up to discussion; nonetheless, mainstream, more conservative accounts of the data accept that, at the very least, the sixth syllable of each hemistich must have a different representation than the preceding five, because it categorically receives a stressed syllable (Dominicy 1992:164).

2.4 Discussion

Verse constituents can be set apart and grouped together on the basis of shared features, contrastive features, and boundary markers. By grouping together adjacent constituents, it is possible to posit intermediate constituency levels, such as the couplet level within an *aabb* rhyming quatrain. Most templates here reviewed show a systematic number of children constituents at some level, such as the number of lines within a stanza, or the number of syllables within a half-line. I have argued that direct control of these numeric regularities (e.g. by counting) is not needed when adjacent constituents have different templatic specifications, because the composer of verse can use these features as indication of which constituent comes next. In cases where adjacent constituents are underlyingly identical, though, the composer is obliged to keep numeric track of the constituents in order to produce the expected amount. I have proposed an index (the ***m**aximum number of **i**dentical **a**djacent **c**onstituents*, or *miac*) to capture the numerical requirements imposed by a given verse constituent.

Many constituents do not show any numeric requirement whatsoever, i.e. miac equals one, or is unbounded (∞). The prototypical case of a miac = 1 are melodic constituents, where the number of syllable holders is fixed, but they differ in pitch, duration or prominence, making it very unlikely that two adjacent holders share exactly the same feature specifications. The prototypical case of a miac = ∞ are strophic songs, where adjacent stanzas are formally identical but they are not required to come in specific numbers. Beyond these two cases, the analysed sample of templates shows robust evidence for miac scores of two, three and four, but not more. This limit lies within the subitizing range (see below), which places the numeric requirements imposed by verse templates in the broader context of human cognition.

Since Kaufman et al. (1949) coined the term, a wealth of psychological studies have discussed the difference between *subitizing*, and counting or estimating (see Mandler & Shebo 1982 and Núñez 2017 for overviews). When presented with a small set of objects, such as three stones, one is immediately aware of the exact quantity; the same accuracy is also attainable for a collection of, say, thirty-seven stones, but it will require a much slower process. In the first scenario, we are assumed to complete the task instantly by subitizing, while in the second we resort to counting. Alternatively, we often solve the second kind of scenario by estimating, which is quicker than counting, but less accurate.

Apart from the speed and accuracy advantage related to the subitizing range, its special status is also supported by developmental and linguistic data. Infants are able to discriminate (both visually and acoustically) between discrete quantities as long as these do not exceed a limit of four (Antell & Keating 1983; Bijelac, Bertoncini & Mehler 1993; Van Loosbroek & Smitsman 1990). This is remarkable, since the ability is reported for infants who have not yet acquired number-related terms or a linguistic system. With larger quantities, six-month-old infants can discriminate sets differing by a ratio such as 8 vs 16, but not by smaller ratios such as 8 vs 12 (Xu & Spelke 2000). These results have led to the proposal of two distinct core knowledge systems: an object tracking system specialised in the subitizing range, and an approximate number system used for higher numerosities (Carey & Xu 2001; Feigenson, Dehaene & Spelke 2004). These systems are considered to belong to a small set of innate, non-species-specific cognitive abilities which both enable and constrain more complex skills and cultural systems such as language or arithmetic (Spelke & Kinzler 2007).

Two kind of linguistic data support the special status of the subitizing range. First, many languages (ca. 1386, Hammarström 2013) have a restricted numeral system, with words for small, exact quantities (e.g. between one and three), and approximate expressions such *several* or *many* for larger quantities (Pica et al. 2004; Epps et al. 2012). Second, languages with more extensive numeral systems often treat numeric expressions within the subitizing range in a grammatically distinct way, e.g. by using gender or case agreement, which is then neutralised for higher numerals (Greenberg 2005:42; Hurford 1987).

Given the body of behavioural, developmental and linguistic evidence, researchers have proposed that subitizing and counting involve separate cognitive processes, probably recruiting distinct neural resources (Burr, Turi & Anobile 2010; Vuokko, Niemivirta & Helenius 2013). Others, however, argue that increased effort, reaction times and neural activity are associated with higher numerosities both within and beyond the subitizing range, suggesting a gradual (rather than categorical) transition between the two tasks (Balakrishnan & Ashby 1992; Piazza et al. 2002).

An important difference between most studies on numeric cognition and the domain of verse is that the former focus on visual, static patterns, while verse involves temporal patterns which are not perceivable simultaneously. However, even if marginally, developmental studies also argue that when we perceive simple double and triple musical rhythms we are engaging in a task comparable to visual subitizing (Von Glasersfeld 1982; Steffe & Cobb 1988). Indeed, phenomenological discussions on numerical control during verse creation reject explicit counting of e.g. syllables, and describe the process in terms of *feeling* in an automatic way the right rhythm to be filled by the text (Gentili 1955 for Greek poetry, Banti & Giannattasio 1996 for Somali sung verse).

The subitizing range, hence, may prove appropriate for lower level constituents such as feet and other sub-line constituents. The fact that larger constituents, such as the number of couplets within a stanza, show similar numeric demands is unlikely to be explained by subitizing constraints; to be sure, automaticity represents an essential trait of subitizing, and it is likely to have time constraints in addition to quantity constraints.

Even beyond constituent levels prone to subitizing, a pervasive feature of verse constituents is that they get chunked. Chunking is a general cognitive strategy used spontaneously and in a wide range of contexts to deal with large quantities efficiently (Gobet et al. 2001). Similar to subitizing, it appears early in ontogeny (Rosenberg & Feigenson 2013), which argues for its basic status in cognition. Crucial to working memory, hierarchical chunking can expand its limits to allocate tens of items at a time (Ericsson, Chase & Faloon 1980).

Fabb (2014:29) argues that verse lines are "held as a whole sequence in the limited capacity of working memory." Nonetheless, other metricists have also expressed the intuition that when lines exceed a certain length limit, they undergo chunking. Regarding French verse, de Cornulier (1995:47) holds that "beyond eight, exact numbers of syllables become inaccessible to perception"[4], longer lines being divided into hemistichs as a consequence. Beltrami (1984) proposes a similar but more restricted 'law of six syllables' for Italian, and Dominicy (1992:161) hypothesises that comparable limits may hold for "many other languages."

These quantities fall within the classical, so-called magical, number of 7±2 for the number of units which can be held in working memory (Miller 1956), a remark which has been made for early Romance poetic genres too (Valenti 2009). Nevertheless, in the sample of templates, even sub-line hemistichs showed more restricted ncc and miac scores, not exceeding the subitizing limit of four. Updated versions of the magical number 7±2 seem to bridge the gap between the two limits. Miller's number is now understood as the number of *uncompressed* objects which can be held in working memory; given that we readily chunk inputs, the actual limit of *compressed* (i.e. chunked) items is set at 4±1 (Cowan 2001; Mathy & Feldman 2012). This number corresponds to the limit observed in this chapter's data; hence we conclude that numeric control demanded by verse constituents is bounded by both the subitizing range, and the working memory limit. That syllables within lines are chunked is evident in templates for which poetic feet are proposed (e.g. Greek, English). Moreover, syllables are grouped into phonological feet and words in virtually every spoken language (Hayes 1995). Hence, lines or hemistichs of up to eight syllables are easily compressable into subitizable chunks.

Finally, it is critical to note that the data here presented serve but to illustrate some problems faced by the study of constituent structure, and a case study involving numeric control. In order to develop this preliminary work, a balanced

[4] Original quotation in French: "En français, au-delà de huit, le nombre syllabique exact est inaccessible à la perception" (de Cornulier 1995:47).

sample of languages needs to be thoroughly surveyed, with detailed analyses of templates encoded into a database. This will enable discovering typological and geographical patterns which can further lead to testable hypotheses on how cognition shapes the verse templates at the core of human songs.

2.5 Conclusion

I have presented a simple taxonomy of constituency cues (parallelism, contrast, and boundary markers), and illustrated them in the verse traditions of unrelated linguistic families. Taking into account these cues we can investigate the numeric requirements imposed by a verse template. In the pilot dataset under study, there is no evidence of verse templates where the number of constituents to keep track of exceeds four. These results conform with landmarks related to numeric cognition, such as the subitizing range. Hence, verse typology can both benefit from and inform the subject of how humans deal with temporal sequences.

References

Agawu, K. 1995. *African Rhythm: A Northern Ewe Perspective*. Vol. 1. Cambridge University Press.

Antell, S. E. & D. P. Keating. 1983. Perception of numerical invariance in neonates. *Child development*. 695–701.

Aroui, J.-L. 2009. Proposals for metrical typology. In J.-L. Aroui & A Arleo (eds.), *Towards a typology of poetic forms*, 1–39. Benjamins.

Aroui, J.-L. 2005. Rime et richesse des rimes en versification française classique. In M. Murat & J. Dangel (eds.), *Poétique de la rime* (Métrique française et comparée), 179–218. Honoré Champion.

Balakrishnan, J. D. & F. G. Ashby. 1992. Subitizing: magical numbers or mere superstition? *Psychological Research* 54. 80–90.

Banti, G & F Giannattasio. 1996. Music and metre in Somali poetry. In R.-J. Hayward & I.-M. Lewis (eds.), *Voice and power: essays in honour of B. W. Andrzejewski*, 83–127. School of Oriental & African Studies.

Beck, R. & M Mous. 2014. Iraqw *slufay* and the power of voice. In H Vögele, U Reuster-Jahn, R Kastenholz & L Diegner (eds.), *From the tana river to lake chad. research in african oratures and literatures. in memoriam thomas geider*, 357–371. Köln: Rüdiger Köppe Verlag.

Beckman, M. E. & J. Edwards. 1990. Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston & M. E. Beckman (eds.), *Papers in laboratory phonology 1: between the grammar and physics of speech*, 152–178. Cambridge University Press.

Beltrami, P. G. 1984. Review: de Cornulier. 1982. Théorie du vers. Rimbaud, Verlaine, Mallarmé. *Rivista di Letteratura Italiana* 2. 587–605.

Beyer, S. V. 1992. *The Classical Tibetan Language*. State University of New York Press.

Bickel, B. 2007. Typology in the 21st century: major current developments. *Linguistic Typology* 11(1). 239–251.

Bickel, B. & J. Nichols. 2002. Autotypologizing databases and their use in fieldwork. In P. Austin, H. Dry & P. Wittenburg (eds.), *Proceedings of the international LREC workshop on resources and tools in field linguistics*. Nijmegen: Max Planck Institute for Psycholinguistics.

Biggs, H. P. 1996. *A statistical analysis of the metrics of the Classic French Decasyllable and the Classic French Alexandrine*. University of California, Los Angeles Doctoral dissertation.

Bijelac, R, J Bertoncini & J Mehler. 1993. How do four-day-old infants categorize multisyllabic utterances. *Developmental Psychology* 29. 253–269.

Blacking, J. 1967. *Venda children's songs: a study in ethnomusicological analysis*. University of Chicago Press.

Blacking, J. 1969. Songs, dances, mimes and symbolism of Venda girls' initiation schools: Part 4: The great Domba song. *African Studies* 28(4). 215–266.

Brandily, M. 1976. Un chant du Tibesti (Tchad). *Journal des africanistes* 46(1). 127–192.

Brown, A.-R., T Gifford & R Davidson. 2012. Tracking levels of closure in melodies. In E Cambouropoulos, C Tsougras, P Mavromatis & K Pastiadis (eds.), *Proceedings of the 12th international conference on music perception and cognition*, 149–152. School of Music Studies, Aristotle University of Thessaloniki.

Brown, S. & J. Jordania. 2011. Universals in the world's musics. *Psychology of Music* 41(2). 229–248.

Brown, S., P. E. Savage, A. M.-S. Ko, M. Stoneking, Y.-C. Ko, J.-H. Loo & J. A. Trejaut. 2014. Correlations in the population structure of music, genes and language. *Proc. R. Soc. B* 281(1774). 1–7.

Burgoyne, J. A., I. Fujinaga & J. S. Downie. 2015. Music information retrieval. In S. Schreibman, R. Siemens & J. Unsworth (eds.), *A new companion to digital humanities*, 213–228. John Wiley & Sons, Ltd.

Burling, R. 1966. The metrics of children's verse: a cross-linguistic study. *American Anthropologist* 68(6). 1418–1441.

Burr, D. C., M. Turi & G. Anobile. 2010. Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision* 10(6). 20.

Cambouropoulos, E. 1997. Musical rhythm: a formal model for determining local boundaries, accents and metre in a melodic surface. In M. Leman (ed.), *Music, gestalt, and computing: studies in cognitive and systematic musicoiogy*, 277–293. Springer.

Carey, S & F Xu. 2001. Infants' knowledge of objects: beyond object files and object tracking. *Cognition* 80. 179–213.

Chen, M. Y. 1979. Metrical structure: evidence from Chinese poetry. *Linguistic Inquiry* 10(3). 371–420.

Cowan, N. 2001. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and brain sciences* 24(01). 87–114.

de Cornulier, B. 1995. *Art poëtique. Notions et problèmes de métrique*. Lyon (France): Presses Universitaires de Lyon.

Dell, F & M Elmedlaoui. 2008. *Poetic meter and musical form in Tashlhiyt Berber songs*. Köln: Rüdiger Köppe.

Deutsch, D. 2013a. Grouping mechanisms in music. In D. Deutsch (ed.), *The psychology of music*, 183–248. Elsevier.

Deutsch, D. 2013b. The processing of pitch combinations. In D. Deutsch (ed.), *The psychology of music*, 249–325. Elsevier.

Dixon, R.-M.-W. & G Koch. 1996. *Dyirbal song poetry. The oral literature of an Australian rainforest people*. University of Queensland Press.

Dominicy, M. 1992. On the meter and prosody of French 12-syllable verse. *Empirical Studies of the Arts* 10(2). 157–181.

Dorronsoro, J. 1995. *Bertso doinutegia*. Donostia: Euskal Herriko Bertsolari Elkartea.

Edwards, O. 2016. *Metathesis and Unmetathesis: Parallelism and Complementarity in Amarasi, Timor*. The Australian National University Doctoral dissertation.

Egaña, A. 2007. The process of creating improvised bertsos. *Oral Tradition* 22(2). 117–142.

Epps, P., C. Bowern, C. A. Hansen, J. H. Hill & J. Zentz. 2012. On numeral complexity in hunter-gatherer languages. *Linguistic Typology* 16(1). 41–109.

Ericsson, K., W. G. Chase & S. Faloon. 1980. Acquisition of a memory skill. *Science* 208(4448). 1181–1182.

Fabb, N. 1997. *Linguistics and literature*. Blackwell Publishers.

Fabb, N. 1999. Verse constituency and the locality of alliteration. *Lingua* 108. 223–245.

Fabb, N. 2014. The verse-line as a whole unit in working memory, ease of processing, and the aesthetic effects of form. *Royal Institute of Philosophy Supplement* 75. 29–50.

Fabb, N. 2015. *What is poetry? Language and memory in the poems of the world*. Cambridge University Press.

Feigenson, L., S. Dehaene & E. Spelke. 2004. Core systems of number. *Trends in cognitive sciences* 8(7). 307–314.

Fletcher, J. 2010. The prosody of speech: timing and rhythm. In W.-J. Hardcastle, J Laver & F.-E. Gibbon (eds.), *The handbook of phonetic sciences*, 523–602. Blackwell Publishers.

Fox, J. J. 2014a. Roman Jakobson and the comparative study of parallelism. In J. J. Fox (ed.), *Explorations in semantic parallelism*, 19–40. First published in 1977. ANU Press.

Fox, J. J. 2014b. Semantic parallelism in Rotenese ritual language. In J. J. Fox (ed.), *Explorations in semantic parallelism*, 91–128. First published in 1971. ANU Press.

Frog. 2014. A preface to parallelism. In Frog (ed.), *Parallelism in verbal art and performance*, 7–28. Folklore Studies, University of Helsinki.

Garzia, J, J Sarasua & A Egaña. 2001. *The art of bertsolaritza: improvised Basque verse singing*. Donostia: Bertsozale Elkartea.

Gasparov, M. L. 1996. *A history of European versification*. G. S. Smith & L. Holford-Strevens (eds.). Oxford University Press.

Gentili, B. 1955. *La metrica dei greci*. Casa Editrice G. D'Anna.

Gobet, F., P. C. Lane, S. Croker, P. C. Cheng, G. Jones, I. Oliver & J. M. Pine. 2001. Chunking mechanisms in human learning. *Trends in cognitive sciences* 5(6). 236–243.

González-Blanco García, E., M. Manailescu & S. Ros. 2016. From syllables, lines and stanzas to linked open data: standardization, interoperability and multilingual challenges for digital humanities. In *Proceedings of the fourth international conference on technological ecosystems for enhancing multiculturality*, 979–983. ACM.

Gouvard, J.-M. 1999. *La versification*. Paris: Presses universitaires de France.

Greenberg, J. H. 2005. *Language universals: with special reference to feature hierarchies*. Walter de Gruyter.

Hammarström, H. 2013. Universals and non-universals in numeral systems in the languages of the world. Unpublished paper presented at IMPRS, Nijmegen (The Netherlands).

Hanson, K. 2006. Shakespeare's lyric and dramatic metrical styles. In B.-E. Dresher & N Friedberg (eds.), *Formal approaches to poetry. recent developments in metrics*, 111–133. Mouton de Gruyter.

Haspelmath, M. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.

Hayes, B.-P. & M MacEachern. 1998. Quatrain form in English folk verse. *Language* 74(3). 473–507.

Hayes, B. 1995. *Metrical stress theory: principles and case studies*. University of Chicago Press.

Hinton, L. 1984. *Havasupai songs: a linguistic perspective*. Vol. 6. John Benjamins Publishing Company.

Hinton, L. 1990. Song metrics. In *Annual meeting of the berkeley linguistics society*, vol. 16, 51–60.

Hull, K. M. 2003. *Verbal art and performance in Ch'orti' and Maya hieroglyphic writing*. University of Texas Doctoral dissertation.

Humphrey, E., J Salamon, O Nieto, J Forsyth, R. Bittner & J. Bello. 2014. JAMS: a JSON annotated music specification for reproducible MIR research. In *Proceedings of the 15th international society for music information retrieval conference (ismir 2014)*, 591–596. ISMIR.

Hurford, J. R. 1987. *Language and number: the emergence of a cognitive system.* Basil Blackwell.

Hymes, D. 1981. "In vain I tried to tell you": essays in Native American ethnopoetics.

Jakobson, R. 1966a. *Selected writings IV. Slavic epic studies.* Mouton de Gruyter.

Jakobson, R. 1966b. Grammatical parallelism and its Russian facet. *Language* 42(2). 399.

Kara, G. 1970. *Chants d'un barde mongol.* Budapest: Akadémiai Kiadó.

Kaufman, E. L., M. W. Lord, T. W. Reese & J. Volkmann. 1949. The discrimination of visual number. *The American journal of psychology* 62(4). 498–525.

Kiparsky, P. 1977. The rhythmic structure of English verse. *Linguistic Inquiry* 8(2). 189–247.

Le Bomin, S., G. Lecointre & E. Heyer. 2016. The evolution of musical diversity: the key role of vertical transmission. *PLoS One* 11(3). 1–17.

Lerdahl, F & R Jackendoff. 1983. *A generative theory of tonal music.* Massachusetts Institute of Technology.

Lomax, A. 1976. *Cantometrics: an approach to the anthropology of music: audiocassettes and a handbook.* Berkeley: University of California Extension Media Center.

Lomax, A. & V. Grauer. 1968. The cantometric coding book. In *Folk song style and culture*, 34–74. American Association for the Advancement of Science.

Lotman, M & M.-K. Lotman. 2011. Toward a statistical analysis of accentual rhythm (with reference to the Estonian trochaic tetrameter). In B.-P. Scherr, J Bailey & E Kazartsev (eds.), *Formal methods of poetics*, 256–294. RAM-Verlag.

Lotz, J. 1960. Metric typology. In T.-A. Sebeok (ed.), *Style in language*, 135–148. Massachusetts Institute of Technology Press.

Lundström, H. 1984. A Kammu song and its structure. *Asian Folklore Studies* 43(1). 29–39.

Maceda, J. 1958. Chants from Sagada Mountain Province, Philippines. *Ethnomusicology* 2(2). 45–55.

MacKenzie, D. N. 1958. Pashto verse. *Bulletin of the School of Oriental and African Studies* 21(02). 319–333.

Mandler, G. & B. J. Shebo. 1982. Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General* 111(1). 1.

Mathy, F. & J. Feldman. 2012. What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition* 122(3). 346–362.

Miller, G. A. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review* 63(2). 81–97.

Moyle, R. M. 1979. *Songs of the Pintupi: musical life in a central Australian society.* Australian institute of aboriginal studies.

Moyle, R. M. 2007. *Songs from the Second Float: A Musical Ethnography of Takū Atoll, Papua New Guinea.* University of Hawaii Press.

Mulder, J. 1994. Structural organization in Coast Tsimshian music. *Ethnomusicology* 38(1). 81–125.

Narmour, E. 1990. *The analysis and cognition of basic melodic structures. The implication-realization model.* Chicago University Press.

Nieto, O. & J. P. Bello. 2016. Systematic exploration of computational music structure research. In *Proceedings of the 17th international society for music information retrieval conference (ismir 2016)*, 547–553.

Niles, D. 2011. Metric melodies and the performance of sung tales in the Hagen area. In A Rumsey & D Niles (eds.), *Sung tales from the papua new guinea highlands. studies in form, meaning, and sociocultural context*, 275–201. The Australian National University.

Nordhoff, S. & H. Hammarström. 2011. Glottolog/Langdoc: Defining dialects, languages, and language families as collections of resources. In *Proceedings of the first international workshop on linked science 2011*, vol. 783 (CEUR Workshop Proceedings).

Núñez, R. E. 2017. Is there really an evolved capacity for number? *Trends in cognitive sciences* 21(6). 409–424.

Overall, S. 2007. *A grammar of Aguaruna.* LaTrobe University Doctoral dissertation. 611.

Patel, A. D. & S. M. Demorest. 2013. Comparative music cognition: cross-species and cross-cultural studies. In D. Deutsch (ed.), *The psychology of music*, 647–681. Elsevier.

Piazza, M., A. Mechelli, B. Butterworth & C. J. Price. 2002. Are subitizing and counting implemented as separate or functionally overlapping processes? *Neuroimage* 15(2). 435–446.

Pica, P., C. Lemer, V. Izard & S. Dehaene. 2004. Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306(5695). 499–503.

Poppe, N. 1958. Der Parallelismus in der epischen Dichtung der Mongolen. *Ural-Altaische Jahrbücher* 30(3-4). 195–228.

Poucha, P. 1950. Le vers tibétain. *Archiv Orientalni* 18(4). 188–235.

Prince, A. 1989. Metrical forms. In P Kiparsky & G Youmans (eds.), *Rhythm and meter*, 45–80. Academic Press.

Rodríguez López, M. 2016. *Automatic melody segmentation.* Utrecht University Doctoral dissertation.

Rosenberg, R. D. & L Feigenson. 2013. Infants hierarchically organize memory representations. *Developmental science* 16(4). 610–621.

Rumsey, A. 2010. A metrical system that defies description by ordinary means. In J. Bowden, N. Himmelmann & M. Ross (eds.), *A journey through Austronesian and Papuan linguistic and cultural space. Papers in honour of Andrew K. Pawley*, 39–56. Canberra: Pacific Linguistics.

Saffran, J. R., E. K. Johnson, R. N. Aslin & E. L. Newport. 1999. Statistical learning of tone sequences by human infants and adults. *Cognition* 70(1). 27–52.

Sassu, P. & L. Sole. 1972. Funzione degli stereotipi nel canto popolare sardo. *Rivista Italiana di Musicologia* 7(1). 115–144.

Savage, P. E., E. Merritt, T. Rzeszutek & S. Brown. 2012. CantoCore: a new cross-cultural song classification scheme. *Analytical Approaches to World Music* 2(1). 87–137.

Savage, P. E., H. Matsumae, H. Oota, M. Stoneking, T. E. Currie, A. Tajima, M. Gillan & S. Brown. 2015a. How 'circumpolar' is Ainu music? Musical and genetic perspectives on the history of the Japanese archipelago. *Ethnomusicology Forum* 24(3). 443–467.

Savage, P. E., S. Brown, E. Sakai & T. E. Currie. 2015b. Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences* 112(29). 8987–8992.

Scherr, B. P. 2006. Structural dynamics in the Onegin stanza. In B.-E. Dresher & N Friedberg (eds.), *Formal approaches to poetry. recent developments in metrics*, 267–284. Mouton de Gruyter.

Seeger, A. 2004. *Why Suyá sing. A musical anthropology of an Amazonian people.* Cambridge University Press.

Sherzer, J. 1982. Poetic structuring of Kuna discourse: the line. *Language in Society* 11. 371–390.

Solomon, T. 1994. Coplas de Todos Santos in Cochabamba: language, music, and performance in Bolivian Quechua song dueling. *The Journal of American Folklore* 107(425). 378–414.

Spelke, E.-S. & K.-D. Kinzler. 2007. Core knowledge. *Developmental Science* 10(1). 89–96.

Steffe, L. & P. Cobb. 1988. *Construction of arithmetical meanings and strategies.* Springer Science & Business Media.

Stone, R. M. (ed.). 2010. *The Garland handbook of African music.* Routledge.

Temperley, D. 2001. *The cognition of basic musical structures.* Cambridge (MA): Massachusetts Institute of Technology.

Turpin, M. 2007. The poetics of Central Australian song. *Australian Aboriginal Studies* 2. 100–115.

Valenti, G. 2009. The magical number seven and the early Romance poetry. *Cognitive Philology* 2.

Van Loosbroek, E. & A. W. Smitsman. 1990. Visual perception of numerosity in infancy. *Developmental Psychology* 26(6). 916–922.

Vekerdi, J. 1952. Some remarks on Tibetan prosody. *Acta Orientalia Academiae Scientiarum Hungaricae* 2. 221–234.

Versace, S. & N. Fabb. 2011. A database as a method of raising typological questions about poetic form. In P. Austin, O. Bond, D. Nathan & L. Marten (eds.), *Proceedings of conference on language documentation & linguistic theory 3*, 289–296.

Von Glasersfeld, E. 1982. Subitizing: the role of figural patterns in the development of numerical concepts. *Archives de Psychologie.*

Vuokko, E., M. Niemivirta & P. Helenius. 2013. Cortical activation patterns during subitizing and counting. *Brain research* 1497. 40–52.

Xu, F. & E. S. Spelke. 2000. Large number discrimination in 6-month-old infants. *Cognition* 74(1). B1–B11.

Yaqub, N. G. 2007. *Pens, swords, and the springs of art: the oral poetry dueling of Palestinian weddings in the Galilee.* Brill.

Zedda, P. 2009. The Southern Sardinian Tradition of the Mutetu Longu: A Functional Analysis. *Oral Tradition* 24(1). 3–40.

Supplementary Information

Example 2.6: The tune *Haizeak bidali du* used in Basque verse to improvise new lines of text (Dorronsoro 1995:339). The beginning of the first two lines show melodic parallelism.

	Language	Family	ISO-639-3	Glottocode	Macroarea
1	Tashlhiyt	Afro-Asiatic; Berber	shi	tach1250	Africa
2	Iraqw	Afro-Asiatic; Cushitic	irk	iraq1241	Africa
3	Somali	Afro-Asiatic; Cushitic	som	soma1255	Africa
4	Ewe	Atlantic-Congo; Volta-Congo	ewe	ewee1241	Africa
5	Venda	Atlantic-Congo; Volta-Congo	ven	vend1245	Africa
6	Tedaga	Saharan; Western Saharan	tuq	teda1241	Africa
7	Dyirbal	Pama-Nyungan	dbl	dyir1250	Australia
8	Pintupi	Pama-Nyungan; Desert Nyungic	piu	pint1250	Australia
9	Arabic	Afro-Asiatic; Semitic	arb	stan1318	Eurasia
10	Khmu	Austroasiatic; Khmuic	kjg	khmu1256	Eurasia
11	Russian	Indo-European; Balto-Slavic	rus	russ1263	Eurasia
12	English	Indo-European; Germanic	eng	stan1293	Eurasia
13	Greek	Indo-European; Graeco-Phrygian	grc	anci1242	Eurasia
14	Pashto	Indo-European; Indo-Iranian	pst	cent1973	Eurasia
15	French	Indo-European; Italic	fra	stan1290	Eurasia
16	Sardinian	Indo-European; Italic	src	logu1236	Eurasia
17	Basque	Isolate	eus	basq1248	Eurasia
18	Mongolian	Mongolic; Eastern Mongolic	khk	halh1238	Eurasia
19	Tibetan	Sino-Tibetan; Bodic	bod	tibe1272	Eurasia
20	Chinese	Sino-Tibetan; Sinitic	lzh	lite1248	Eurasia
21	Estonian	Uralic; Finnic	ekk	esto1258	Eurasia
22	Havasupai	Cochimi-Yuman; Yuman	yuf	hava1248	North America
23	Chorti	Mayan; Core Mayan	caa	chor1273	North America

24	Tsimshian	Tsimshian	tsi	nucl1649	North America
25	Kwakiutl	Wakashan; Northern Wakashan	kwk	kwak1269	North America
26	Kankanay	Austronesian; Nuclear Austronesian	xnn	nort2877	Papunesia
27	Takuu	Austronesian; Nuclear Austronesian	nho	taku1257	Papunesia
28	Termanu	Austronesian; Nuclear Austronesian	twu	term1237	Papunesia
29	Uab Meto	Austronesian; Nuclear Austronesian	aoz	uabm1237	Papunesia
30	Melpa	Nuclear Trans New Guinea; Central East New Guinea Highlands	med	melp1238	Papunesia
31	Kuna	Chibchan; Core Chibchan	cuk	sanb1242	South America
32	Aguaruna	Jivaroan	agr	agua1253	South America
33	Suyá	Nuclear-Macro-Je; Je	suy	suya1243	South America
34	Quechua	Quechuan; Quechua II	qul	nort2976	South America

Table 2.1: Languages whose verse system is mentioned in the chapter.

3 The emergence of verse templates through iterated learning

3.1 Introduction

Language is present in every human society, in a variety of forms and contexts. Beyond every-day speech, language is also widely found in verse form, for instance in songs, poetry, chant or nursery rhymes. All these phenomena set words to some kind of non-linguistic template, such as a musical tune, a beat pattern, or a poetic metre.

For example, one feature of verse templates is that they are typically composed of chunks (cf. constituent structure as discussed in Chapter 2). The template underlying a nursery rhyme like *Eeny meeny miny moe* could be described as consisting of one big entity with 28 syllable-holders (Example 3.1). However, it can be argued that the template contains further sub-groupings. Those 28 syllables can be divided into 4 parallel lines of 7 syllables each. The 7 syllables in each sequence are also structured: they follow an alternating strong-weak pattern (strong syllables are marked with bold and italics in the table). Knowing this, one can produce a new line by looking for words which match the template.

The relative prominence of syllables is one of the strategies used to divide verse

Example 3.1: The English nursery rhyme *Eeny meeny miny moe.*

1	2	3	4	5	6	7
ee	ny	***mee***	ny	***mi***	ny	***moe***
8	9	10	11	12	13	14
catch	a	***ti***	ger	***by***	the	***toe***
15	16	17	18	19	20	21
if	he	***how***	lers	***let***	him	***go***
22	23	24	25	26	27	28
ee	ny	***mee***	ny	***mi***	ny	***moe***

templates into groups and sub-groups. The lines of this nursery rhyme can be analysed as having four constituents: (+ -)(+ -)(+ -)(+ -), where each parenthesis represents a pair of strong-weak beats in musical terms. Given the pattern of the first three constituents, a final eighth syllable could fill the last weak beat, but we see that this is not done in the case of *Eeny meeny miny moe*. Leaving this empty gap leads to the grouping of the series of 16 strong-weak pairs of the song into 4 lines. Besides, a parallelism between the last sounds of each line (*moe, toe, go, moe* all share the rhyme -/əʊ/) further strengthens the segmentation into lines. The regular alternation of prominence, structural parallelism and systematic use of gaps are recurrent chunking cues used in verse systems; for an overview of these and other cues, see Section 2.2 and Fabb (1997).

How are verse templates created? We address the issue of their emergence by looking into how patterns of syllables evolve in the process of cultural transmission. Previous studies have used an iterated learning paradigm to learn how cultural transmission can make systematic structure emerge out of initially unstructured stimuli (Kirby, Cornish & Smith 2008). Studies of this kind typically show some material for a participant to learn (e.g. random associations between graphical objects and pseudo-words), and then ask the participant to use or reproduce the newly-learnt material. Subsequent subjects are given the output of the preceding participant as their input, so that small changes introduced by individuals can be transmitted from participant to participant, and the overall shape of the initial material evolves as a consequence of the accumulation of these changes. Within this paradigm, each experimental subject serves as a model for a generation in the process of cultural evolution.

Our experiment builds particularly on a previous study (Cornish, Smith & Kirby 2013) where random sequences of colour signals become more learnable and more structured in the transmission process between participants. We follow a similar procedure but employ sequences of syllables as stimuli, which resemble more closely the building blocks of verse. The main question we address is whether random sequences of syllables can evolve into structured patterns; for instance, by using some kind of chunking such as the one found in verse templates (Table 3.1). Our main hypothesis is that the syllable-sequences produced by later generations will be more structured, and hence easier to recall.

By not using words or sentences in the experiment, we abstract away from most real-world verse. Nevertheless, this enables us to have a better control over the experimental variables, since we avoid participants creating patterns on syntactic or semantic basis. More importantly, by using units with no semantic content, our model matches more faithfully the level of the abstract verse template (i.e. the schemas underlying songs and poems) than the level of the lyrics. To be sure, many traditions use nonsense syllables as a strategy to represent, communicate or realise verse templates. English typically represents a weak-strong pattern with the syllables *da-dum* (Fabb & Halle 2008); in Tashlhiyt Berber, verse lines can be composed by using a set of syllables for strong positions (*ay, lay(l), day(l)*), and a different one for weak positions (*a, la(l), da(l)*) (Jouad & Lortat-Jacob 1982; Dell & Elmedlaoui 2008). Similar systems of mnemonic syllables are used in Hindustani and Karnatic music (Clayton 2000; Reina 2013), and West-African drumming traditions (Knight 1984; Stone 1985; Euba 1990). In Example 3.2 we can see how these kinds of syllables are combined into small chunks to produce metrical templates or cycles used to create new songs.

Example 3.2: Sample sequences from two different musical traditions where nonsense syllables are used to realise metrical templates. Vertical lines indicate grouping of syllables within the sequence.

Berber tradition:	*a-lay-da* | *la-la-lay* | *da-lay-la* | *lal*
Hindustani tradition:	*dhin-nā* | *dhin-dhin-nā* | *tin-nā* | *dhin-dhin-nā*

3.2 Method

The experiment follows an iterated learning approach (Mesoudi & Whiten 2008; Cornish, Smith & Kirby 2013; Kirby, Cornish & Smith 2008), where the set of stimuli presented to a participant (i.e. the input) is the set of responses given by the previous participant (i.e. the previous output). Each participant imitates whatever the predecessor has produced, not unlike the routine of Chinese whispers or the Telephone game. Four participants are given an initial set of stimuli created by the experimenter, and each of these four sets further develops independently through transmission chains.

3.2.1 Participants

In total, 40 participants took part in the experiment (mean age = 24.3 years; 22 females, 18 males; left-handed = 7). All of them were native speakers of Dutch, and nine spoke an additional language natively. Participants were recruited at Leiden University and Radboud University (The Netherlands) to take part in a *Syllable Imitation Game*; all signed an informed consent before performing the task (in accordance to Leiden University's LUCL procedure). Each participant was assigned randomly to one of the four transmission chains, but at no point during the experiment were they informed that their input and output stimuli belonged to a chain connecting several subjects.

3.2.2 Stimuli

The first player of each of the four transmission chains received a different collection of semi-random sequences of syllables. A set of four syllables was used to generate all the sequences: {*ban, bi, ta, tin*}. Each of these syllables can be defined as a concatenation of three phonological units: an onset (i.e. the initial consonant), a nucleus (i.e. the vowel), and a coda (i.e. the final consonant or lack thereof); this is summarised in Table 3.1. In our case, each of these features takes one of two possible values: the onset can be [b] or [t]; the nucleus can be [a] or [i]; the coda can be present ([n]) or absent (-). These four syllables were recorded by a female native speaker of Dutch, and were normalised for pitch and intensity. Length was not kept constant because the two items lacking a coda (*bi, ta*) were meant to be shorter.

One important property of the set is that each syllable shares one and only one feature with each of the other three syllables in the set. This means that one can group the syllables in pairs according to their onset, nucleus or coda, resulting in three distinct similarity configurations to which subjects can be sensitive. The choice of onset and nucleus contrasts ([b] vs [t], and [a] vs [i]) seeks to maximise the perceptual distance between syllables.

The third dimension, namely introducing syllables with and without the [n] coda, intends to provide the participants with some cue for prominence. Syllables ending in a consonant tend to attract stress in the world's languages (Gordon 2006), and this is also holds for the Dutch lexicon (Hulst 1984; van Heuven, Hagman, et al. 1988).[1] Making available syllables with potentially different degrees of perceived prominence is relevant given that Dutch poetry uses regular alternations of syllabic prominence (de Groot 1936), and Dutch songs place prominent and non-prominent syllables in a systematic way with respect to melodies (deCastro-Arrazola, van Kranenburg & Janssen 2015).

The initial stimuli for each chain consisted of 30 sequences of 12 syllables each. The sequences were generated by randomly permuting a pool containing 3 tokens of each of the 4 syllable types. The time interval between the onset of a syllable

[1] Other acoustic features such as pitch, duration or spectral balance do provide more unambiguous cues for stress (Heuven & Jonge 2011), but we have not introduced these variables in order to avoid a strong bias in the stimuli.

Table 3.1: Set of syllables used to create the initial sequences. Here displayed according to their defining units; shading represents the presence of the coda [n].

	nucleus = [a]	nucleus = [i]
onset = [b]	*ban*	*bi-*
onset = [t]	*ta-*	*tin*

and the onset of the following syllable was kept constant at 600 ms. Figure 3.10 of the Supplementary Information shows all 30 sequences produced as the initial generation of chain 1 and presented to the first participant of the chain.

3.2.3 Procedure

Participants are instructed that they will listen to sequences of syllables, and they are asked to reproduce them using a keyboard. The sequences are presented in auditory form through headphones (Beyerdynamic DT 880), and, after each of them, an on-screen microphone symbol indicates that it is their turn to reproduce the sequence. This is done using four keys positioned in a row, which correspond to the four syllable types. Each participant is assigned a random key-to-syllable mapping, kept constant throughout the task; if the same mapping was given to every participant, a motor bias or preference for particular keys could be transmitted and amplified over generations.

Before the task starts, participants are given the chance to try the keys in order to adjust the volume and familiarise themselves with the way of typing in syllables. Then, a first training round is presented. The aim of this round is to ensure that they are competent in the key-to-syllable mapping; this mapping has to be memorised, as no visual cues are provided. This round presents (in random order) all the 16 two-syllable combinations which can be generated with the four syllable types.

Subjects are asked to reproduce each two-syllable pattern immediately after it has been played. If the subject presses an incorrect key, a written message appears in the screen requesting to try again. Once the pattern is reproduced correctly, the following pattern gets played.

After the training round is finished, the set of 30 twelve-syllable-long experimental sequences is presented twice, with a pause-screen in between. The routine for these two experimental rounds is similar to the training one: (1) the computer plays a pattern, (2) a microphone sign asks the player to reproduce the pattern,

(3) the score for that individual trial is shown in the screen.[2] Thus, unlike in the training round, a score is shown and, even if mistakes are made, the following sequence is played and the task continues.

After the player has typed in some syllables, if no key is pressed for a period of four seconds, the sequence gets recorded and the following sequence begins. As a way of filtering out obvious slips of memory, sequences of six or less syllables are not registered as a legal response and the same sequence is presented again at a random, later point of the round.

Finally, only the second round of experimental trials is kept for the main analyses below; also, these are the trials which are given as an input to the following participant. The first experimental round, hence, serves as a practice phase for our purposes, but subjects are not told so.

3.2.4 Descriptive measures and statistical analyses

For reproducibility purposes, all the responses (i.e. syllable sequences), as well as the R scripts used to perform the analyses, are included at the online Supplementary Information.[3]

3.2.4.1 Similarity and divergence

Several of the analyses require measuring similarity between sequences, and its counterpart, divergence. For instance, to assess how accurately subjects reproduce a sequence of syllables they have heard, the input and output sequences have to be compared, and a similarity score computed. We do this using a normalised Levenshtein distance metric (Levenshtein 1966). First, the minimum number of insertions, deletions and substitutions to get from sequence A to sequence B is computed. Then, this value is divided by the length (i.e. number of syllables) of the longest sequence (i.e. A or B). The value obtained ranges from 0 (i.e. the sequences are identical), to 1 (i.e. the sequences are maximally divergent). We call this measure the *normalised divergence* or *ndiv*. Its counterpart ($1 - ndiv$) represents the *normalised similarity* measure (*nsim*).

Here, the main use of the normalised similarity is to assess the accuracy with which a subject reproduces an input sequence. The average accuracy for a set of

[2] The score is a value between 0 and 100 reflecting how similar the sequence typed by the participant is with respect to the target sequence. This value corresponds to the *normalised divergence* measure described in Section 3.2.4.1.

[3] https://github.com/vdca/hch2.

sequences can be interpreted as a measure of *learnability* of the set under consideration; the higher the similarity between input and output, the higher the learnability of the input.

3.2.4.2 Measures of structure

Following previous studies (Mathy & Feldman 2012; Cornish, Smith & Kirby 2013), we hypothesise that higher accuracy is partly a result of a more structured, less random input. A number of metrics are used to quantify the amount of structure in a set of sequences.

The normalised divergence is used to evaluate the *dispersion* of a set of sequences. All the sequences in a set are compared in a pairwise manner, and the mean normalised divergence is calculated. In a minimally disperse set, all the sequences would be identical, yielding a normalised divergence score of 0.

We hypothesise that dispersion decreases over generations, i.e. sequences within a set look more alike in later generations, and that this results in a higher overall accuracy. The basis for the dispersion-advantage is that repeated exposure to similar sequences facilitates their recollection. Following this reasoning, if certain syllable patterns (sub-sequences) occur frequently, subjects will identify and recall them with less effort.

We further test this advantage at the level of the individual sequence by analysing the sequence-internal dispersion. We slice each sequence at its midpoint and compute the normalised divergence between the two sections. Sequences with lower internal dispersion indicate a higher degree of repeated material and are expected to develop in later generations.

Less disperse sets of sequences are also more compressible from an information theory point of view, which is related to a lower Kolmogorov complexity (Kolmogorov 1963). File compression algorithms rely on chunking in order to represent the same information in a more efficient way. If the pattern {112233} repeats itself very often in a set of sequences, it can be stored once using a less verbose symbol (e.g. *a*), and then be referred back to every time it is encountered.

In order to obtain a working compressibility measure we use a computer file-compression method. First, we write all the sequences produced by a subject into a file. Then, we compress the file using the Zlib algorithm (Gailly & Adler 2016). Finally, we divide the size of the compressed file by the size of the original file to obtain a compression ratio. Lower values indicate that a file is more compressible because more structure (i.e. more repeated chunks) has been detected by the algorithm.

The emergent regularities in the chains can be a consequence of two general processes: (1) a global bias common to all the participants (due to e.g. general cognition or linguistic bias), and (2) a random bias amplified in a chain-specific way. If the second process is producing at least some of the regularities, the chains should be seen to diverge over time. A way of assessing this is to calculate the evolution of sequence-identifiability. A sequence is identifiable as belonging to its set if the similarity with the sequences in the set (*within-group-nsim*) is higher than the similarity with sequences from other sets in the same generation (*across-group-nsim*). This measure of sequence-identifiability, also known as *lineage divergence* (Matthews, Roberts & Caldwell 2012), is formalised as a proportion:

$$\textit{within-group-nsim}/(\textit{within-group-nsim} + \textit{across-group-nsim}) \tag{3.1}$$

3.2.4.3 Mixed effects models

So far, we have discussed a number of measures which describe some aspect of the (sets of) sequences produced by each subject. The main hypothesis of the experiment is that some of the variation in these measures can be explained by the subject's generation, i.e. by the relative position of the subject within its chain of transmission. Later generations are expected to produce more structured sequences of syllables. We build mixed effects models to assess the amount of variation in the data due to the effect of generation while controlling for variability across chains (Winter & Wieling 2016), using the statistical package `lme4` (Bates et al. 2015).

All the tests follow the same general structure. The outcome of the model is the descriptive measure (e.g. the normalised similarity of an output sequence); the fixed predictor is the generation the measure belongs to; the random predictor is an intercept and slope specific to each of the four chains of transmission.[4] Each of these models is compared to a null model where the generation has been removed but the rest of the predictors are kept unchanged. The fit of each model to the data is assessed through a likelihood ratio test to determine whether the full model bears greater explanatory power, hence showing support for the predictor under consideration (Roberts, Winters & Chen 2015).

[4] In the cases where the metric refers to the production of a subject as a whole, e.g. dispersion, only a random intercept was included, because the available degrees of freedom did not allow for random slopes.

3.2.4.4 Interesting patterns

The main analyses involve testing whether structure increases within sets of sequences produced by a participant, or within individual sequences, and we tackle this issue by employing a number of proxies for structure (Section 3.2.4.2). These analyses only attempt to explain whether the initial randomness of the computer-generated sets is somehow reduced by the transmission process; however, we also want to inspect the concrete regularities in a principled way by searching for emerging syllable patterns.

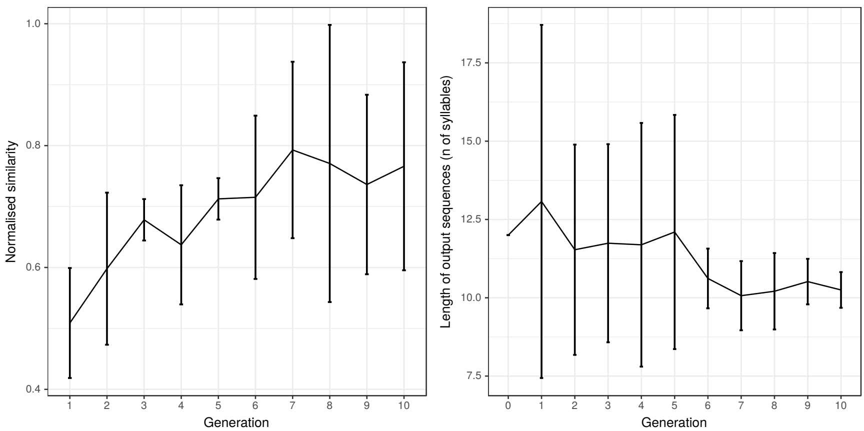
In order to examine the properties of the emerging structures, we analyse the extent to which each possible ngram of size 2 through 4 is over- or under-represented within each subject. We first create a baseline of expected frequencies consisting of one million sequences generated in the same way the sequences for the initial generation of each chain were generated; i.e. a base sequence containing three instances of each of the four kinds of syllables (*ban*×3, *bi*×3, *ta*×3, *tin*×3) is randomly shuffled one million times. We then compare the frequencies observed in each subject to the baseline frequency.

For each possible ngram and for each of the ten sets of sequences produced in a chain, we calculate how many sequences contain the ngram.[5] The raw count is divided by the number of sequences in the set: 30 in the experimental subjects, and one million in the random baseline. Additive smoothing is applied in order to avoid zero probabilities (Chen & Goodman 1996). The ngram frequency for the subject is then divided by the baseline frequency to obtain an odds ratio, which we log transform for visualisation purposes. Ngrams with a positive ratio are considered to be over-represented compared to the random baseline.

In order to identify ngrams with a robust increase in popularity across chains, we build a mixed model with generation as a predictor of ngram frequency ratio, and chain as a random effect. We also run equivalent linear regression models for each chain to determine ngrams with a chain-specific increase. In both kinds of models, we focus on ngrams where generation is a significant predictor of the ngram becoming over-represented, and the ngram reaches a mean frequency of at least 0.2 by the last generation. We calculate the significance threshold by applying a Bonferroni correction based on the total number of ngrams which can be generated using the available syllables.

Sequence boundaries represent a context with a particular potential of develop-

[5] We intentionally do not count the total number of instances of the ngram because we want to assess how representative individual patterns are of the set produced by a subject as a whole. This method avoids an ngram's frequency being inflated by its repetition within a single sequence (Conklin 2010).



(a) Learnability. **(b)** Length.

Figure 3.1: Evolution of sequence learnability over generations. Error bars indicate the 95% confidence interval.

ing fixed or conventional patterns, as illustrated by cadence or rhyme in both music and poetry. We test whether sequence openings become increasingly different from endings by applying the identifiability measure described above. For each subject, we compute an index of how identifiable the first syllable is as belonging to the sequence-opening syllables, as opposed to the syllable-closing syllables.

3.3 Results

3.3.1 Learnability

Overall, subjects belonging to later generations reproduce their input sequences more accurately (Figure 3.1a). Hence, we can say that the original random sets of sequences given to the first generation of each chain get more learnable as they get modified by participants. Figure 3.9 of the Supplementary Information exemplifies this by showing the evolution of a single sequence in the first chain; after the fifth generation, the sequence stabilises and subsequent subjects reproduce it very accurately. Overall, subjects in the initial generation score an average of 0.51, reaching a score of 0.77 by generation 10. When comparing the null model to a model with generation as a fixed predictor (cf. Section 3.2.4.3), we obtain a statistically significant improvement in prediction (Table 3.2, Similarity).

Table 3.2: Results of the full mixed models compared to the correspondent null model where the predictor of interest (generation) has been removed.

	Model	Estimate	Std. Error	χ^2	Pr $(> \chi^2)$
1	Similarity	0.0255	0.00771	5.28	0.0215
2	Similarity with length control	0.023	0.00708	5.17	0.023
3	Length	-0.249	0.0952	3.99	0.0459
4	Set dispersion	-0.00333	0.00155	4.36	0.0367
5	Sequence dispersion	-0.0153	0.00188	65.3	6.56e-16
6	Compression	-0.00263	0.00092	7.45	0.00633
7	Identifiability	0.000545	0.00019	8.16	0.00428
8	Boundary identifiability	0.0209	0.00204	101	9.71e-24

3.3.2 Length

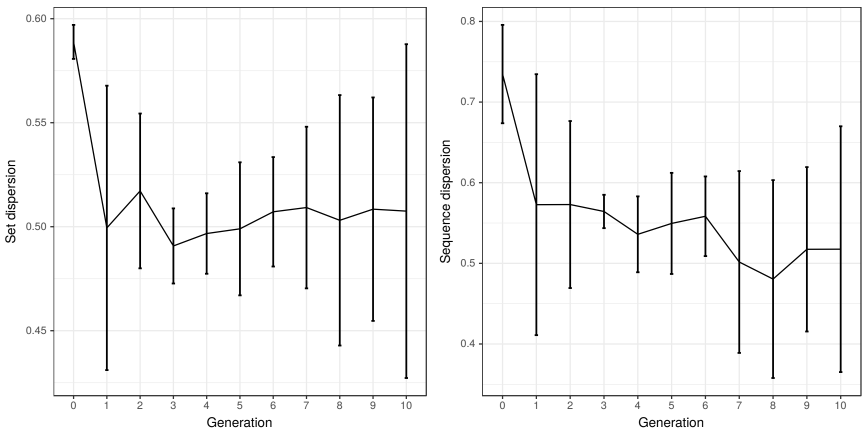
The random sequences given to the first generation are all twelve-syllable long. However, their mean length decreases over time, stabilising at a length of ∼10 syllables (Figure 3.1b). Longer sequences will be typically harder to remember, i.e. length is inversely correlated with learnability ($r = -.28$). Hence, the number of syllables in the input sequences needs to be controlled for in order to assess whether the improvement in accuracy (Figure 3.1a) is just a function of length (Figure 3.1b), or depends on some other factor.

A mixed effects model with generation *and* length of the input as predictors of accuracy (and random effects for chain) performs significantly better than a model with only length as a fixed predictor (Table 3.2, Similarity with length control). Hence, length alone cannot account for the increase in learnability of the sequences.

3.3.3 Dispersion

If all the sequences in a set look alike, it can become easier for a subject to reproduce them accurately. We hypothesise that sequence *sets* of later generations are more learnable because they have less internal variation. We test whether indeed dispersion decreases over time, correlating with the increasing accuracy shown above. Sets of sequences do become less disperse, but the decrease is robust only when going from the initial random state to the first generation of participants; later generations keep a steady dispersion measure of ∼0.5 (Figure 3.2a).

Adding generation as a predictor of set dispersion significantly improves the



(a) Within-set dispersion.

(b) Within-sequence dispersion.

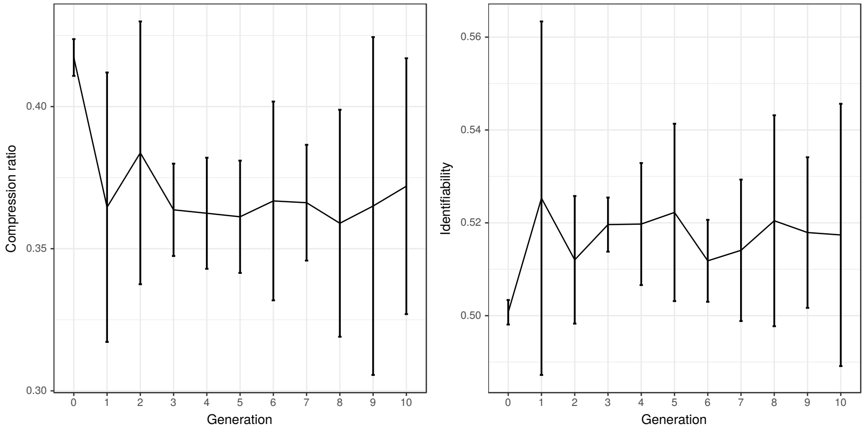
Figure 3.2: Evolution of dispersion within sequence sets and within individual sequences, over generations. Error bars indicate the 95% confidence interval.

explanatory power of the null mixed effects model (Table 3.2, Set dispersion). Nevertheless, the effect disappears if the initial random state (generation 0) is removed, indicating that the decrease takes place as soon as a human subject intervenes, but is not amplified as a function of iterated learning (a variant of the models in Table 3.2 excluding the initial generation is reproduced in Table 3.5 of the Supplementary Information).

The decrease of sequence-internal dispersion, however, shows a more robust cumulative effect over generations, as shown in Figure 3.2b. The first and second halves of sequences resemble each other more in later generations, indicating that each subject increases the amount of repetition of sequence-internal patterns. In Table 3.5 of the Supplementary Information we confirm that generation remains a significant predictor of the decrease in sequence-internal dispersion even when the computer-produced generation is excluded from the analysis. This means that dispersion does not decrease categorically, but shows a cumulative effect.

3.3.4 Compression

The evolution of compressibility resembles that of within-set dispersion: human-produced sets of sequences are more compressible than the randomly generated ones. However, once the compression ratio drops with the first participant, it does not further decrease in a robust way across chains (Figure 3.3a).



(a) Compression. **(b)** Identifiability.

Figure 3.3: Evolution of sequence compression and identifiability over generations. Error bars indicate the 95% confidence interval.

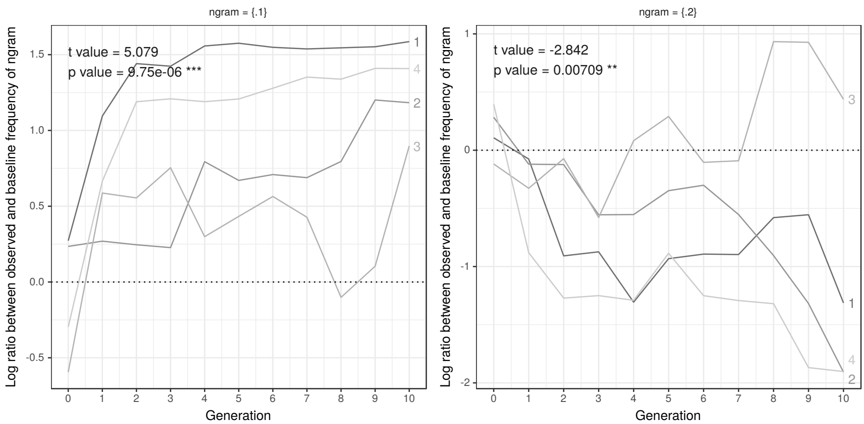
Adding generation as a predictor of set compressibility significantly improves the explanatory power of the null mixed effects model when including the computer-generation (Table 3.2, Compression), but not when excluding it (Table 3.5, Supplementary Information). If we compare the compression ratio of the random sets, to those produced by the participants, we obtain a mean difference of 0.051 ($t = 12.676, p = 5.027e - 13$).

3.3.5 Identifiability

Overall, sequences from later generations are more identifiable as belonging to their chain (Figure 3.3b). This suggests that at least some of the strategies by which chains develop structure are chain-specific. The average identifiability index for the initial random generation approximates 0.5; i.e. within-group similarity is as high as across-group similarity. Some of the increase in identifiability can be attributed to the effect of generation (Table 3.2, Identifiability).

3.3.6 Interesting patterns

For convenience during the analyses, syllables are encoded with the integers {1, 2, 3, 4} corresponding to the syllables {*ban, bi, ta, tin*}. Besides, the start and end of sequences are encoded by a special boundary symbol $\{.\}$, so a bigram like $\{.1\}$



(a) Global increase of {.1} (*ban*).

(b) Idiosyncratic increase of {.2} (*bi*).

Figure 3.4: Frequency of the sequence-opening patterns {.1} (*ban*) and {.2} (*bi*) relative to the baseline. A log ratio above 0 means that the pattern is over-represented in that generation.

represents the single syllable *ban* opening a sequence.

We have investigated distinctive patterns at four different levels: at the level of the syllable, and at three sub-syllabic dimensions: onset, nucleus and coda (see Section 3.2.2 for details on the syllabic structure). Table 3.3 displays the ngrams of size 2, 3 and 4 which are increasingly over-represented over generations.[6] Note that all of these contain a boundary symbol, meaning that they belong to the sequence-initial or sequence-final contexts. Figure 3.4a illustrates the increase in popularity for the opening pattern {.1} (*ban*) over generations.

The emergent pervasiveness of this opening can be observed by comparing the first and last generations of chain 1 (Figures 3.10 and 3.11 of the Supplementary Information). More generally, by visually inspecting the sets of sequences plotted as phase-space diagrams (Figure 3.12 of the Supplementary Information), we can notice how the initial generations produce all syllable combinations with a similar frequency, while later participants persist on a few ngrams, reflected in the emerging geometric patterns (Ravignani 2017).

In some instances, an individual chain can develop a preference for an ngram, while the general trend of the other three chains is to gradually disprefer the pattern. These chain-specific patterns are listed in Table 3.4, and Figure 3.4b shows

[6] The Bonferroni-corrected significance thresholds differ for syllabic patterns (α = 9.92e − 05) and sub-syllabic feature patterns (α = 8.93e − 04); see Section 3.2.4.4.

Table 3.3: Patterns which increase robustly across generations. The numeric codes follow the syllables in alphabetical order: 1 = *ban*, 2 = *bi*, 3 = *ta*, 4 = *tin*. A hyphen indicates that the syllable lacks a coda.

	Feature	Context	Pattern	t statistic	Pr ($>t$)
1	syllable	initial	`.1`	5.08	9.75e-06
2	onset	initial	`.b`	6.82	3.82e-08
3	onset	initial	`.bb`	4.36	9.13e-05
4	onset	final	`t.`	3.76	0.000561
5	nucleus	initial	`.aii`	4.45	7.05e-05
6	coda	initial	`.n-n`	3.63	0.000817
7	coda	final	`-.`	4.16	0.000169

Table 3.4: Ngrams with a significant increase in preference one chain, coupled with a global trend to disprefer the pattern. Chain-specific t values are the result of linear regressions on a single chain; global t values are computed with a mixed model including all chains.

	Context	Chain	Pattern	Chain's t	Global t	Pr ($>t$)
1	internal	4	`134`	2.48	-2.42	0.0203
2	initial	3	`.2`	3.12	-2.84	0.00709
3	initial	2	`.4`	2.33	-2.29	0.0273
4	initial	3	`.--`	2.27	-2.1	0.0423
5	initial	2	`.41`	5.68	-2.19	0.0349
6	final	4	`iaa.`	2.44	-2.76	0.00884

the evolution of a sample ngram.

Chain 3, for instance, develops a preference for sequences starting with the syllable *bi* or two light syllables (i.e. syllables without a final [n]), while the general tendency is to decrease these openings (Table 3.4) in favour of a pattern with a heavy syllable like *ban* (Table 3.3). On the sequence-final context, there is a global tendency to end with the nuclei pattern {`ia.`}, but chain 4 idiosyncratically favours the ending {`iaa.`}.

As the previous results reveal, distinctive patterns tend to emerge at the boundaries of sequences, but the over-represented opening syllables seem to differ from the closing ones. By running the identifiability analysis (Section 3.2.4.2) on the opening and closing unigrams, we can test whether sequence-initial syllables pro-

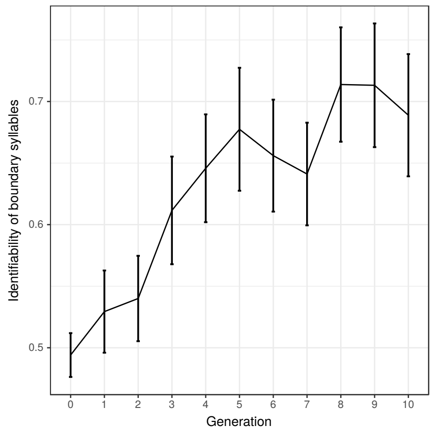


Figure 3.5: Identifiability of syllables as sequence openers or closers. All four chains averaged. Error bars indicate 95% confidence interval.

gressively become more similar to each other, and more unlike the sequence-final syllables. Figure 3.5 indicates that, indeed, opening and closing syllables become increasingly polarised as a function of generation. Opening syllables start off being indistinguishable from closing syllables (mean identifiability at generation 0 = 0.49), and exhibit a steady divergence over generations which proves robust across all four chains (Figure 3.5 and Table 3.2, Boundary identifiability).

3.4 Discussion

The starting point of the experiment are sequences which randomly alternate the syllables *ban, bi, ta, tin*. After a process of iterated learning involving four chains of transmission and ten generations of participants, the original sequences become (1) easier to recall, (2) shorter, (3) more structured. As suggested by the analysis of significant patterns and sequence identifiability, some of the emerging regularities are common to all four chains, while others are chain-specific.

Tendencies which emerge across the board are most likely attributed to biases shared by all the participants. These biases can be related to (1) basic aspects of human cognition involved in sequence perception and recall, (2) phonological properties of the Dutch language, which all participants speak natively.

Regarding basic cognitive biases, we can highlight that all chains become consistently more compressible and less disperse than the initial random sets. This is the result of a number of syllabic patterns gaining popularity at the expense of others. Hence, we can infer that sequences are not processed as unitary entities; instead, sub-patterns within sequences (i.e. chunks) are recognised and reproduced, leading to an increasingly uneven distribution of ngrams.

The data suggest that participants are engaging in a chunking strategy to deal with the task at hand. Essentially, they are asked to remember a sequence too long to fit in working memory, and then reproduce it. Working memory can hold around four items (Mathy & Feldman 2012; Chen & Cowan 2005), yet, crucially, items need not be unitary but can contain further items within themselves. This effectively can expand our working memory capacity to a span of tens of items (Ericsson, Chase & Faloon 1980). We apply chunking strategies unconsciously, and even 14-month-old infants combine chunks into super-chunks under experimental conditions in order to expand the limits of working memory (Rosenberg & Feigenson 2013).

Given this readiness to divide temporal sequences, it is unsurprising that human music and poetry rely heavily on segmenting and repeating motifs (cf. Chapter 2; Tierney, Russo & Patel 2011; Rubin 1995). Moreover, this aspect of cognition is not restricted to humans; a number of bird species (e.g. bullfinches, nightingales) learn and reproduce sound sequences, and are shown to engage in chunking too (Nicolai et al. 2014).

Unlike in the Simon game experiment (Cornish, Smith & Kirby 2013), however, the chunking bias does not show a cumulative effect on within-set dispersion and compressibility: the effect appears in the very first participant of a chain, and then remains stable along the following generations. Given that both experiments only run over ten generations, we do not know whether the dispersion, for instance, would continue to decrease or remain stable in further generations.

This earlier decrease and stabilisation compared to the colour experiment may stem from a greater difficulty in the task. This can force the participants to focus on less detail, effectively boosting the chunking effect. Crucially, the experiments differ in the modality used for stimuli presentation (visual vs auditory), and the input method used by the subjects (visual cues vs no cues), which can make the task more challenging.

A measure where a cumulative effect does take place is the gradual divergence of opening and closing syllables (Figure 3.5); syllables get specialised in all chains. On the one hand, the fact that a specialisation takes place can still be driven by some aspect of general cognition; on the other hand, the specific macro-phonotactics of which syllables are preferred on the left or right boundary are arguably language-specific.

In this experiment, we hypothesise that a Dutch bias drives the emerging preference for starting sequences with the syllable *ban* (more generally, a heavy syllable), and ending sequences with the syllable *ta* (more generally, a light syllable). One kind of support for this bias comes from the properties of the Dutch lexicon: heavy syllables attract stress, and most content words have initial stress (van Heuven, Hagman, et al. 1988). Another kind of evidence is provided by acquisition data: children learning Dutch produce mostly disyllabic words starting with a stressed syllable (Fikkert 1994). This trochaic bias has been described for other Germanic languages (Pater 1997), but some non-Germanic languages like Hebrew or Portuguese show either no preference between iambic or trochaic disyllables (Santos 2003), or a preference in the opposite direction (Segal & Kishon-Rabin 2012). Follow-up experiments can exploit these differences to set apart general biases from those related to particular phonological systems.

So far, we have only discussed tendencies which consistently appear in all four chains. Certain patterns, nevertheless, gain preference in a single chain, while the other three follow the opposite direction (Table 3.4). We can refer to these as *arbitrary preferences*, since, if they were determined by general cognitive or linguistic biases, all four chains should have developed them. Instead, we can think of these biases as pressures which shape the pool of possible patterns in a particular direction. The pressures get amplified in the process of cultural evolution, but can not explain in a deterministic way the exact patterns which will prevail. Even in a non-creative task as the one we present here, individual subjects move the syllable sequences in idiosyncratic directions, some of which are picked up and amplified by further generations.

The emergence and evolution of verse patterns in the world's languages can be conceptualised in this way. Among the virtually infinite combinations of e.g. syllables, phonological features, or drum patterns, our shared cognitive and linguistic background creates a biased baseline. Only a subset of these combinations is used as the basis to create songs and poems, and this subset is continually innovated and filtered in the process of cultural evolution.

3.5 Conclusion

In the present study we show that random sets of syllables develop an increasingly systematic structure through iterated learning, where individuals try to reproduce the stimuli produced by the predecessor. Cognitive, linguistic and other subject-specific biases shape the sequences in particular ways, while the learn-and-reproduce procedure magnifies the biases inherited from previous generations. Both the emergent features and the iterated learning mechanism resemble aspects of versification pervasive in human societies, making this paradigm a suitable one to model the emergence of verse templates.

References

Bates, D., M. Mächler, B. Bolker & S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Chen, S. F. & J. Goodman. 1996. An empirical study of smoothing techniques for language modeling. In *Proceedings of the 34th annual meeting on association for computational linguistics* (ACL '96), 310–318. Stroudsburg, PA, USA: Association for Computational Linguistics.

Chen, Z & N Cowan. 2005. Chunk limits and length limits in immediate recall: a reconciliation. *Journal of experimental psychology, learning, memory, and cognition* 31(6). 1235–1249.

Clayton, M. 2000. *Time in Indian music: rhythm, metre, and form in North Indian rāg performance*. Oxford University Press.

Conklin, D. 2010. Discovery of distinctive patterns in music. *Intelligent Data Analysis* 14(5). 547–554.

Cornish, H, K Smith & S Kirby. 2013. Systems from sequences: an iterated learning account of the emergence of systematic structure in a non-linguistic task. In M Knauff, M Pauen, N Sebanz & I Wachsmuth (eds.), *Proceedings of the 35th annual conference of the cognitive science society*, 340–345. Austin (TX): Cognitive Science Society.

de Groot, A. W. 1936. De structuur van het vers. *De Nieuwe Taalgids* 30. 197–212.

deCastro-Arrazola, V, P van Kranenburg & B Janssen. 2015. Computational textsetting analysis of Dutch folk songs. In O Adam & D Cazau (eds.), *Proceedings of the 5th international workshop on folk music analysis*, 51–55. Paris (France): Association Dirac.

Dell, F & M Elmedlaoui. 2008. *Poetic meter and musical form in Tashlhiyt Berber songs*. Köln: Rüdiger Köppe.

Ericsson, K., W. G. Chase & S. Faloon. 1980. Acquisition of a memory skill. *Science* 208(4448). 1181–1182.

Euba, A. 1990. *Yorùbá drumming: the dùndún tradition*. Vol. 21. E. Breitinger, Bayreuth University.

Fabb, N. 1997. *Linguistics and literature*. Blackwell Publishers.

Fabb, N & M Halle. 2008. *Meter in poetry: a new theory. With a chapter on Southern Romance meters by Carlos Piera*. Cambridge University Press.

Fikkert, P. 1994. *On the acquisition of prosodic structure*. Holland Institute of Generative Linguistics.

Gailly, J.-l. & M. Adler. 2016. Zlib compression library.

Gordon, M. K. 2006. *Syllable weight: phonetics, phonology, typology*. New York: Routledge.

Heuven, V. J. van & M. de Jonge. 2011. Spectral and temporal reduction as stress cues in Dutch. *Phonetica* 68(3). 120–132.

Hulst, H. Van der. 1984. *Syllable structure and stress in Dutch*. Foris.

Jouad, H. & B. Lortat-Jacob. 1982. Les modèles métriques dans la poésie de tradition orale et leur traitement musical. *Revue de musicologie* 67(1-2). 174–197.

Kirby, S, H Cornish & K Smith. 2008. Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31). 10681–10686.

Knight, R. 1984. Music in Africa: the Manding contexts. *Performance practice: Ethnomusicological perspectives* (12). 53.

Kolmogorov, A. N. 1963. On tables of random numbers. *Sankhyā: The Indian Journal of Statistics, Series A*. 369–376.

Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics–Doklady* 10(8). 707–710.

Mathy, F. & J. Feldman. 2012. What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition* 122(3). 346–362.

Matthews, C., G. Roberts & C. A. Caldwell. 2012. Opportunity to assimilate and pressure to discriminate can generate cultural divergence in the laboratory. *Evolution and Human Behavior* 33(6). 759–770.

Mesoudi, A. & A. Whiten. 2008. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 363(1509). 3489–3501.

Nicolai, J., C. Gundacker, K. Teeselink & H. R. Güttinger. 2014. Human melody singing by bullfinches (*Pyrrhula pyrrula*) gives hints about a cognitive note sequence processing. *Animal cognition* 17(1). 143–155.

Pater, J. 1997. Minimal violation and phonological development. *Language Acquisition* 6(3). 201–253.

Ravignani, A. 2017. Visualizing and interpreting rhythmic patterns using phase space plots. *Music Perception: An Interdisciplinary Journal* 34(5). 557–568.

Reina, R. 2013. *Karnatic rhythmical structures as a source for new thinking in Western music*. Amsterdamse Hogeschool voor de Kunsten, Conservatorium van Amsterdam Doctoral dissertation.

Roberts, S. G., J. Winters & K. Chen. 2015. Future tense and economic decisions: controlling for cultural evolution. *PloS one* 10(7). e0132145.

Rosenberg, R. D. & L Feigenson. 2013. Infants hierarchically organize memory representations. *Developmental science* 16(4). 610–621.

Rubin, D.-C. 1995. *Memory in oral traditions: the cognitive psychology of epic, ballads, and counting-out rhymes*. Oxford University Press.

Santos, R. S. 2003. Bootstrapping in the acquisition of word stress in Brazilian Portuguese. *Journal of Portuguese Linguistics* 2(1). 93–114.

Segal, O. & L. Kishon-Rabin. 2012. Evidence for language-specific influence on the preference of stress patterns in infants learning an iambic language (Hebrew). *Journal of Speech, Language, and Hearing Research* 55(5). 1329–1341.

Stone, R.-M. 1985. In search of time in African music. *Music Theory Scpectrum* 7. 139–148.

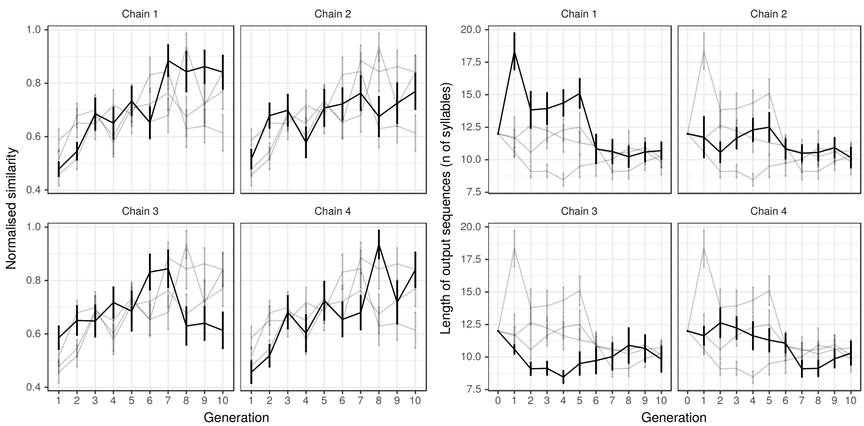
Tierney, A. T., F. A. Russo & A. D. Patel. 2011. The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences* 108(37). 15510–15515.

van Heuven, V., P. J. Hagman, et al. 1988. Lexical statistics and spoken word recognition in Dutch. *Linguistics in the Netherlands*. 59–68.

Winter, B. & M. Wieling. 2016. How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *Journal of Language Evolution* 1(1). 7–18.

Supplementary Information

In the following figures, we plot the evolution of the different metrics for each chain separately. This enables the tracking of global and local trends with more detail. Note that the error bars of the main-text plots indicated the 95% confidence interval based on the average of all 4 chains, while the following intervals are based on 30-sequence sets of single chains.

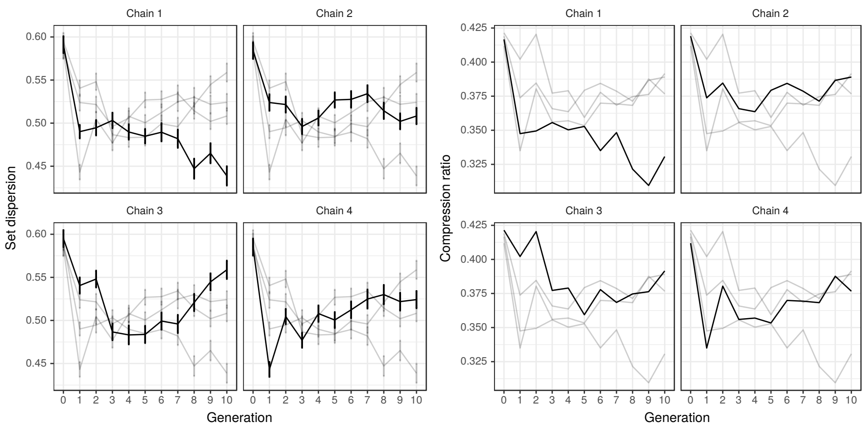


(a) Learnability. **(b)** Length.

Figure 3.6: Evolution of sequence learnability and length over generations. Error bars indicate the 95% confidence interval.

Table 3.5: Results of the full mixed models compared to the correspondent null model where the predictor of interest (generation) has been removed. These models only include human-generated data, i.e. the initial generation has been excluded.

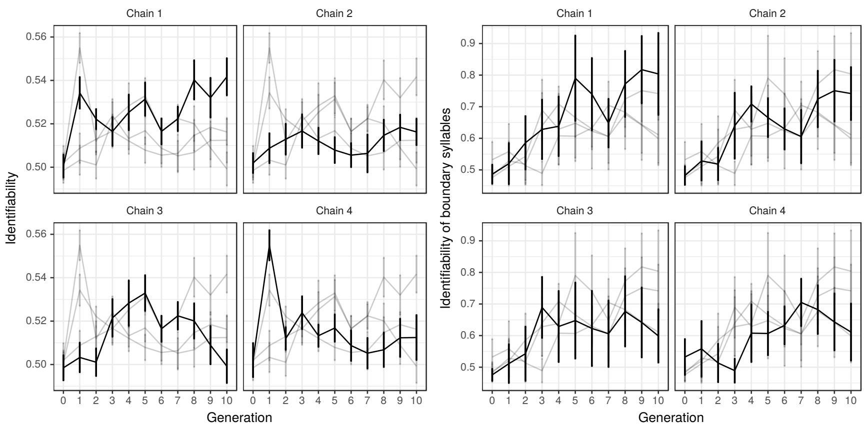
	Model	Estimate	Std. Error	χ^2	Pr $(> \chi^2)$
1	Length	-0.282	0.158	2.34	0.126
2	Set dispersion	0.000718	0.00129	0.311	0.577
3	Sequence dispersion	-0.00848	0.00217	15.1	0.000102
4	Compression	-0.000432	0.000803	0.288	0.591
5	Identifiability	-0.000323	0.000219	2.18	0.14
6	Boundary identifiability	0.0189	0.00245	58.4	2.15e-14



(a) Dispersion.

(b) Compression.

Figure 3.7: Evolution of sequence dispersion and compression over generations. Error bars indicate the 95% confidence interval.



(a) Sequence identifiability.

(b) Boundary identifiability.

Figure 3.8: Evolution of sequence identifiability, and identifiability of syllables as sequence openers or closers, over generations. Error bars indicate the 95% confidence interval.

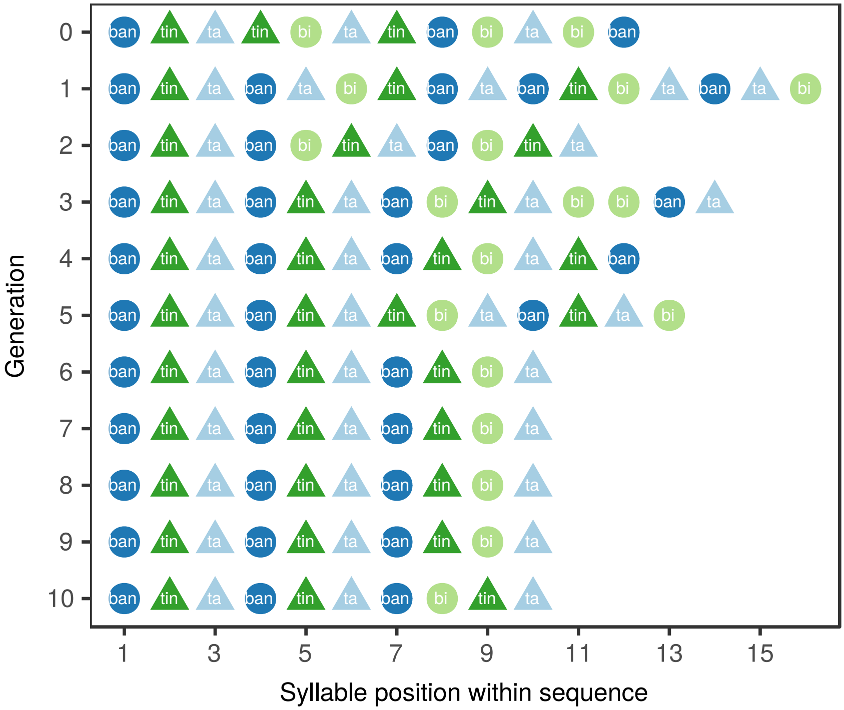


Figure 3.9: Evolution of sequence lineage 15 in chain 1, from generation 0 (random), to generation 10.

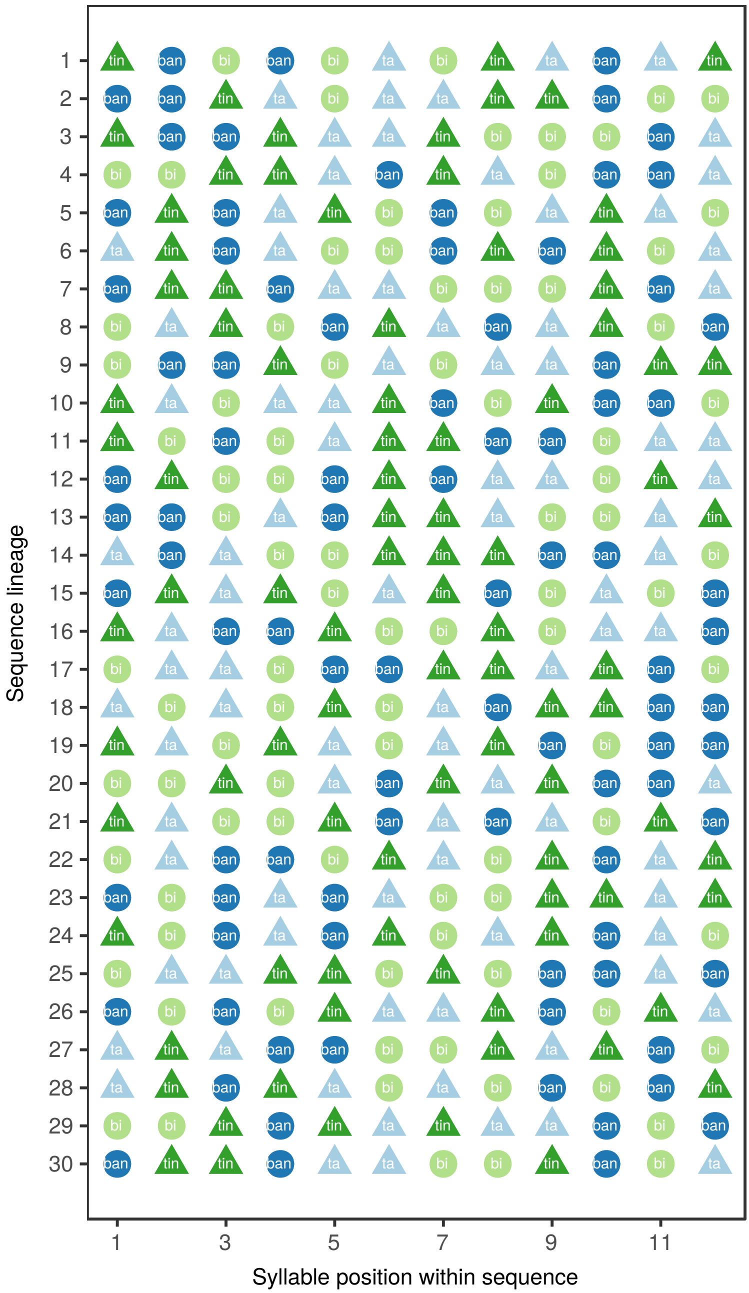


Figure 3.10: All 30 computer-generated sequences (i.e. initial state) of chain 1.

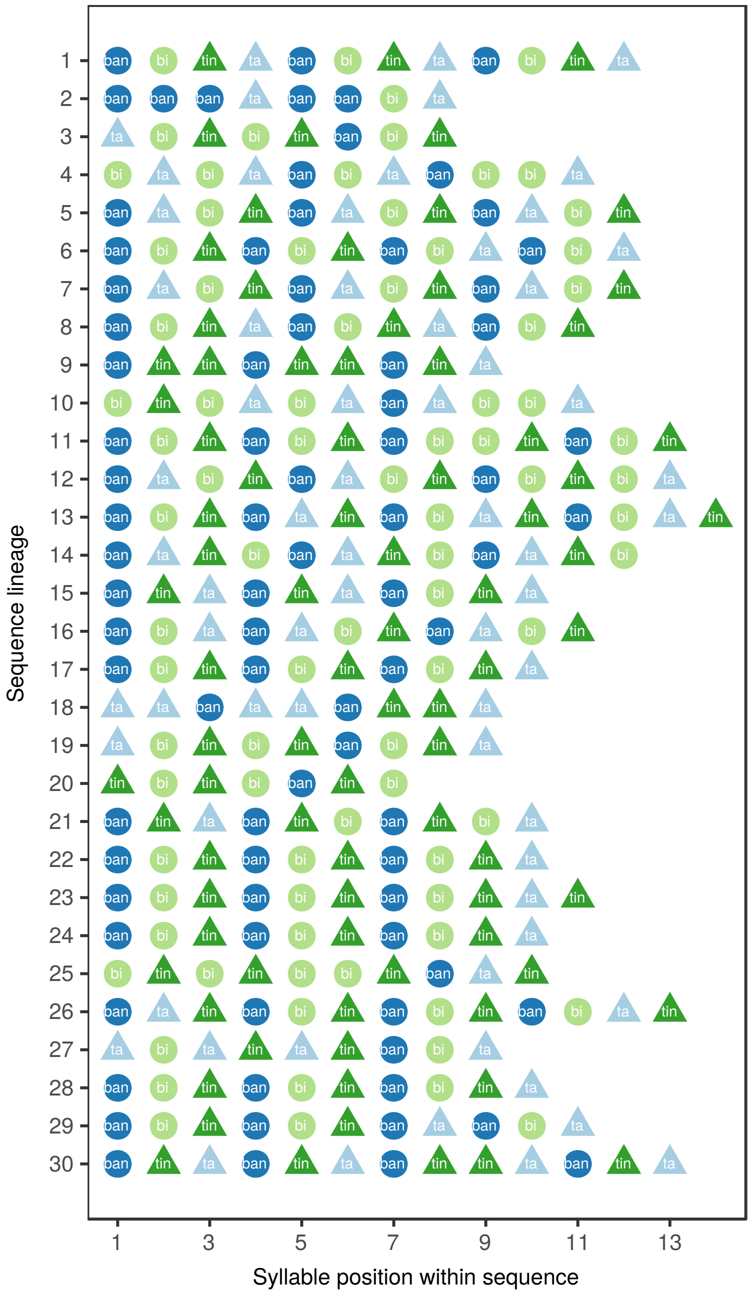


Figure 3.11: All 30 sequences produced by the last subject in chain 1.

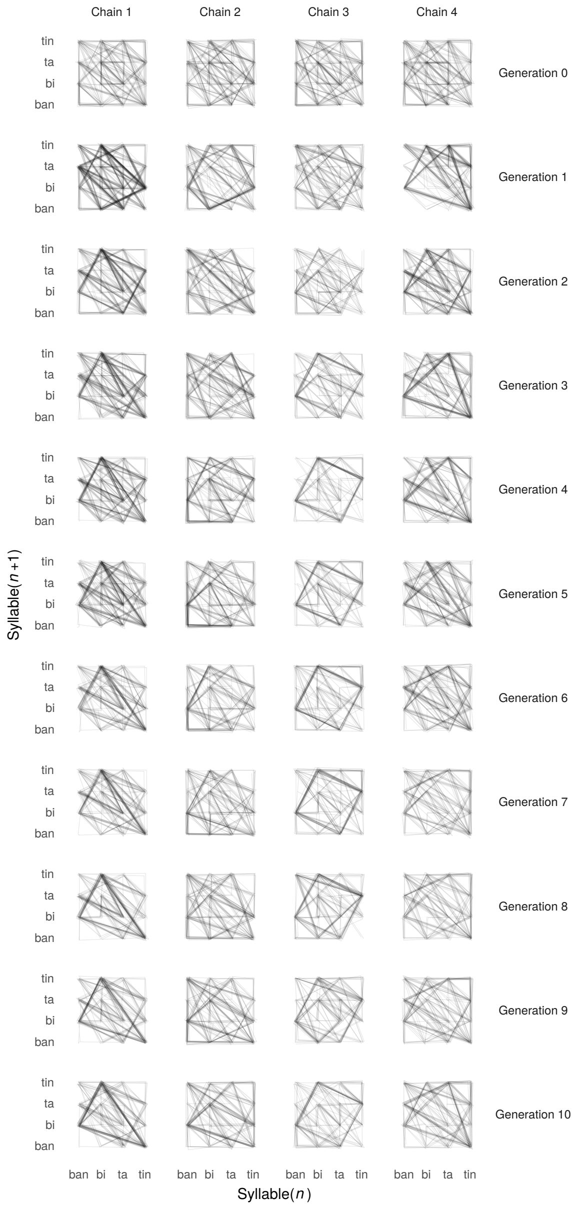


Figure 3.12: Visualisation of all the sequences as phase-space diagrams (Ravignani 2017). Each syllable receives a coordinate based on its own value (x-axis), and the value of the following syllable (y-axis). Consecutive syllables are connected with a line, and emerging geometric patterns represent often-visited syllabic ngrams.

Part II

How verse templates are instantiated

4 A corpus study of final strictness in verse

4.1 Introduction

The words used in verse are subject to a number of constraints which are absent in everyday speech. By analysing how songs and poems are structured we can observe, for instance, that the discourse is organised into lines of similar length, that a pulse can be perceived by the regular alternation of strong and weak syllables, or that a number of adjacent lines end with exactly the same phonemes.

However, there seems to exist an asymmetry in the way these constraints are arranged: the beginning of lines are left relatively free, and later parts of the line are more constrained. This can be the result of constraints specific to the end of lines (e.g. rhyme), or due to some general constraint (e.g. alternate weak and strong syllables) being more stringent later in the line.

The phenomenon has been most prominently mentioned with reference to early verse corpora, such as Ancient Greek (Prince 1989) or Sanskrit (Arnold 1905). However, it has also been noted that it is an "almost constant feature of numerous widely differing metrical systems of the world" (Kiparsky 1968:138), which has led to the hypothesis that it "is in fact just a specific manifestation of a universal principle" (Hayes 1983:388).

The Strict End Hypothesis (henceforth, SEH) covers a number of predictions, roughly summarised by the statement that "correspondence to a metrical pattern tends to be lax at the beginnings of units; strict at the ends" (Hayes 1983:373). In order to verify the extent of the hypothesis, one needs to be able to verify whether a given sample of verse conforms to it or not. As I will argue in the following section (4.2), the concept of final strictness is too coarse to undertake a typological study. As defended by Bickel (2007:247), the same problem applies to many traditional descriptive variables in linguistic typology (e.g. incorporation); instead of seeking universal definitions of such variables, it would be more productive, he maintains, to encode finer-grained variables: "such variables allow capturing rather than ignoring diversity, and they stand a greater chance to be codable in replicable ways across many languages."

Section 4.2 discusses the properties of a number of instances of final strictness from different languages, and proposes a formal division into two main classes of strictness. Beyond the characterisation of the sub-variables composing strictness in verse, there is also the issue of gradience: it is unfeasible to describe with precision a *relative* higher strictness in a *relatively* late position of a line without some sort of quantification. Section 4.3 showcases how a particular type of final strictness (the kind found in Vedic verse) can be quantified. This enables a finer-grained characterisation of the degree of strictness and finality, and allows systematic cross-linguistic comparisons. In the final discussion, I present a number of potential explanations for the different types of final strictness, and point to further ways of testing them empirically.

4.2 Types of strictness

The Strict End Hypothesis has a number of possible definitions, which means that it is flexible enough to cover a wide range of phenomena. However, if we want to systematically survey and compare like with like manifestations of the SEH, we need to be precise about the scope of the term. Characterising one or several types of final strictness becomes crucial if we want to (1) verify or falsify the universality of the phenomenon, and (2) investigate the possible causes of its pervasiveness. There are at least three problematic issues with the notion of SEH.

First, it is unclear whether the strictness applies only at one particular constituent level (e.g. the line), or to any kind of metrical constituent (e.g. stanzas, hemistichs, feet). This can be a source of confusion, since evidence and counter-evidence for the hypothesis can refer to completely different metrical domains. The article by Zwicky & Zwicky (1986) *Patterns first, exceptions later* offers a potential counter-example to the hypothesis, but in this case the strictness applies at the level of the stanza: limericks show greater regularity at the beginning lines. Still, most of the remarks on final strictness make reference to the level of the line. Besides, lines are possibly the only defining constituents of verse (Fabb 2015), so it is a suitable domain for a working definition of SEH.

A second source of heterogeneity in the phenomena is where the focus of the strictness asymmetry is set: (1) an exceptional freedom at the very beginning, (2) an exceptional strictness at the very end, (3) a gradual increase of strictness (or decrease of freedom) encompassing the line as a whole. The trochaic inversion common in English iambic pentametre refers to an exceptional looseness at the beginning of the line (Hayes 1983). Rhyme constraints, on the other hand, usually target just the end of lines. Finally, other instances of SEH are reported to apply gradually, with strictness increasing from the beginning to the end of the line (e.g. Finnish *Kalevala* verse, Kiparsky 1968).

Third, in order to speak of strictness, one needs to posit a rule or restriction which can then be satisfied or violated; however, a variety of features (e.g. syllable weight, phonemes) can be subject to restrictions in verse templates. The cognitive representation of these features differ, and it is unclear whether strictness phenomena on different features can be directly compared as produced by a shared cause. Next, I review the main types of restrictions for which the SEH has been mentioned.[1]

4.2.1 Final strictness phenomena

Syllable type

The prototypical example of final strictness relies on the existence of a template which shows restrictions on syllable type. In Somali [Afro-Asiatic; Cushitic][2] *geeraar* verse lines, for instance, the third and fourth syllabic positions are required to be heavy and light respectively; the preceding positions, however, do not show a strong preference for either of the syllable types (Banti & Giannattasio 1996:99). The same kind of stricter syllabic weight pattern is found in Sanskrit [Indo-European; Indo-Iranian] (Arnold 1905). Languages where syllabic stress rather than weight is constrained in verse, e.g. English or Dutch [Indo-European; Germanic], also show lines with looser beginnings (Hayes 1983; de Groot 1936).

Syllable-to-position association

The number of syllables or morae associated to a metrical position is fixed in many traditions. However, certain deviations are permitted, such as the resolution, where a strong position, typically filled with a heavy syllable, is realised instead as two weak syllables. This kind of freedom is frequent in Greek [Indo-European; Graeco-Phrygian] iambic trimetre, but it does not occur in the final two strong positions (Prince 1989:61). Similarly, in Somali *gabay* metre, the strong positions of the first half-line can be resolved as two light syllables; in the second

[1] Several of the examples I mention are discussed by Fabb (2002), with additional analysis on the SEH.

[2] In the first mention of a language in the current section, I include its genealogical information between brackets, extracted from Glottolog (Nordhoff & Hammarström 2011).

half, however, only one of the three strong positions can be resolved, because the total number of syllables is fixed to six (Johnson 1996:76). This means that the second half of the line is more restricted than the first. Another kind of Somali verse, *masafo*, constrains the number of morae per position in all but the very first position of the line, where 2, 3 or 4 morae can be realised (Banti & Giannattasio 1996:94).

Word length

A number of verse traditions restrict the length of the line-final word, leaving the rest of words unconstrained in terms of length. In the Irish [Indo-European; Celtic] *Ae freislighe* quatrains, odd lines end in trisyllables, and even lines in dissyllables (Knott 1957:13); Dyirbal [Pama-Nyungan] *Marrga* songs require line-final words to be bisyllabic (Dixon & Koch 1996:181); in Finnish [Uralic; Finnic] *Kalevala* verse, monosyllables are "not permitted at the end of a line" (Kiparsky 1968:138).

Phonemes

Rhyme is a very common feature in the poetry and singing of many languages. This involves limiting the choices of phonemes one can use at the end of lines; i.e. a relative free choice of phonemes in every position of the line, except at the end. Rhyme, hence, can be interpreted as a case of strict end.

Some languages, nonetheless, follow the opposite pattern: they restrict the phonemes occurring at the beginning of lines, and not elsewhere, by using line-initial alliteration. This type of strict beginning is attested throughout the Mongolic language family (Krueger 1961; Kara 2011), as well as in Welsh [Indo-European; Celtic] (Greenhill 2011). From a typological point of view, however, line-final rhyme is much more frequent than line-initial alliteration (Fabb 1999). This asymmetry further strengthens the SEH.

Words

Even more stringent are the verse templates where specific words or kind of words are required to close the line. In a Sardinian [Indo-European; Italic] *anninnia* from Bosa, each line in the song is closed by the formula *ninna ninna* (Sassu & Sole 1972:121); in the Melpa [Nuclear Trans New Guinea; Central East New Guinea Highlands] *kang rom* style of songs, lines are composed by regular lexical words, but one of a restricted set of vocables (i.e. meaningless words) is produced at the end of each line (Niles 2011:284). Further typological work is needed to assess whether these line-final formulae are more frequent than line-initial ones, as the latter are also attested in languages such as Kuna [Chibchan; Core Chibchan] (Sherzer 1982:373).

Melody

The majority of the world's verse is performed in sung form; this usually entails the use of stable pitch classes (i.e. in the form of melodies). Robust typological evidence lacking, there are indications that fixed patterns of melodic pitch classes are more frequent at the end of lines than elsewhere. In the Nambudiri tradition of Veda recitation (in Sanskrit), lines are closed by a conventional melodic cadence over the last vowel or nasal consonant (Staal 1961:50). Many verse systems show a similar pitch-cadence phenomenon, where a fixed (low) pitch closes each line (e.g. Huli [Nuclear Trans New Guinea; Enga-Kewa-Huli] songs, Pugh-Kitingan 1984:107). While these traditions display particularly invariant melodic material in the *last* few syllables, others display an asymmetry by singing the *first* few notes with an undetermined, speech-like pitch, while the rest of the line employs stable notes (e.g. Tedaga [Saharan; Western Saharan] songs, Brandily 1976:176).

Rhythm

A related line-final effect is the lengthening of the very last syllable of the line. This is observed e.g. in Warlpiri [Pama-Nyungan; Desert Nyungic], Somali or Sardinian (Turpin & Laughren 2013; Banti & Giannattasio 1996; Sassu & Sole 1972), and has probably a widespread typological distribution (Nettl 1956:66). One of the results of lengthening is that the temporal interval between the last syllable of a line and the first of the following line is increased with respect to the preceding inter-syllabic intervals. This same effect is also achieved by a simple pause, or by leaving an empty beat, as seen at the end of lines in many nursery rhymes, such as *Eeny, meeny, miny, moe*, where each strong syllable is followed by a weak one except at the end of lines, or *Hickory dickory*, where all lines except the third show a line-final empty beat. This kind of truncation is used preferentially at the end of lines, as it provides a structural pause (Hayes & MacEachern 1996; Hayes & Kaun 1996).

4.2.2 Template and instance strictness

All these phenomena can be interpreted as evidence for the SEH in one way or another. However, I want to argue that it is useful to be specific about the kind of features which are constrained, and to characterise the stringency of the constraints.

Some of the cited examples consist of completely stringent, fixed phenomena which mark the right edge of lines somehow, e.g. by requiring the presence of specific phonemes, words or pitches. We can refer to them as examples of *template strictness*.

The constraints related to syllable type and syllable-to-position association, on the other hand, usually exhibit varying degrees of stringency. We can call them examples of *instance strictness*. Requiring a specific rhyme to close the line is arguably a categorical feature encoded in the verse template; the increasing consistency in using light or heavy syllables, though, is a gradual effect observed when a collection of instances of the same template are analysed.

This binary classification of final strictness phenomena can already be helpful in order to better understand their possible cause or function. Nevertheless, quantifying the stringency of any kind of strictness is still crucial if one wants to verify whether it is categorically localised at one of the edges, whether it is gradual, and, if gradual, the extent to which the different positions in the line are restricted. The next section develops a case study where the instance strictness of syllable type is quantified and compared across verse samples in five different languages. This is the prototypical case of final strictness, and, given its gradual nature, it constitutes a suitable object for quantitative examination.

4.3 Measuring strictness

4.3.1 Materials

We analyse data from five languages: three Indo-European languages from two different branches (English and Dutch from the Germanic branch, and Sanskrit from the Indo-Iranian branch), the Uralic language Estonian, and the Afro-Asiatic language Tashlhiyt Berber. To be sure, the sample is not broad enough to make strong claims about the universality of final strictness; however, I describe a methodology which can be easily extended to include further data in future studies. The choice of languages attempts to maximise the typological coverage, while being constrained by the availability of sizeable digitized corpora.

Table 4.1: Summary of corpora used in the analyses.

Language	ISO	Lines	Samples	Family	Branch
English	eng	4198	2	Indo-European	Germanic
Estonian	est	8811	20	Uralic	Finnic
Dutch	nld	9079	2	Indo-European	Germanic
Sanskrit	san	37908	3	Indo-European	Indo-Iranian
Tashlhiyt Berber	shi	314	7	Afro-Asiatic	Berber

The kind of strictness being measured here is the one about restrictions on syllable type (cf. Section 4.2.1). Hence, any verse sample where at least certain positions require a particular syllabic feature (e.g. weight or stress) are suitable for the analysis. Table 4.1 lists the five languages used in the analysis, together with the total number of lines, and the number of samples (e.g. authors) for each language.

The Sanskrit sample includes lines from the *Ṛgveda*, a text composed in the second millennium BC, probably earlier than 1200 BC (Witzel 1995). The templates used are quantitative, i.e. they contrast heavy and light syllables. For the current analysis I have employed the summary statistics provided by Gunkel & Ryan (2011), who list the proportion of heavy syllables for each position in metres of eight, eleven and twelve syllables. All three metres follow a general iambic pattern. However, that the verse lines from the *Ṛgveda* show final strictness has been known for a long time: "in all metres the rhythm of the latter part of the verse is much more rigidly defined than that of the earlier part" (Arnold 1905:9). The present analysis involves almost thirty-eight thousand lines of verse.

The Dutch sample includes 9079 lines by two 20th-century poets: J. P. Kal (b. 1946) and C. O. Jellema (1936–2003). Most of the lines belong to 14-line-long sonnets and follow an iambic pentametre template. For purposes of syllable-position identification, lines longer or shorter than ten syllables have been excluded from the original corpus. Each line has been automatically scanned using a scansion algorithm (van Oostendorp 2014) which takes into account the syllable's lexical stress and its environment, and yields a binary result for each syllable: metrically stressed (1) or unstressed (0). The iambic pentametre template predicts that odd positions will contain unstressed values, and even positions stressed values.

The English sample contains 4198 lines by John Milton (1608–1674) and William Shakespeare (1564–1616). Our analysis is based on the digital text annotated by Bruce Hayes, which assigns one of four stress levels to each syllable: 0 = unstressed, 1 = secondary stress, 2 = primary, 3 = phrasal (Hayes, Wilson & Shisko 2012). Given that most accounts of English verse only distinguish a binary opposition of stress, I have collapsed values 1, 2 and 3 into a single category of stress, opposed to unstressed (0). The lines used for the analysis all follow an iambic pentametre template. The sample by Shakespeare is derived from his 154 sonnets, excluding sonnet 145 which is composed in iambic tetrametre, and excluding lines longer than 10 syllables (i.e. ending in feminine rhyme). The sample by Milton is drawn from books 9 and 10 of his work *Paradise Lost*, also applying the filter to retain only 10-syllable-long lines.

The Estonian sample summarises 8811 lines composed by 20 different authors from the late 19th, early 20th centuries. The data are taken from statistics provided by Lotman & Lotman (2013), where each syllable is assigned a stress value ranging from 0 (unstressed) to 4 (phrasal stress). As with the English sample, only a binary distinction between stressed and unstressed has been retained. All lines follow a trochaic tetrametre template, where odd positions generally contain stressed syllables, and even positions unstressed.

The Tashlhiyt sample contains 314 lines of verse belonging to seven different songs. Each song follows a different template, but all of them are quantitative; hence, positions are expected to contain either a heavy or a light syllable (as in the Sanskrit corpus). The song texts, their scansion and thorough analyses have been published by Dell & Elmedlaoui (2008).

In all five corpora we observe templates where two classes of syllables are used in a controlled way. Still, the nature of these two classes depends on the phonological features of each language, and the interpretation of which syllables constitute deviations from the template depends on the method of analysis used for each corpus. In the Estonian corpus, the alternating syllable classes are based on word stress, which is always word-initial in this language (Harms 2017). The syllable classes in the English and Dutch corpora are also based on word stress, but this feature plays a more important role than in Estonian, since its placement within the word is not fully predictable (Hulst 1984). On the other hand, the Sanskrit and Tashlhiyt syllable classes are not based on stress but on weight, where syllables ending in a coda (and/or in a long vowel for Sanskrit) are considered heavy (Dell & Elmedlaoui 2002; Arnold 1905).

In terms of the kinds of methods used to detect deviant syllables, the Dutch corpus differs from the other four in that for each syllable the scansion algorithm takes into account the neighbouring syllables and the ideal metrical template to determine its prominence value. In the other corpora, whether a syllable is considered weak or strong does not depend on the metrical context, but is determined exclusively on linguistic grounds.

4.3.2 Statistical analyses

For each syllabic position in the corpora as coded here, a binary feature (0, 1) indicates the prominence value (related to stress or weight) for that syllable. It is assumed all lines in a sample follow the same template, which regulates the placement of stress or weight. In order to measure the consistency of this regulation, I compute the entropy for each position of the template (Shannon 1948). This produces a measure of how certain we can be about the prominence value of a syllable given it occurs in a particular position; values lie between 0 (complete consistency), and 1 (highest uncertainty).

This measure does not capture higher order dependencies between positions, such that an unstressed syllable in the second position of an iambic pentametre is likely preceded by a stressed syllable (i.e. trochaic inversion). Nevertheless, it does allow us to analyse the relative unigram consistency related to syllable type regulation across line positions, and across corpora.

In order to assess whether the relative syllabic position (*relpos*) predicts a decrease in entropy (i.e. less heterogeneity towards the end of the line), I fit a mixed effects model to the whole set of samples combined. The different language corpora are added as random factors (see the complete model under Equation 4.1 as called in `R` using the `lme4` package, Bates et al. 2015). In order to assess with more detail the robustness of the SEH for each individual language, I perform separate linear regressions for each language corpus.

$$entropy \sim relpos + (1 + relpos + prominence | language) \tag{4.1}$$

4.3.3 Results

Figure 4.1 plots the entropy values for each syllabic position within each sample, grouped by languages. Separate regression lines are drawn for positions treated as strong and weak in their respective templates.

We can observe a general downtrend in entropy for all languages, suggesting that the initial hypothesis that there is an increase in consistency in the use of weight and stress is correct. In the Estonian sample, the trend for strong positions is completely flat, with no apparent increasing strictness. In fact, deviations from the ideal template (stressed syllables on odd positions) are close to null. The Sanskrit sample also shows a flatter and lower regression line for strong positions. However, this is not the case for the other three languages.

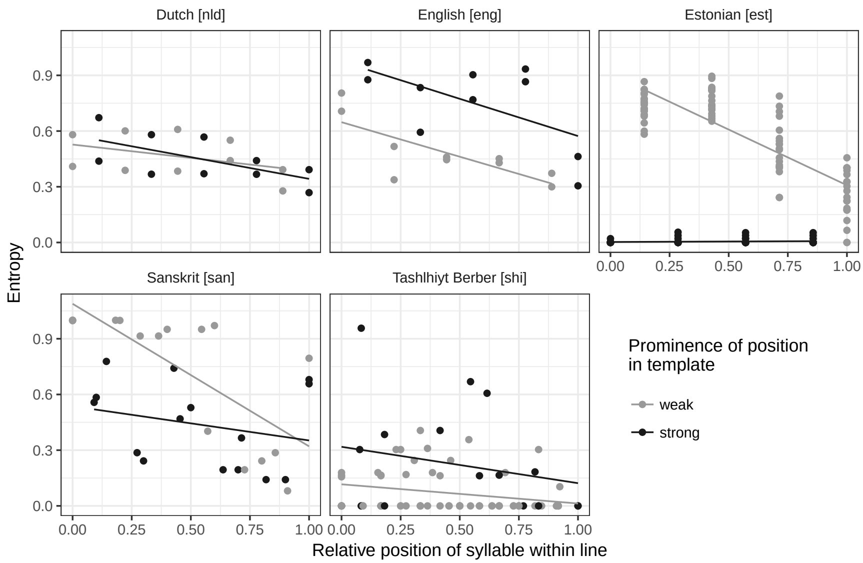


Figure 4.1: Entropy of syllabic feature values (stressed vs unstressed; heavy vs light) for each position in each sample of verse.

Visual inspection reveals that the final position of Sanskrit verse breaks with the increasing strictness tendency, a phenomenon well described in the literature (Arnold 1905). Furthermore, a relative increase in entropy towards the middle of the line (e.g. English and Sanskrit) suggests the presence of division of the line in two half-lines.

Table 4.2 shows the results of the mixed effects model fitted to the data. Relative position within the line proves to be a good predictor of entropy, with a strong negative estimate indicating a decrease in uncertainty as the syllable position increases. A comparison of this model with the corresponding null model without the fixed predictor indicates that the prediction of entropy significantly improves by adding relative syllable position as a predictor (χ^2 = 9.63, Pr ($> \chi^2$) = 0.0019).

Figure 4.2 shows how these estimates are to be adjusted for each of the subcorpora; that is, it visualizes the random part of the mixed model under Equation 4.1. The vertical dashed line represents the average, baseline results; red/blue figures indicate that a language shows a lower/higher value with respect to the baseline.

In the leftmost panel (Intercept) we observe that Sanskrit shows the highest overall entropy, and Tashlhiyt Berber the lowest, confirming the plots in Fig-

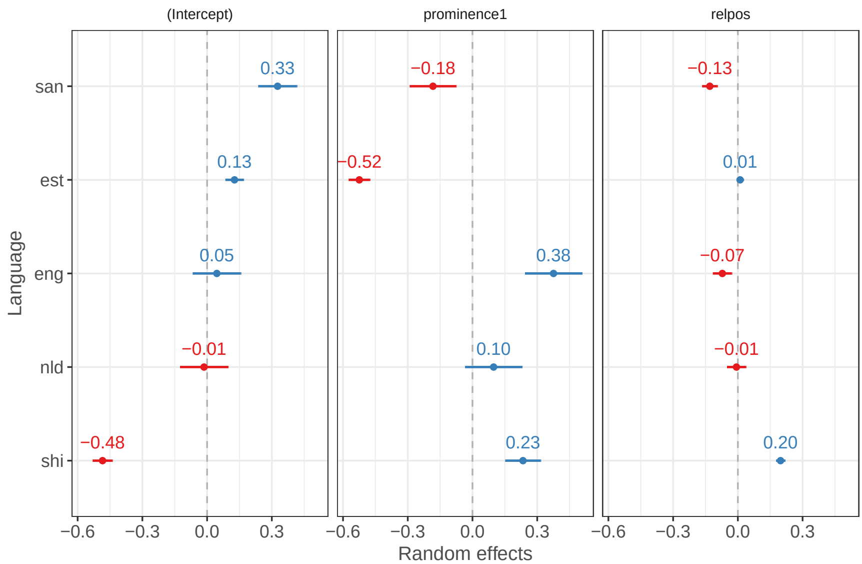


Figure 4.2: Random effects for the five different languages in the sample. The plot indicates the amount by which the model estimates are adjusted according to the language of the data under analysis. Negative adjustments with respect to the baseline are plotted in red.

ure 4.1. In the second row of Table 4.2 we had observed that later syllables have a negative estimate, i.e. they are predicted to have a lower entropy. The rightmost panel (relpos) reveals that Sanskrit is the language where this decreasing-looseness effect is strongest, i.e. it shows the steepest slope from high entropy at the beginning of the line, to low entropy at the end. Finally, the central panel (prominence1) displays the varying effect of prominence on entropy. Here, we corroborate that strong positions in Estonian (and in Sanskrit to a lower extent) are much stricter than weak positions (i.e. are expected to show a lower entropy), whereas the opposite effect is found in English.

Individual linear regressions per language sample (Table 4.3 of the Supplementary Information) confirm that the robust overall decrease in entropy also holds for each language corpus. The prominence value of the position (whether weak or strong) further improves the model for all languages except Dutch, where weak and strong positions follow a similar downtrend.

Table 4.2: Results of the full mixed model, with relative syllable position as the fixed predictor, and random slope and intercept for the effect of prominence and syllable position in each language.

	Term	Estimate	Std. Error	t value	Pr ($> t$)
1	(Intercept)	0.575	0.117	4.91	0.00804
2	relpos	-0.305	0.0616	-4.95	0.00995

4.4 Discussion

Binary restrictions on the type of syllable used at certain positions of the line are common to many verse traditions. As proposed by e.g. Jakobson (1966), metrical rules are often regarded as binary; nevertheless, Ryan (2011) argues that this binarity appears to be gradient. Moreover, the direction of the gradience can be predicted to a certain extent using the relative position of the syllable. In the samples of verse here analysed, there is a robust trend for the binary restriction on weight or stress to be more stringent later in the line. This supports a specific type of the SEH, namely, if a verse tradition regulates a syllabic feature, it will be more consistent the later the syllable occurs in the line.

Visual inspection of the Estonian corpus suggested a second predictor of entropy: prominence. Even though four out of five corpora show distinct entropy profiles for weak and strong positions, there is no overall prediction: in some cases strong positions show higher entropy (English, Tashlhiyt), and in others, lower entropy (Estonian, Sanskrit). A note of caution is needed when using position prominence as a predictor of entropy. At least two potential confounds can drive the direction of the effect of prominence on entropy: (1) the structure of the lexicon, (2) the coding algorithm.

In a language with higher proportion of light syllables, consistently using light syllables in weak positions proves easier, hence strong positions are predicted to have a higher entropy (i.e. it is more likely that they contain light syllables, than the other way round). On the other hand, as explained in the description of the materials, the algorithms used to decide the weight/stress feature of a syllable differ, e.g. in Estonian, even the lightest degree of stress is coded as stressed; in Dutch, the neighbouring syllables are considered in order to maximise the fit to the ideal template. The facts expressed in those two caveats hinder a direct interpretation on the predictive effect of position prominence on entropy. Further research is needed which takes lexico-statistical data from the feature of interest as a predictor of the effect of prominence.

If a robust typological tendency is established, we can hypothesise that it derives from features shared by the whole population under study, like some aspect of basic human cognition. In Section 4.2 I propose that the diverse phenomena under the general SEH can be subdivided into two main types: template strictness, and instance strictness. The contrast between the categorical nature of the first type, and the gradual nature of the second suggests different causes. From a general perspective, I propose that categorical asymmetries encoded in the templates (e.g. rhyme, cadence) may have a facilitating *function*, and that the gradual asymmetries (e.g. selection of syllable type) are a *result* of some cognitive bias.

A plausible, low-level cognitive bias is the gradual increase of attention as new temporal stimuli are processed. Each verse template diplays a regular alternation of features, but the regularity gets a stronger representation as the line develops and more syllables satisfy the template, as proposed by the Bayesian predictive coding framework (Vuust & Witek 2014). Alternatively, if the creator or recipient of the verse lines entrains to some regular temporal sequence (e.g. of syllables or beats), the dynamic attending theory (Jones et al. 2002) predicts that clusters of neurons will synchronise with that regularity. The synchronisation strengthens as more stimuli are processed, and, if attention peaks correlate with neuronal firing, one is expected to be extra sensitive by the end of the line.[3]

Hypotheses on increase of attention can be suitable for gradual phenomena, and particularly for loose beginnings, where full-entrainment has not yet taken place. Nevertheless, it does not fit well with categorically final phenomena, such as rhyme. A straightforward interpretation of these is that they work as boundary-markers, making the constituent structure of verse easier to parse. However, an equivalent explanation is available for left-boundary markers. Hence, a number of functional accounts specific to template final strictness can be put forward, but they all remain tentative in the lack of robust empirical work, probably experimental.

A feature common to all the categorical asymmetries I have discussed in Section 4.2 is that they limit the number of choices at the very end of the line, by restricting the choice of phonemes, the word-length, or requiring a specific closing-word. This effectively reduces the cognitive load of the performer. Still, this cognitive advantage at the end of lines can be seen as serving various alternative *functions*.

[3] These arguments could also be applied to larger chunks of verse, such as couplets or stanzas. However, there is good evidence that lines are treated as whole units in working memory (Fabb 2014), making it a suitable candidate for the unit of attention increase.

The first is proposed, for instance, by Niles (2011) when discussing the fixed vocables closing lines of *kang rom* Melpa songs: "because of their regularity in delimiting a line of text, these vocables perhaps also allow the performer a brief rest and chance to mentally compose his or her thoughts for the next line." This explanation may be particularly relevant for improvised traditions, where the following line needs to be composed while still singing the current one; yet, it also applies to non-improvised performances, where saving cognitive load can facilitate the recall of the next line (Rubin 1995). Similar functions have been attributed to the pervasive final lengthening observed in everyday speech (Fletcher 2010).

A related advantage of marking right boundaries with predictable material is that it enables a smoother turn-taking. The previous hypothesis worked best in the context of solo performance; however, there are traditions where two poets engage in dialogue-like exchange of lines (Egaña 2007). In this context, one needs to compose a line *while* the other is singing. Hence, predicting the end of a line gives the poet some advantage in order to plan the next line and execute it without delay. Again, this has a parallel in everyday speech, where it has been shown that the gap between turns is so short, that speakers must plan in advance and accurately predict the end of the interlocutor's utterance (Stivers et al. 2009).

An alternative proposal by Fabb (2014) is that reduced cognitive load at the end of a line facilitates the recall of earlier elements within that same line. The hypothesis relies on the idea that lines are processed as single units within working memory. The final part of the line already has a recency-effect advantage, and having a reduced cognitive demand at the end would leave more room for keeping in mind earlier linguistic content. Earlier in this section, I have argued that final strictness may ease the planning of the *following* line, while Fabb proposes that final strictness eases the remembering of the *current* line.

Given the diversity of final strictness phenomena, it is unlikely that they all have the same cause or function, so apparently contradictory hypotheses may in fact prove complementary. In this specific case, Fabb's proposal covers better the function of final strictness during verse perception, while the planning hypothesis applies to the process of verse production, as exemplified by the quote on Melpa singing (Niles 2011). In order to test the coverage of these hypotheses, behavioural experiments can be conducted which manipulate line-final elements and compare recall and reaction times. Moreover, the predominant ecological context where verse is created and consumed in a given tradition may also require dissimilar cognitive explanations; in particular, the visual aid available when composing texts in written form, such as the Dutch or Estonian samples here analysed, would need to be taken into account.

Finally, the pervasive presence of predictable material at the end of lines can have aesthetic reasons. As argued by Huron (2006), a source of pleasure in music lies in the fulfilment of expectations. Moreover, added aesthetic value may be produced when expectations are violated or kept on hold for some time, and satisfying them. Although gathering empirical evidence of aesthetic value poses methodological challenges, physiological measures such as skin conductance response (Mas-Herrero et al. 2014) can potentially be used in order to test whether line-final fixed elements are a more significant source of aesthetic pleasure compared to predictable material on other locations in the line.

4.5 Conclusion

The Strict End Hypothesis has been discussed in the field of metrics for more than a century. Furthermore, it has been posited that it may be universal in nature. In this chapter I address two challenges which precede the typological verification of the claim. First, I argue that a range of diverse phenomena are categorised within the SEH, making the hypothesis effectively intractable. At the very least, two types of strictness should be distinguished: template and instance strictness. Second, in order to measure the extent and degree of each type of strictness, quantification or other kinds of fine-grained description are necessary. I showcase this by characterising gradual strictness in a dataset including verse from five languages, using entropy as a proxy for strictness. Finally, I discuss a number of testable cognitive explanations of the phenomena, which move towards the goal of understanding the *why* of final strictness.

References

Arnold, E.-V. 1905. *Vedic metre in its historical development*. Cambridge University Press.

Banti, G & F Giannattasio. 1996. Music and metre in Somali poetry. In R.-J. Hayward & I.-M. Lewis (eds.), *Voice and power: essays in honour of B. W. Andrzejewski*, 83–127. School of Oriental & African Studies.

Bates, D., M. Mächler, B. Bolker & S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Bickel, B. 2007. Typology in the 21st century: major current developments. *Linguistic Typology* 11(1). 239–251.

Brandily, M. 1976. Un chant du Tibesti (Tchad). *Journal des africanistes* 46(1). 127–192.

de Groot, A. W. 1936. De structuur van het vers. *De Nieuwe Taalgids* 30. 197–212.

Dell, F & M Elmedlaoui. 2008. *Poetic meter and musical form in Tashlhiyt Berber songs*. Köln: Rüdiger Köppe.

Dell, F. & M. Elmedlaoui. 2002. *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht: Kluwer Academic Publishers.

Dixon, R.-M.-W. & G Koch. 1996. *Dyirbal song poetry. The oral literature of an Australian rainforest people*. University of Queensland Press.

Egaña, A. 2007. The process of creating improvised bertsos. *Oral Tradition* 22(2). 117–142.

Fabb, N. 1999. Verse constituency and the locality of alliteration. *Lingua* 108. 223–245.

Fabb, N. 2002. *Language and literary structure: the linguistic analysis of form in verse and narrative*. Cambridge University Press.

Fabb, N. 2014. The verse-line as a whole unit in working memory, ease of processing, and the aesthetic effects of form. *Royal Institute of Philosophy Supplement* 75. 29–50.

Fabb, N. 2015. *What is poetry? Language and memory in the poems of the world*. Cambridge University Press.

Fletcher, J. 2010. The prosody of speech: timing and rhythm. In W.-J. Hardcastle, J Laver & F.-E. Gibbon (eds.), *The handbook of phonetic sciences*, 523–602. Blackwell Publishers.

Greenhill, P. 2011. Bardic rhythm: the implications from *Cerdd Dant* studies. *Studia Celtica* 45. 131–153.

Gunkel, D & K Ryan. 2011. Hiatus avoidance and metrification in the *Rigveda*. In S.-W. Jamison, H.-C. Melchert & B Vine (eds.), *Proceedings of the 22nd annual ucla indo-european conference*, 53–68. Hempen.

Harms, R. T. 2017. *Estonian grammar*. Routledge.

Hayes, B. 1983. A grid-based theory of English meter. *Linguistic Inquiry* 14(3). 357–394.

Hayes, B. & A. Kaun. 1996. The role of phonological phrasing in sung and chanted verse. *The Linguistic Review* 13. 243–303.

Hayes, B. & M. MacEachern. 1996. Are there lines in folk poetry? *UCLA Working Papers in Phonology* 1. 125–142.

Hayes, B., C. Wilson & A. Shisko. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88(4). 691–731.

Hulst, H. Van der. 1984. *Syllable structure and stress in Dutch*. Foris.

Huron, D. 2006. *Sweet anticipation: music and the psychology of expectation*. Massachusetts Institute of Technology.

Jakobson, R. 1966. *Selected writings IV. Slavic epic studies*. Mouton de Gruyter.

Johnson, J.-W. 1996. Musico-moro-syllabic relationships in the scansion of Somali oral poetry. In R.-J. Hayward & I.-M. Lewis (eds.), *Voice and power: essays in honour of B. W. Andrzejewski*, 73–82. School of Oriental & African Studies.

Jones, M. R., H. Moynihan, N. MacKenzie & J. Puente. 2002. Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science* 13(4). 313–319.

Kara, G. 2011. Alliteration in Mongol poetry. In J. Roper (ed.), *Alliteration in culture*, 156–179. London: Palgrave Macmillan UK.

Kiparsky, P. 1968. Metrics and morphophonemics in the *kalevala*. In C Gribble (ed.), *Studies presented to roman jakobson by his students*, 137–148. Cambridge (MA, USA): Slavica.

Knott, E. 1957. *An introduction to Irish syllabic poetry of the period 1200–1600*. Dublin Institute for Advanced Studies.

Krueger, J. R. 1961. *Poetical passages in the* Erdeni-yin Tobči*: a Mongolian chronicle of the year 1662 by Sayang Sečen*. 's-Gravenhage: Mouton.

Lotman, M.-K. & M Lotman. 2013. The quantitative structure of Estonian syllabic-accentual trochaic tetrameter. *Trames* 17(3). 243–272.

Mas-Herrero, E, R.-J. Zatorre, A Rodríguez-Fornells & J Marco-Pallarés. 2014. Dissociation between musical and monetary reward responses in specific musical anhedonia. *Current Biology* 24. 1–6.

Nettl, B. 1956. *Music in primitive culture*. Harvard University Press.

Niles, D. 2011. Metric melodies and the performance of sung tales in the Hagen area. In A Rumsey & D Niles (eds.), *Sung tales from the papua new guinea highlands. studies in form, meaning, and sociocultural context*, 275–201. The Australian National University.

Nordhoff, S. & H. Hammarström. 2011. Glottolog/Langdoc: Defining dialects, languages, and language families as collections of resources. In *Proceedings of the first international workshop on linked science 2011*, vol. 783 (CEUR Workshop Proceedings).

Prince, A. 1989. Metrical forms. In P Kiparsky & G Youmans (eds.), *Rhythm and meter*, 45–80. Academic Press.

Pugh-Kitingan, J. 1984. Speech-tone realisation in Huli music. In J. C. Kassler & J. Stubington (eds.), *Problems & Solutions: Occasional Essays in Musicology presented to Alice M. Moyle*, 95–120. Sydney: Hale & Iremonger.

Rubin, D.-C. 1995. *Memory in oral traditions: the cognitive psychology of epic, ballads, and counting-out rhymes.* Oxford University Press.

Ryan, K. M. 2011. Gradient syllable weight and weight universals in quantitative metrics. *Phonology* 28(03). 413–454.

Sassu, P. & L. Sole. 1972. Funzione degli stereotipi nel canto popolare sardo. *Rivista Italiana di Musicologia* 7(1). 115–144.

Shannon, C. E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27(3). 379–423.

Sherzer, J. 1982. Poetic structuring of Kuna discourse: the line. *Language in Society* 11. 371–390.

Staal, J.-F. 1961. *Nambudiri Veda recitation.* Mouton & Co.

Stivers, T., N. J. Enfield, P. Brown, C. Englert, M. Hayashi, T. Heinemann, G. Hoymann, F. Rossano, J. P. de Ruiter, K.-E. Yoon & S. C. Levinson. 2009. Universals and cultural variation in turn-taking in conversation. eng. *PNAS* 106(26). 10587–10592.

Turpin, M & M Laughren. 2013. Edge effects in Warlpiri *yawulyu* songs: resyllabification, epenthesis, final vowel modification. *Australian Journal of Linguistics* 33(4). 399–425.

van Oostendorp, M. 2014. Introducing a scansion machine for Dutch poetry and prose. *Loquens* 1(1). 1–7.

Vuust, P. & M. A. Witek. 2014. Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Frontiers in psychology* 5.

Witzel, M. 1995. Early Sanskritization. Origins and development of the Kuru state. *Electronic Journal of Vedic Studies (EJVS)* 1(4). 1–26.

Zwicky, A.-M. & A.-D. Zwicky. 1986. Patterns first, exceptions later. In R Channon & L Shockey (eds.), *In honor of Ilse Lehiste*, 525–537. Foris.

Supplementary Information

Table 4.3: Results of the linear model applied to each language corpus, with position prominence and relative syllable position as predictors.

	Language	Term	Estimate	Std. Error	t value	Pr $(> t)$
1	eng	(Intercept)	0.654	0.0668	9.79	2.1e-08
2	eng	relpos	-0.386	0.106	-3.64	0.00204
3	eng	prominence1	0.312	0.0678	4.6	0.000258
4	est	(Intercept)	0.735	0.0248	29.6	3.24e-66
5	est	relpos	-0.297	0.034	-8.73	3.5e-15
6	est	prominence1	-0.603	0.0223	-27.1	5.31e-61
7	nld	(Intercept)	0.547	0.0464	11.8	1.29e-09
8	nld	relpos	-0.189	0.0737	-2.56	0.0204
9	nld	prominence1	0.00408	0.0471	0.0866	0.932
10	san	(Intercept)	0.966	0.0915	10.6	2.91e-11
11	san	relpos	-0.504	0.143	-3.53	0.00147
12	san	prominence1	-0.257	0.091	-2.83	0.00859
13	shi	(Intercept)	0.127	0.034	3.73	0.000351
14	shi	relpos	-0.125	0.0552	-2.26	0.0264
15	shi	prominence1	0.158	0.0431	3.65	0.000459

5 The detection of deviants in pseudo-verse lines

5.1 Introduction

Verse encompasses a variety of forms, such as song, poetry, chant or nursery rhymes. All of these contain words, but they also include a feature which is absent from everyday speech, viz. in verse, words are set to templates. These constrain the verbal material one wants to use in several ways. Often, for instance, the length of lines in poems or songs is limited by a fixed number of syllables or beats. In many languages, the relative prominence of the syllables is also constrained, so that e.g. when creating an English sonnet in iambic pentameter, lines are usually opened with an unstressed syllable, followed by a stressed one.

Notwithstanding these structural constraints, verse is not completely rigid, as poets and singers often deviate from the templates, introducing unexpected elements which can be exploited for aesthetic purposes by generating interest or surprise in the listener (Huron 2006). Studies of verse corpora show that, still, it is possible to generalise where deviations tend to occur: they are most frequent at the beginning of lines, and their incidence (progressively) decreases (Fabb 2002: 173–177). This phenomenon is referred to by terms such as *final strictness* or *initial looseness*, bringing out the fact that the asymmetry can stem from exceptional events at either edge of the line.

Despite the lack of a systematic typological survey, robust final strictness phenomena are reported for languages from unrelated families, such as Sanskrit (Arnold 1905), Finnish (Kiparsky 1968), Berber (Dell & Elmedlaoui 2008), and Greek (Allen 1973; Golston & Riad 2000). Chapter 4 provides further details on final strictness, as well as an overview of other phenomena usually considered examples of final strictness, such as rhyme or melodic cadence. Hence, there is some evidence that final strictness is not a property linked to a limited set of related languages which accidentally developed the tendency. Instead, the range of independent observations of final strictness, and the lack of a robust set of languages showing the opposite pattern (i.e. initial strictness) asks for a common explanation. There is the possibility, for instance, that the effect is driven by some aspect of cognition shared across populations. In the present chapter we explore one hypothesis within this context, namely, that the decrease in the frequency of deviations is due to an increase of attention along the line, which is disrupted between lines.

Previous studies show that if the occurrence of an event can be predicted, its processing is facilitated (Jongsma, Desain & Honing 2004; Niemi & Näätänen 1981). The internal regularities characteristic of verse lines allow for prediction building to take place. Nevertheless, this process may be disrupted by line boundaries, which are a defining feature of verse (Fabb 2015). Other constituent levels such as the stanza or the hemistich (see Chapter 2) may also show comparable disruptions, but we focus on the line because it is the constituent for which (1) final strictness is most often described, (2) universality has been argued.

In the present chapter we use sequences of drum strokes as a model of verse lines. Subjects are asked to detect deviations from a pattern under three different experimental conditions. We manipulate the relative timing of the strokes in order to test the extent to which the regularity of the stimuli and the pause between the sequences is driving the variation in reaction times. Overall, the results show that it takes longer to detect deviations closer to the beginning of the line, mirroring the data from verse corpora.

5.2 Method

5.2.1 Participants

A total of 45 subjects took part in the experiment (mean age = 23.1 year; 26 males, 19 females; all native Dutch speakers). Each participant was randomly assigned to one of the three conditions, reaching a total of 15 subjects per condition. The recruitment was done at Leiden University and Radboud University (The Netherlands). All participants signed an informed consent before performing the task (in accordance with Leiden University's LUCL procedure).

5.2.2 Procedure

The general procedure of the experiment was the same for all three conditions, i.e. subjects performed an auditory odd-ball experiment. Each participant listened to a total of 576 drum strokes; these could be of two kinds: (1) a probe stroke (n

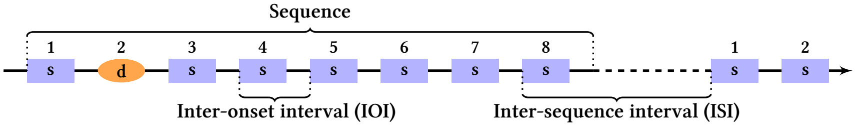


Figure 5.1: Temporal presentation of the stimuli, with time running from left to right following the arrow. The s symbol represents a standard drum stroke, and the d symbol represents a deviant stroke, i.e. the probe.

= 48), or a standard stroke (n = 528). Both sounds are publicly available studio recordings of a mridangam drum (Anantapadmanabhan, Bellur & Murthy 2013), comparable in frequency and intensity, but with differing timbre.[1]

Participants were instructed to press a key as soon as they detected a probe stroke. The general temporal configuration of the 576 strokes was similar across conditions: the strokes are played sequentially, with a short silent gap after every stroke, and a longer gap after every eighth stroke. Figure 5.1 depicts the temporal presentation of the stimuli, with time running from left to right following the arrow.

As seen in Figure 5.1, we refer to the group of eight strokes separated by a longer gap as a *sequence*. The duration from the beginning of a stroke to the beginning of the following stroke is called the *inter-onset interval* (IOI). The longer gap between sequences is called the *inter-sequence interval* (ISI). Each participant listened to a total of 72 sequences, two thirds of which (n = 48) contained a probe, and the remaining third (n = 24) served as fillers with no probe. None of the sequences contained more than one probe. The key measure taken during the experiment is the reaction time to detect the probe, i.e. the lapse of time between the onset of the probe and the subject pressing the key.

All three conditions contain the same number of sequences and probes, but they differ in their temporal presentation, as summarised in Table 5.1. The IOI is kept constant in conditions 1 and 2, i.e. strokes within sequences are isochronous. In condition 3, the IOI varies randomly and can take any value between 250 and 500 milliseconds. The difference between conditions 1 and 2 lies in the ISI, which is kept constant for condition 1, but varies in condition 2 between 1200 and 1800 milliseconds.

[1] Specifically, the standard sound is the *ta* stroke with identifier 224350, and the probe tone is the *num* stroke with identifier 224279.

Table 5.1: Summary of the parametres which define the three experimental conditions.

	Inter-onset interval (IOI)	Inter-sequence interval (ISI)
Condition 1	300 ms	1500 ms
Condition 2	300 ms	1200 $\sim$ 1800 ms
Condition 3	250 $\sim$ 500 ms	1500 ms

5.2.3 Statistical analyses

The main test we perform assesses whether probes occurring earlier within a sequence were detected more slowly than later probes. This is based on the observation that deviant syllables are more likely to occur earlier in verse lines, as formulated by the strict end hypothesis (Chapter 4). Thus, we build a mixed effects model with reaction time as the dependent variable, probe position as a fixed effect, and subject as a random effect. Subsequently, we run more complex models adding the experimental conditions as fixed effects, and controlling for potential confounds.

All the mixed models are implemented in `R` (R Core Team 2017) using the statistical package `lme4` (Bates et al. 2015). Significance of the predictors is calculated in two ways. First, we conduct maximum likelihood t-tests using Satterthwaite approximations to degrees of freedom, as implemented in the package `lmerTest` (Kuznetsova, Bruun Brockhoff & Haubo Bojesen Christensen 2016). Second, we build a null model, identical to the full model except that the variable of interest has been excluded. The fit of the model to the data is compared through a likelihood ratio test to determine whether the full model bears greater explanatory power, hence showing support for the predictor under consideration (Roberts, Winters & Chen 2015).

5.3 Results

Visual inspection of the reaction times to the probe plotted against the probe position within the line (Figure 5.2) reveals a strong negative correlation: probes occurring later in the line require less time to be detected.

However, there is an important confound to control for. Given the design of the experiment (maximally one probe per sequence), probes occurring earlier in the line can have a preceding probe closer by (i.e. if the previous sequence contains

Table 5.2: Summary of the fixed effects in the mixed model analysis.

Predictor	Estimate	Std. Error	df	t	Pr $(> t)$
(Intercept)	0.271	0.0925	51.85	2.93	0.00496
probe.dist	-0.0277	0.00534	1949.06	-5.18	2.39e-07
probe	-0.0346	0.0171	61.09	-2.02	0.0476
condition2	0.105	0.127	46.08	0.83	0.411
condition3	0.362	0.127	45.69	2.86	0.00637
probe:condition2	-0.0236	0.0228	48.29	-1.04	0.306
probe:condition3	-0.0791	0.0227	47.58	-3.48	0.00109

a late probe). Let *probe distance* be the number of strokes between a probe and its preceding probe. On the one hand, probes in position 1 have a mean probe distance of 4.5, while the mean probe distance in position 8 is 12.1 ($r = .64$). On the other hand, probe distance negatively correlates with reaction time: the longer the probe distance, the shorter it takes to react to a probe ($r = -.3$).

Our full model (see Equation 5.1) includes probe distance as a predictor of reaction time, plus an interaction between probe position and experimental condition. A random slope and intercept for the effect of probe position per subject is added. The results of the model are summarised in Table 5.2. Condition 1 is taken as a baseline with which the other two conditions are compared. It can be observed that probe position remains a robust predictor of reaction time after controlling for the confounding effect of probe distance. Besides, the effect of probe position is significantly greater under condition 3.

$$RT \sim probe.dist + probe.pos * condition + (1 + probe.pos | subjectID) \quad (5.1)$$

The statistical significance of the fixed effects on reaction time is confirmed by the comparison of the full model with equivalent null models where the predictor of interest has been removed. The addition of probe position improves the model ($\chi^2 = 38.5$, $p = 2.2e-08$), the addition of the experimental condition does so too ($\chi^2 = 13.4$, $p = 0.0096$), and the model with an interaction between probe position and condition provides a better fit too ($\chi^2 = 11.3$, $p = 0.0035$).

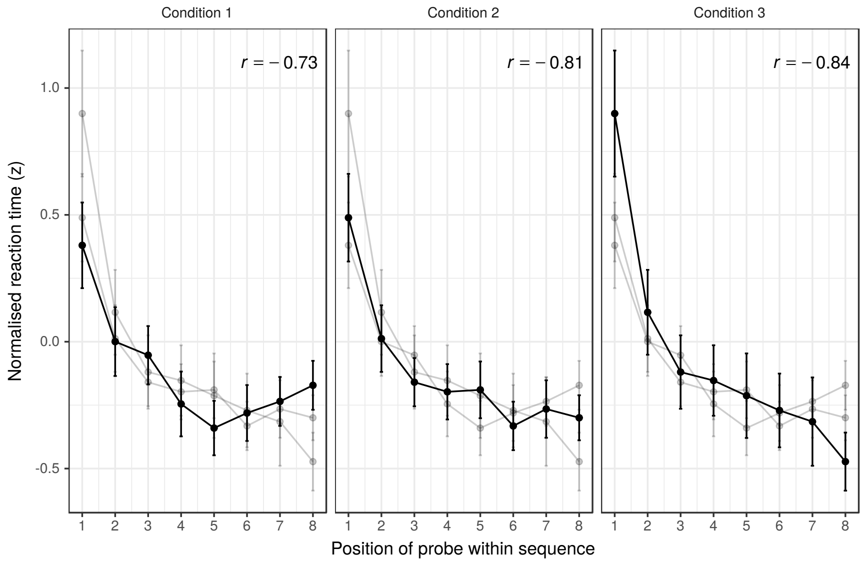


Figure 5.2: Mean reaction time to detect the probe in each of the eight possible positions within the sequence. Pearson's correlation coefficient added separately for each of the conditions. Error bars indicate the 95% confidence interval. In each panel, the condition of interest is highlighted, and the other two conditions are greyed out as a visual reference.

5.4 Discussion

Overall, our experimental results correlate with the strict end hypothesis posited for verse: (1) later in the verse line, deviations are less frequent, and (2) later in the experimental sequence, deviations are detected faster. Nevertheless, the three conditions under inspection do not pinpoint the defining rhythmic context under which the decreasing reaction time takes place.

Unlike similar odd-ball experiments (Schwartze, Farrugia & Kotz 2013; Bouwer & Honing 2015), our stimuli were organised into sequences separated by a longer silent gap. To be sure, the stimuli used in Bouwer & Honing (2015) have a recurring metrical structure which does evoke eight-beat sequences. Nevertheless, and despite the differences in design, we can hypothesise that the crucial difference which produces the decreasing reaction times lies in the sequence-dividing silent gaps.

That being said, we do find some noteworthy differences between the three experimental conditions. The effect of probe position on reaction time is smallest under condition 1, and it increments gradually under conditions 2 and 3.

The first condition is maximally regular, i.e. both the IOI and the ISI are kept constant throughout. This regularity entails that the timing of the events is also maximally predictable. In the second condition, the sounds within sequences are regularly spaced, but the onset of each sequence is unpredictable. Hence, by comparing conditions 1 and 2, we test whether the crucial factor producing a decrease in reaction time lies in the uncertainty of knowing *when* the first stroke of the sequence will be heard. An unpredictable beginning would produce a sequence-initial disadvantage, which would then disappear as further strokes are played with predictable timing. The results do confirm a slightly bigger initial-disadvantage under condition 2 compared to condition 1. Still, (1) with the current sample size, the difference fails to reach statistical significance, and, more importantly (2) the first condition still shows an initial disadvantage, even if the onsets of sequences are completely predictable.

The predictability-driven initial disadvantage (and final advantage) relies on the general readiness principle: if one can predict *when* an event will occur, the speed and accuracy with which we respond to the event is enhanced (Woodworth 1938; Niemi & Näätänen 1981). Despite the difference between the first two conditions, the longer gap which precedes sequences in both cases can be interpreted as a disruption of readiness.

Beyond readiness, finer-grained models of how attention is modulated as a function of predictability become relevant. According to the dynamic attention model (Large & Jones 1999), when we track an external regular rhythm such as a beat sequence, our attention is modulated at the same rate as the rhythm via entrainment. Empirical work (Jones et al. 2002; Fitzroy & Sanders 2015; Jongsma, Desain & Honing 2004) has shown that performance (a proxy for attention) peaks at the points where a beat is predicted, and decreases elsewhere. As the underlying mechanism, it is hypothesised that neural populations synchronise to the external rhythm by firing at the same rate.

The dynamic attention account can explain an increasing advantage later in the line, as the neural entrainment comes into place and attention tracks incoming strokes more precisely. However, this account relies on the regularity of the strokes for the increasing advantage to take place. An alternative account which does not rely on the isochrony of the input is the Bayesian predictive coding model (Vuust & Witek 2014). In this case, our prediction of events gets continuously updated as new stimuli are processed, regardless of isochrony. New strokes of the same kind reinforce our prediction, and performance is enhanced as a consequence.

The third experimental condition tackles this critical point: the IOI or within-sequence regularity. Our results show that, compared to the other two conditions, the initial disadvantage is further increased (Figure 5.2), and the overall effect of probe position on reaction time is significantly greater (Table 5.2). These results fit better the predictive coding rather than the dynamic attention model. Under the latter, the lack of isochrony of condition 3 would predict that the final advantage is diminished, but the opposite is true. The alternative explanatory mechanism is more general, since it relies on the building of predictions based on previous regularities, though not necessarily temporal. This has the potential of being applicable to a broader range of verse types, not restricted to prototypical metrical songs (where a beat can be felt), but including, for instance, non-isochronous poetry recitation.

Fluctuations of attention across verse lines offer a possible explanation of final strictness defined as a decrease in the frequency of deviations. Nevertheless, there exist other phenomena related to final strictness, such as rhyme, or the categorical control of word-length at the end of the line (Fabb 2002:174). An increasingly efficient use of attention is not well suited for these other kinds of final strictness, where the very end of the line is targeted. We should conclude, instead, that verse final strictness is driven by a variety of factors, including attention and the highlighting of constituent boundaries.

Further work is required in order to bridge two critical gaps. First, the low-level odd-ball task used here should be followed up with more ecological stimuli using e.g. verbal material (i.e. syllables), and rhythmic sequences. Many poetic metres, for instance, rely on the alternation of strong and weak positions; hence, deviations from the norm in that kind of context are more complex than in the present paradigm, where violations deviate from a single standard tone. Second, the gap between perception and production needs to be addressed, since the final strictness evidence which motivated the study relies on how poets and singers *produce* their lines of verse, not on how they *perceive* them. Unavoidably, the extent to which the attention mechanisms described here are applicable in a comparable way during the generation of lines (or other non-linguistic sound sequences) needs to be determined by production experiments.

5.5 Conclusion

Versification systems are cultural phenomena shaped by a complex interaction of factors. Typological tendencies such as final strictness can shed light on some of the underlying principles which both make possible *and* constrain the production of songs and poems. When the subjects in our experiment were asked to track the sequences of drum strokes and react to the deviant ones, their performance consistently dropped after the sequence-dividing gap. We propose that a similar drop of attention can play a role in the reduced faithfulness to templates found in songs and poems. Nonetheless, it should be kept in mind that, besides cognitive or anatomical constraints, verse is also shaped by aesthetic ideas which purposefully satisfy and violate our expectations.

References

Allen, W. 1973. *Accent and rhythm: prosodic features of Latin and Greek: a study in theory and reconstruction*. Cambridge University Press.

Anantapadmanabhan, A., A. Bellur & H. A. Murthy. 2013. Modal analysis and transcription of strokes of the mridangam using non-negative matrix factorization. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing: proceedings; 2013 May 26-31; Vancouver, USA. Piscataway, NJ: IEEE, 2013. p. 181-185.* Institute of Electrical & Electronics Engineers (IEEE).

Arnold, E.-V. 1905. *Vedic metre in its historical development*. Cambridge University Press.

Bates, D., M. Mächler, B. Bolker & S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.

Bouwer, F. L. & H. Honing. 2015. Temporal attending and prediction influence the perception of metrical rhythm: evidence from reaction times and ERPs. *Frontiers in psychology* 6.

Dell, F & M Elmedlaoui. 2008. *Poetic meter and musical form in Tashlhiyt Berber songs*. Köln: Rüdiger Köppe.

Fabb, N. 2002. *Language and literary structure: the linguistic analysis of form in verse and narrative*. Cambridge University Press.

Fabb, N. 2015. *What is poetry? Language and memory in the poems of the world*. Cambridge University Press.

Fitzroy, A. B. & L. D. Sanders. 2015. Musical meter modulates the allocation of attention across time. *Journal of cognitive neuroscience*.

Golston, C. & T. Riad. 2000. The phonology of Classical Greek meter. *Linguistics* 38(1). 99–167.

Huron, D. 2006. *Sweet anticipation: music and the psychology of expectation*. Massachusetts Institute of Technology.

Jones, M. R., H. Moynihan, N. MacKenzie & J. Puente. 2002. Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science* 13(4). 313–319.

Jongsma, M. L. A., P. Desain & H. Honing. 2004. Rhythmic context influences the auditory evoked potentials of musicians and nonmusicians. *Biological Psychology* 66(2). 129–152.

Kiparsky, P. 1968. Metrics and morphophonemics in the *kalevala*. In C Gribble (ed.), *Studies presented to roman jakobson by his students*, 137–148. Cambridge (MA, USA): Slavica.

Kuznetsova, A., P. Bruun Brockhoff & R. Haubo Bojesen Christensen. 2016. *lmerTest: Tests in Linear Mixed Effects Models*. R package version 2.0-32.

Large, E.-W. & M.-R. Jones. 1999. The dynamics of attending: how people track time-varying events. *Psychological Review* 106(1). 119–159.

Niemi, P. & R. Näätänen. 1981. Foreperiod and simple reaction time. *Psychological Bulletin* 89(1). 133–162.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Roberts, S. G., J. Winters & K. Chen. 2015. Future tense and economic decisions: controlling for cultural evolution. *PloS one* 10(7). e0132145.

Schwartze, M., N. Farrugia & S. A. Kotz. 2013. Dissociation of formal and temporal predictability in early auditory evoked potentials. *Neuropsychologia* 51(2). 320–325.

Vuust, P. & M. A. Witek. 2014. Rhythmic complexity and predictive coding: a novel approach to modeling rhythm and meter perception in music. *Frontiers in psychology* 5.

Woodworth, R. 1938. *Experimental psychology*. New York: Henry Holt.

Part III

How words are set to templates

6 A computational analysis of textsetting

6.1 Introduction

Songs can be perceived globally as homogeneous objects, but we can also consider them composite objects with two main components: a text and a tune. The autonomous status of the tune can be observed, for instance, in cases where the same melody is used several times with varying lyrics, be it in the same or in different songs. Example 6.1 shows such a case (song `NLB073196`) taken from the MTC-FS corpus of Dutch songs (van Kranenburg et al. 2014) on which we have based this study. Tunes can therefore be considered abstract templates to which words are set. This makes songs similar to poems in that a template (e.g. iambic pentameter) is used productively to create new instances of verse. Compared to the instantiation of templates discussed in Chapter 4, though, song templates are less abstract (i.e. they can be whistled), but they are also more complex in that they contain additional variables (e.g. pitch).

The analysis of textsetting describes how the two components of songs (i.e. text and tune) are combined (see Proto 2015 for an overview). This is often done by formulating textsetting rules, which state whether particular combinations of linguistic and musical features are preferred or avoided in a given musical tradition. Textsetting studies typically focus on three dimensions common to language and music which can be either aligned or misaligned when setting words to a tune: (1) prominence, (2) pitch, and (3) constituency.

In their seminal paper on the analysis of English textsetting, Halle & Lerdahl (1993) develop an algorithm which describes how a line of text is to be aligned with a given tune, addressing exclusively the dimension of prominence (further elaborated in Halle 1999). As a case study, they discuss an example where a single tune is set to a wide range of words: the well-known sea-chanty *The Drunken Sailor*. The first step in their textsetting algorithm, slightly simplified, specifies: "assign all accented syllables to available [strong musical] beats from left to right"

Example 6.1: Songs can be analysed as composite objects combining a text and a tune. This tune (with slight variations) is used several times in the same song (`NLB073196`, van Kranenburg et al. 2014), while different words are set to it.

(Halle & Lerdahl 1993:15). This can be observed under Example 6.2, where both lines are aligned so that their stressed syllables (marked with an accent) match the four beats (marked with asterisks) of the *drunken sailor* tune. Note, for instance, that the weak musical position where the word *his* is placed (first line) gets 'skipped' by the stressed syllable *there* in the second line so that it falls on the beat.

The kind of procedure described by Halle & Lerdahl (1993) seeks to produce text-to-tune alignments which are accepted by subjects familiar with the tradition by avoiding, for instance, placing stressed syllables in musically weak positions. This is based on the intuition that music and language share some notion of prominence (Jackendoff 2009), and that textsetting grammars avoid 'incongruent' alignments (Hayes 2009).

Pitch is another dimension shared by speech and music, and it also plays a role in the textsetting grammar of at least some languages. Constraints on pitch alignment have been proposed particularly for tonal languages (for a review, see Schellenberg 2012), where pitch contours are used to create lexical contrasts (Yip 2002). A standard method of evaluating the degree of textsetting control is to calculate the proportion of cases in a song where the tonal and melodic contours move in a similar direction. In a sample of 20 Vietnamese popular songs, Kirby & Ladd (2016) found that cases of opposing pitch contours (i.e. linguistic and musical pitch moving in opposite directions) are exceedingly rare, occurring less than 5% of the times. Similar degrees of correspondence have been reported for other Asian (e.g. Cantonese) and African (e.g. Ewe) tonal languages (Schellenberg 2012).

Finally, grouping and constituent structure is yet another dimension shared by language (Nespor & Vogel 2007; Matthews 2007) and music (Deutsch 2013; Lerdahl & Jackendoff 1983). Compared to stress and tone, however, the alignment

Example 6.2: Two sample lines of the song *The Drunken Sailor*. Stressed syllables (marked with accents) are aligned with the musical beats (marked with asterisks). See Halle & Lerdahl (1993) for details.

*			*		*		*	
stíck	*on*	*his*	*báck*	*a*	*mús*	*tard*	*plás*	*ter*
kéep	*him*		*thére*	*and*	*máke*	*him*	*báil*	*her*

of constituent boundaries has received less attention in textsetting studies (but see e.g. Dell & Halle 2009; Halle 2004). Nevertheless, the logic to evaluate the correspondence between text and template remains comparable to the domains of prominence and pitch; the (mis)alignment of constituent boundaries is taken as evidence for textsetting control or lack thereof.

This chapter deals exclusively with the first dimension, i.e. the alignment of linguistic stress and musical prominence, in Dutch folk songs. The Dutch language uses word stress (Booij 1995); e.g. the first syllable in a word like *bóter* 'butter' is more prominent than the second syllable, exactly as in its English translation. Similarly, in metered music some positions are more prominent than others (London 2004). In the remaining of the chapter, we will use *stress* or its abbreviation *s* to talk about linguistic prominence, and *prominence* or just *p* to talk about musical prominence. As a notational shorthand, we indicate linguistic stress with an acute accent ($\sigma\acute{\sigma}$), and underline the syllable which receives the highest metrical prominence within a given context ($\sigma\underline{\sigma}$).

The goal of this chapter is twofold. From a methodological point of view, it presents a systematic way of addressing the textsetting problem computationally (most studies on the topic rely on manual analyses; but see, e.g. Temperley & Temperley 2013). Standardised methods of this kind are critical to test and compare the extent to which textsetting rules apply in the languages of the world. Secondly, as a case study, it provides a first description of textsetting in Dutch folk songs. We demonstrate that the alignment of stress and prominence is not random, and that it interacts with other musical and linguistic factors (presence of melisma and part of speech).

6.2 Method

6.2.1 Material

In order to study the textsetting rules of Dutch folk songs we analysed 3,724 songs from the MTC-FS corpus (van Kranenburg et al. 2014). Most of the songs were collected through fieldwork between the 1950s and the 1990s as part of the radio programme *Onder de groene linde* led by Ate Doornbosch. The corpus also contains similar songs taken from 19th and 20th century songbooks.[1]

The original corpus contains 3,861 songs. However, the features we focus on (stress and prominence) were not always obtainable. Songs encoded as having free rhythm (n = 125) were excluded because they lack the feature of prominence. Linguistic stress for the lyrics was obtained through a nearest-neighbour lookup in the e-Lex lexical database[2] (as specified in van Kranenburg & Karsdorp 2014). Thus, the database lookup is robust against minor variations in spelling. Cases in which the nearest neighbour in the e-Lex database has a different number of syllables than the word in the lyrics were discarded. Any phrase containing one such word has also been excluded from the analysis (n = 2,451 phrases).

Every song in the dataset is divided into stanzas; stanzas are divided into phrases; phrases contain notes, which can then be associated to syllables. For the purposes of this chapter, stanzas are equivalent to musical strophes, and phrases are also referred to as lines. The filtered dataset contains 3,724 songs, 3,973 stanzas, 20,662 phrases, 185,263 notes, and 176,708 syllables. Syllables and notes are often in a one-to-one correspondence. Some syllables, though, span over more than one note; such a syllable is referred to as a *melisma*. In the filtered dataset, 4.46% of the syllables are melismas.

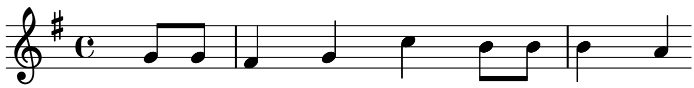
6.2.2 Corpus annotation

Stress is not a feature present in the original dataset, it was looked up at the e-Lex database. Stress is encoded in a binary way in the database; each syllable gets a value of either `0` (unstressed) or `1` (stressed). Dutch is described as containing secondary stress too (Booij 1995:105), but this is not explicitly encoded in the e-Lex database. Example 6.3 illustrates how stress and the following features related to stress and prominence have been automatically annotated. We also added part-of-speech information using the CELEX database (Baayen, Piepenbrock & van Rijn

[1] The melody, text and metadata for each song are openly available in several formats at www.liederenbank.nl/mtc.

[2] http://tst-centrale.org/en/producten/lexica/e-lex/7-25

Example 6.3: Sample annotation of stress, prominence and their respective contours.



prominence	.25	.12	1	.25	.5	.25	.12	1	.25
p. contour		−	+	−	+	−	−	+	−
syllables	*om*	*haar*	*bo* -	*ter*	*duur*	*te*	*ver* -	*ko* -	*pen*
stress	1	1	1	0	1	0	0	1	0
s. contour		=	=	−	+	−	=	+	−

1995). Words were categorised into two classes: content words (nouns, verbs), and function words (articles, prepositions, conjunctions).

Musical prominence is also not explicitly encoded as a feature in the MTC-FS dataset. However, this feature can be inferred from the symbolic representation of the tunes. For each note, we know its position within the musical bar, and the time signature this bar belongs to (e.g. 6/8). Given that information, relative prominence can be derived (Lerdahl & Jackendoff 1983). This was done using the *music21* software (Cuthbert & Ariza 2010). Prominence values range from `0` to `1`, the first position of the bar being assigned a `1`.

Both stress and prominence are relative notions, that is, given a syllable in isolation, its raw stress/prominence value cannot be determined. Hence, to capture how stress and prominence are aligned, it becomes necessary to compare a syllable with its neighbours. We have achieved this by computing the transition for the stress and prominence values of each syllable compared to the values of the preceding syllable. This produces three possible stress/prominence contours: decreasing (–), level (=) and increasing (+), as illustrated in Example 6.3. Note that the first syllable of a song does not have any preceding syllable to compare its stress or prominence to; hence, its transitions values are not computed (set to `NA`).

6.2.3 Statistical analyses

The goal of the analyses is to find alignment patterns of linguistic features and musical features which are arguably avoided by the individuals who created the songs in the corpus (and other speakers of Dutch as an extension). A simplified illustration of our analysis goes us follows. We count the number of syllables following a combination of a linguistic stress contour and a musical prominence contour (e.g. `s–p+`), we then calculate the expected count of syllables with the given combination assuming a random alignment of text and tune, and finally compare the observed frequency to the expected one. If a given alignment shows a much lower frequency in the corpus compared to the expected baseline, we conclude that it is an avoided alignment in the textsetting grammar under study.

In order to calculate the baseline probability of a combination of any 2 features f_1 and f_2, we compute the product of their individual frequencies: $p(f_1, f_2) = p(f_1) \times p(f_2)$. For instance, in a subset of our data, the p+ prominence contour shows up in 35% of the syllables, and the s– stress contour in 15%. The expected frequency of s–p+ syllables would be $.35 \times .15 = .05$, i.e. roughly 5% of the data. The *observed* frequency, however, is substantially lower, at 0.3%, indicative of it being an ill-formed setting. In principle, the same general procedure applies for any additional number of features, allowing us to introduce further linguistic and musical variables and fine-tune the analysis:

$$p(f_1 \dots f_n) = \prod_{i=1}^{n} f_i \tag{6.1}$$

Following this approach, however, does not take into account that tunes and texts can behave as independent objects, as evidenced e.g. by everyday conversation and instrumental music. In order to evaluate the well-formedness of a text-to-tune *alignment*, it is irrelevant whether a particular musical or linguistic context is exceptional on its own. To avoid that, we treat the frequency of the set of musical features ($m_{1\dots j}$), and the frequency of the set of linguistic features ($l_{1\dots k}$) in an autonomous fashion. Their product provides the expected frequency of a textsetting alignment assuming the lyrics and tunes are, individually, well-formed objects:

$$p(m_{1\dots j}, l_{1\dots k}) = p(m_{1\dots j}) \times p(l_{1\dots k}) \tag{6.2}$$

Once we have obtained the expected frequencies for all the combinations of musical and linguistic features we are interested in, we compute the relation (in the form of a log ratio) between observed and expected frequencies. In cases where the log ratio is close to zero, we lack evidence to argue that that particular alignment is controlled by textsetting constraints, because the observed frequency of the alignment corresponds to that expected by a random textsetting. Here, instead, we focus on under-represented alignments, setting a conservative significance threshold at a log ratio of -2, i.e. cases where the observed frequency is less than four times the expected one (Agresti 2013:56).

Besides the main variables of interest (stress and prominence), we refine the analysis by including two additional linguistic features (word boundary and word

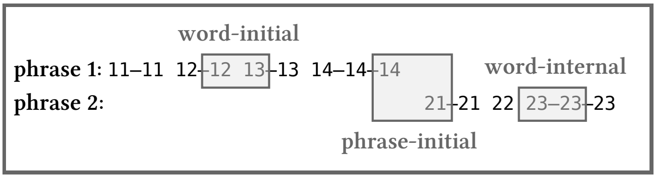


Figure 6.1: We define the textsetting domain with a two-syllable window: a target syllable on the right, and its preceding syllable. In the above figure, two abstract phrases are represented as a sequence of two-digit codes standing for syllables; the first digit indicates the phrase, and the second digit the word; syllables within the same word are connected by a hyphen. In the *word-initial* domain, the target syllable is preceded by a syllable from a different word; in the *phrase-initial* domain, by a syllable from a different phrase; and in the *word-internal* domain, by a syllable from the same word.

class), and two musical features (phrase boundary and melisma).

The boundary features serve to narrow down the relevant textsetting domain. Given that the stress and prominence variables are contour values over a two-syllable window, we can test whether textsetting constraints apply *across* musical/linguistic constituents. Does the prominence contour between the last note of a phrase and the first of the following line constrain the choice of words to be set? We compare three different domains, as illustrated in Figure 6.1: phrase-initial, word-initial, and word-internal.

Finally, we add two additional features which can affect the stringency of the textsetting constraints. We hypothesise that these are less stringent for so-called function or grammatical words, as it has been observed in Italian nursery rhymes (Proto & Dell 2013:112). Similarly, melismas can interact with the alignment of stress and prominence. In French, for instance, final unstressed syllables set to a relatively prominent note may be more acceptable if preceded by a melismatic syllable (Dell & Halle 2009:71).

6.3 Results

By combining the six features in our analysis, we obtain a six-dimensional textsetting model with all the 288 possible feature-value combinations and their associated log ratio, i.e. the extent to which this set of songs differs from a random textsetting alignment. Some of these combinations are completely absent from the corpus (e.g. there is no word-internal `s+p–` syllable preceded by a melisma).

Others are exceedingly rare, occurring only once in the whole corpus (e.g. a melismatic, word-internal s–p+ syllable belonging to a function word). A problem common to all corpus-based studies is that they are unable to provide negative evidence (Schütze 2011; 2016); hence, it is hard to determine whether absent alignments are accidental gaps, or whether they offer meaningful insights about forms perceived as ill-formed. Similarly, alignments with very low counts (such as single occurrences) present very wide confidence intervals, which make their interpretation unreliable.

6.3.1 Textsetting domain

In order to determine the domain in which textsetting is most constrained, we focus on a large subset of the data (content words with no melismas, n = 16,552 syllables) and compare the alignment of stress and prominence in the three domains depicted in Figure 6.1: line-initial, word-initial and word-internal. As shown in the Tables 6.1a–c (Supplementary Information), the alignment of stress and prominence is increasingly constrained in these three domains. None of the alignments exceeds the under-representation threshold (log ratio = −2) in the line-initial window. For the word-initial domain only one type of alignment is significantly avoided (s+p–, log ratio = −2.74). This alignment shows an even more extreme log ratio in the word-internal domain (−3.60), where an additional combination, s–p+, also exceeds the significance ratio (−4.14). This alignment, hence, represents the primary difference between a word-initial and a word-internal target syllable.

In absolute terms, the alignment s–p+ occurs approximately as often word-internally as word-initially (270 and 282 respectively). However, the word-internal examples are much more exceptional in comparison, because, assuming a random alignment, word-internal s–p+ are expected to occur more than four times as often as word-initial s–p+ (4,768 vs 1,092). This is explained by the trochaic bias in the Dutch lexicon, i.e. the prevalence of words with initial stress. Hence, the word *méisje* 'girl' with a s–p+ alignment on the second syllable (e.g. in song NLB072300) is more likely to be perceived as ill-formed, than the word *gemóed* 'mind' where the first, unstressed syllable is preceded by a stressed syllable, also exhibiting a s–p+ alignment (e.g. in song NLB161811; see Example 6.4; this and all subsequent examples are included in the Supplementary Information).

Among the two opposing alignments avoided word-internally, s–p+ is more strongly under-represented than s+p–, as established by the log ratios (−4.14 vs

−3.60, see Table 6.1c, Supplementary Information). This suggests native speakers would deem the alignment of the word *dróef<u>heid</u>* 'sorrow' in song NLB074530 as ill-formed more readily than the alignment of the word *<u>ge</u>lóof* 'faith' in song NLB111653 (Example 6.5). In the remaining analyses, we focus exclusively on word-internal syllables, as this context shows the strongest evidence for alignment constraints.

6.3.2 Linguistic context: content and function words

In a sample of 37 Italian nursery rhymes analysed by Proto & Dell (2013:112), function words exhibit textsetting misalignments more often than content words. Our results (Figure 6.2) support this difference in textsetting stringency between word types.

In the top panel we observe that only the s–p+ alignment is significantly avoided in function words (e.g. *ónder* 'under'), while the bottom panel shows that both s–p+ *and* s+p– are avoided in content words (e.g. *méisje* 'girl', *tambóer* 'drummer'). The total number of target syllables belonging to function words is much lower than those belonging to content words (814 vs 16,552), as reflected in the wider confidence intervals. Most observations of function words exhibit s– contours (top leftmost facet), which provides sufficient evidence for the avoidance of s–p+. Note, for instance, that the prepositions *óver* 'over' and *ónder* 'under' appear 227 and 188 times respectively, of which more than 90% follow a s–p– alignment. Function words with increasing stress contours are less frequent, and are often compounds, e.g. *totdát* 'until', *zoláng* 'as long as'. More than 20% of function words with an increasing stress contour exhibit a s+p– alignment, which does not allow for a robust avoidance rule. Consider, for instance, the word *totdát* 'until', which is the most frequent iambic function word in our sample: 9 cases follow a s+p+ alignment, and 12 cases follow a s+p– alignment (Example 6.6). This particular example and the overall results indicate that we cannot posit a definite textsetting rule for iambic function words.

6.3.3 Musical context: melisma

The association between prominence and stress can also be mediated by features of the musical context. Here we address the problem of melismatic syllables, i.e. syllables sung to more than one musical note. Overall, syllables preceded by a melisma are (1) less frequent, and (2) less controlled from a textsetting point of view, as shown in Figure 6.3. This loosening of rules introduced by melisma replicates previous observations made for French textsetting (Dell & Halle 2009:71).

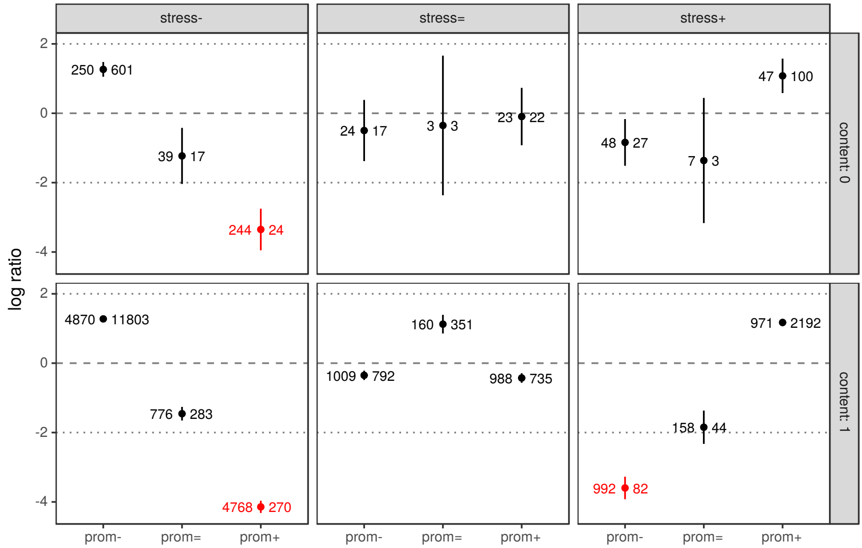


Figure 6.2: Association between stress and prominence in content words (bottom panel) vs function words (top panel). For each combination of stress and prominence values, the data points plot the log ratio between the observed frequency (number to the right of the point) and the expected frequency (left number). Error bars indicate the 95% confidence interval. Combinations below a log ratio of -2 are highlighted and interpreted as being significantly avoided.

As with function words, the relatively low frequency of contexts where a melismatic syllable precedes the target syllable produces less reliable results, as indicated by the larger error bars in the lower panel of Figure 6.3. The melismatic context is approximately 15 times less frequent than the non melismatic one, with the vast majority of the cases (92%) involving a decreasing stress contour. This lower leftmost facet provides the most reliable results: all three possible alignments (`s–p–`, `s–p=`, `s–p+`) are over-represented, even the `s–p+`, which is robustly avoided in the non-melismatic context (top leftmost facet). Consider the word *méisje* 'girl': in non-melismatic contexts, only 1% of the cases follow a `s–p+` alignment, while the proportion rises to 16% in the melismatic context.

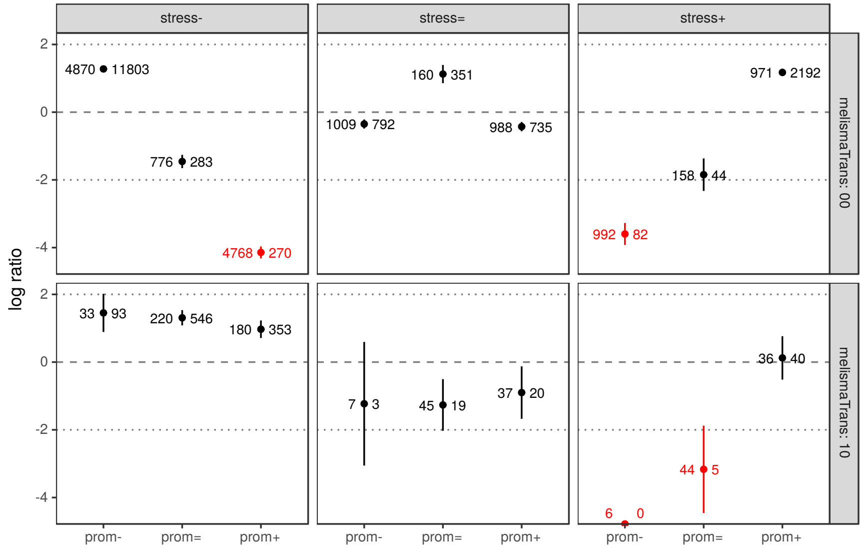


Figure 6.3: Association between stress and prominence in syllables preceded by a melisma (bottom panel) vs preceded by a non-melismatic syllable (top panel). For each combination of stress and prominence values, the data points plot the log ratio between the observed frequency (number to the right of the point) and the expected frequency (left number). Error bars indicate the 95% confidence interval. Combinations below a log ratio of −2 are highlighted and interpreted as being significantly avoided.

6.4 Discussion

The avoided combinations described in the previous section can be reformulated in terms of constraints which are active when a Dutch speaker creates a new song. Analogously, these constraints are arguably used in perception to judge a text-to-tune alignment as well- or ill-formed. Corpus studies like this one can put forward working hypotheses on such textsetting constraints, but their psychological reality is better tested through production (Hayes & Kaun 1996) and perception experiments (Gordon, Magne & Large 2011, and Chapter 7).

The fact that stress and prominence are relative entails that, minimally, the window where their alignment is compared must comprise two adjacent syllables. Our results suggest that this window is reset afresh at the beginning of each phrase, as evidenced by the lack of avoided alignments phrase-initially. Word-initially, the alignment is somehow controlled, although less so than within word-internal windows. Further analysis is needed in order to determine with more precision the relevant domain where textsetting rules apply (e.g. the phonological phrase, as argued for Italian by Proto & Dell 2013). In this respect, the current annotation of the corpus only offers information on word boundaries; in some cases, however, a word boundary can also mark the beginning/end of a higher-order constituent, such as a phonological or intonational phrase (Nespor & Vogel 2007). The decrease of stringency we observe word-initially may be the result of a gradual loosening of textsetting constraints across increasingly higher-order constituents.

In a comparable way to the textsetting rules proposed for tonal languages (Schellenberg 2009; Kirby & Ladd 2016), the general rule in our corpus is that stress and prominence should not move in opposite directions. Further, there is an asymmetry between s–p+ and s+p– alignments, the former being more heavily avoided. These tendencies, however, apply to a lesser extent in function words, and after the presence of a melismatic syllable.

By including these two finer-grained linguistic and musical contexts to the analysis, we have illustrated how the simple alignment of stress and prominence can be conditioned by further variables. The methodology straightforwardly allows for the inclusion of additional features, although interpreting the interactions between numerous variables can become complex, especially because it leads to an increasing number of absent (and thus uninformative) combinations in the corpus. All in all, there are at least two additional features which can prove fruitful for future studies.

On the linguistic side, we have followed a binary treatment of stress, as provided by most lexical datasets, but more detailed annotations can include further degrees, taking into account secondary stress and phrasal context (Hulst 1984; Booij 1995). Moving beyond stress binarity places the linguistic prominence on a gradual scale comparable to that of the musical metre (see Hayes, Wilson & Shisko 2012 for such an analysis in a corpus of English verse).

On the musical side, note duration and pitch are essential elements of the template, which we have ignored from the analysis so far. This omission is common in textsetting studies of stress languages, as it is generally regarded that pitch is less relevant for non-tonal languages (Halle & Lerdahl 1993, but see Särg & Ambrazevičius 2007). Nonetheless, there are reasons to further investigate the alignment of words to the melodic tier. First, pitch is one of the main cues for stress in languages such as Dutch or English (Sluijter & Heuven 1996), even if the associated contours are not as fixed as in tone languages (Hyman 2006). Second, pitch contours are also a source of prominence in musical tunes, creating so-called melodic accents (Thomassen 1982; Müllensiefen, Pfleiderer & Frieler 2009). These usually correlate and hence reinforce metrically prominent positions, but, crucially, they have the potential of producing prominence independently of metre (Särg & Ambrazevičius 2007 for a case study in Estonian and Lithuanian songs).

By discovering the musical and linguistic features relevant for text-to-tune alignment, we may shed some light on the cognitive processes involved in the simultaneous processing of music and language. Ultimately, the fact that speakers of a language are readily sensitive to the co-occurrence of features from these two domains needs to be accounted for by cognitive models aiming at explaining the shared or independent neural resources involved (Zatorre & Baum 2012; Hausen et al. 2013; Peretz 2012).

Finally, a thorough description of the textsetting grammar of any song style represents a valuable back-engineering tool. When presented with a historical set of lyrics for which the original tune is missing, the alignment constraints can be effectively used to search among a set of plausible melodies the one yielding the least violations (see e.g. the contributions within Proto, Canettieri & Valenti 2015). Even in the absence of potential melodies, a textsetting grammar can serve as a base model to generate, from scratch, well-formed templates (Conklin 2016), constrained by the specifications of the tune-less lyrics we are interested in. Naturally, the same procedure applies to the reverse process of generating a new text (or locating a missing one) given an available melody.

6.5 Conclusion

We have described a method which is general enough to describe in a systematic way textsetting constraints in the song traditions of different languages. The MTC-FS corpus of Dutch songs has been used as a case study, where we have described a number of avoided combinations and how they are affected by the linguistic and musical context. Further experimental validation and cross-linguistic studies will determine the most relevant features involved in textsetting grammars, and their typological prevalence.

References

Agresti, A. 2013. *Categorical data analysis*. Hoboken (New Jersey): John Wiley & Sons.

Baayen, R. H., R Piepenbrock & H van Rijn. 1995. *The CELEX lexical database (release 2)*. Web version (last consulted: 2016.06.24). Philadelphia: Linguistic Data Consortium.

Booij, G. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.

Conklin, D. 2016. Chord sequence generation with semiotic patterns. *Journal of Mathematics and Music* 10(2). 92–106.

Cuthbert, M. S. & C. Ariza. 2010. Music21: a toolkit for computer-aided musicology and symbolic music data. In *Proceedings of the 11th international society for music information retrieval conference (ismir 2010)*, 637–642. ISMIR.

Dell, F. & J. Halle. 2009. Comparing musical textsetting in French and English songs. In J.-L. Aroui & A Arleo (eds.), *Towards a typology of poetic forms*, 63–78. Benjamins.

Deutsch, D. 2013. Grouping mechanisms in music. In D. Deutsch (ed.), *The psychology of music*, 183–248. Elsevier.

Gordon, R. L., C. L. Magne & E. W. Large. 2011. EEG correlates of song prosody: A new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology* 2(352).

Halle, J & F Lerdahl. 1993. A generative textsetting model. *Current Musicology* 55. 3–23.

Halle, J. 1999. *A grammar of improvised textsetting*. Columbia University Doctoral dissertation.

Halle, J. 2004. Constituency matching in metrical texts. In *Proceedings of the Words and Music conference*. University of Missouri-Columbia.

Hausen, M., R. Torppa, V. R. Salmela, M. Vainio & T. Särkämö. 2013. Music and speech prosody: a common rhythm. *Frontiers in psychology* 4. 1–16.

Hayes, B. 2009. Textsetting as constraint conflict. In J.-L. Aroui & A Arleo (eds.), *Towards a typology of poetic forms*, 43–61. Benjamins.

Hayes, B. & A. Kaun. 1996. The role of phonological phrasing in sung and chanted verse. *The Linguistic Review* 13. 243–303.

Hayes, B., C. Wilson & A. Shisko. 2012. Maxent grammars for the metrics of Shakespeare and Milton. *Language* 88(4). 691–731.

Hulst, H. Van der. 1984. *Syllable structure and stress in Dutch*. Foris.

Hyman, L. M. 2006. Word-prosodic typology. *Phonology* 23(2). 225–257.

Jackendoff, R. 2009. Parallels and nonparallels between language and music. *Music Perception* 26(3). 195–204.

Kirby, J. & D. R. Ladd. 2016. Tone-melody correspondence in Vietnamese popular song. *Tonal Aspects of Languages 2016.* 48–51.

Lerdahl, F & R Jackendoff. 1983. *A generative theory of tonal music.* Massachusetts Institute of Technology.

London, J. 2004. *Hearing in time, psychological aspects of musical meter.* Oxford University Press.

Matthews, P. H. 2007. *Syntactic relations: a critical survey.* Vol. 114. Cambridge University Press.

Müllensiefen, D, M Pfleiderer & K Frieler. 2009. The perception of accents in pop music melodies. *Journal of New Music Research* 38(1). 19–44.

Nespor, M. & I. Vogel. 2007. *Prosodic phonology: with a new foreword.* Walter de Gruyter.

Peretz, I. 2012. Music, language, and modularity in action. In P Rebuschat, M Rohrmeier, J.-A. Hawkins & I Cross (eds.), *Language and music as cognitive systems*, 254–268. Oxford University Press.

Proto, T & F Dell. 2013. The structure of metrical patterns in tunes and in literary verse. Evidence from discrepancies between musical and linguistic rhythm in Italian songs. *Probus* 25(1). 105–138.

Proto, T. 2015. Prosody, melody and rhythm in vocal music: the problem of textsetting in a linguistic perspective. *Linguistics in the Netherlands* 32(1). 116–129.

Proto, T., P. Canettieri & G. Valenti. 2015. *Text and tune: on the association of music and lyrics in sung verse.* Bern: Peter Lang.

Särg, T & R Ambrazevičius. 2007. Melodic accent in Estonian and Lithuanian folk songs. In K Maimets-Volt, R Parncutt, M Marin & J Ross (eds.), *Proceedings of the 3rd conference on interdisciplinary musicology.* Published at: http://www-gewi.uni-graz.at/cim07. Tallinn (Estonia): Estonian Academy of Music & Theatre.

Schellenberg, M. 2009. Singing in a tone language: Shona. In *Selected proceedings of the 39th annual conference on African linguistics*, 137–144. Cascadilla Proceedings Project Somerville, MA.

Schellenberg, M. 2012. Does language determine music in tone languages? *Ethnomusicology* 56(2). 266–278.

Schütze, C. T. 2011. Linguistic evidence and grammatical theory. *Wiley Interdisciplinary Reviews: Cognitive Science* 2(2). 206–221.

Schütze, C. T. 2016. *The empirical base of linguistics: grammaticality judgments and linguistic methodology.* Berlin: Language Science Press.

Sluijter, A. M. & V. J. van Heuven. 1996. Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Fourth international conference on spoken language, icslp 96*, vol. 2, 630–633. IEEE.

Temperley, N & D Temperley. 2013. Stress-meter alignment ine French vocal music. *Journal of the Acoustical Society of America* 134(1). 520–527.

Thomassen, J.-M. 1982. Melodic accent: experiments and a tentative model. *Journal of the Acoustical Society of America* 71. 1596–1605.

van Kranenburg, P & F Karsdorp. 2014. Cadence detection in Western traditional stanzaic songs using melodic and textual features. In *Proceedings of the 15th international society for music information retrieval conference (ismir 2014)*, 391–396. ISMIR.

van Kranenburg, P, M de Bruin, L. de Grijp & F Wiering. 2014. The Meertens Tune Collections. *Meertens Online Reports* 1.

Yip, M. 2002. *Tone*. Cambridge University Press.

Zatorre, R. J. & S. R. Baum. 2012. Musical melody and speech intonation: singing a different tune. *PLoS Biology* 10(7). e1001372.

Supplementary Information

Table 6.1: Log ratios between observed and expected frequencies of stress and prominence alignment in three different domains: line-initial, word-initial, and word-internal.

(a) Line-initial domain (n = 2,032).

	prom-	prom=	prom+
stress-	-0.28	-1.65	-0.33
stress=	-1.43	-1.12	-0.40
stress+	-1.70	-0.13	0.72

(b) Word-initial domain (n = 31,616).

	prom-	prom=	prom+
stress-	0.90	0.15	-1.95
stress=	-1.10	-0.60	0.64
stress+	-2.74	-0.08	1.03

(c) Word-internal domain (n = 16,552).

	prom-	prom=	prom+
stress-	1.28	-1.46	-4.14
stress=	-0.35	1.13	-0.43
stress+	-3.60	-1.85	1.17

Table 6.2: Log ratios between the observed and expected frequencies of stress and prominence alignment in function and content words.

(a) Target syllable belongs to a function word (n = 814).

	prom-	prom=	prom+
stress-	1.27	-1.23	-3.35
stress=	-0.50	-0.35	-0.10
stress+	-0.84	-1.36	1.08

(b) Target syllable belongs to a content word (n = 16,552).

	prom-	prom=	prom+
stress-	1.28	-1.46	-4.14
stress=	-0.35	1.13	-0.43
stress+	-3.60	-1.85	1.17

Table 6.3: Log ratios between the observed and expected frequencies of stress and prominence alignment in two musical contexts: preceded by a melismatic or by a non-melismatic syllable.

(a) Target syllable preceded by a melismatic syllable (n = 1,079).

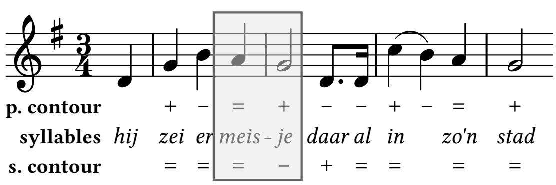
	prom-	prom=	prom+
stress-	1.45	1.31	0.97
stress=	-1.23	-1.27	-0.90
stress+	-Inf	-3.17	0.12

(b) Target syllable preceded by a non-melismatic syllable (n = 16,552).

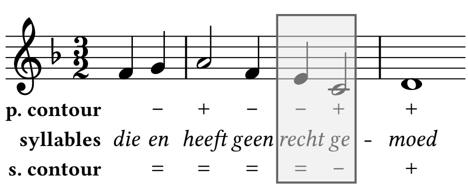
	prom-	prom=	prom+
stress-	1.28	-1.46	-4.14
stress=	-0.35	1.13	-0.43
stress+	-3.60	-1.85	1.17

Example 6.4: A decreasing stress contour aligned with an increasing prominence (s–p+) is heavily avoided word-internally, though not word-initially.

(a) Word-internal s–p+ (song NLB072300).

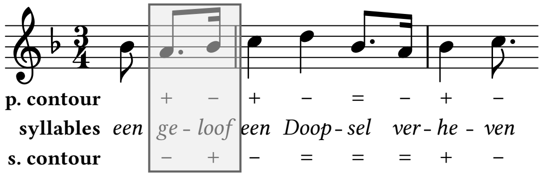


(b) Word-initial s–p+ (song NLB161811).

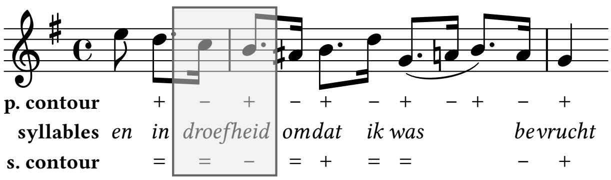


Example 6.5: In content words with no melisma, the alignment s–p+ is more heavily avoided than s+p–.

(a) The word *geloof* with s+p– (song NLB111653).

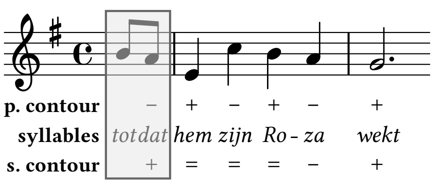


(b) The word *droefheid* with s–p+ (song NLB074530).

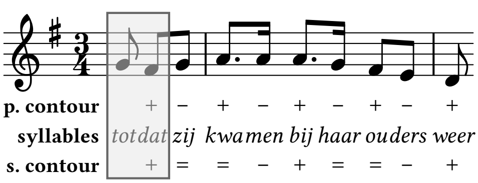


Example 6.6: In function words, the alignment s+p– is not robustly avoided. The word *totdat*, for example, shows 9 instances of s+p+, and 12 instances of s+p–.

(a) Function word *totdat* with s+p– (song NLB070749).

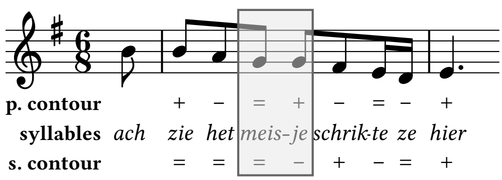


(b) Function word *totdat* with s+p+ (song NLB075303).

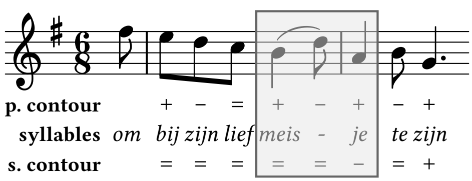


Example 6.7: The alignment `s–p+` is not avoided when the target syllable is preceded by a melisma. This example illustrates the case of the word *meisje*: in non-melismatic contexts, only 1% of the cases follow a `s–p+` alignment, while the proportion rises to 16% in the melismatic context.

(a) The word *meisje* with `s–p+` in a non-melismatic context (song `NLB070805`).



(b) The word *meisje* with `s–p+` in a melismatic context (song `NLB072415`).



7 Experimental testing of sensitivity to textsetting rules

7.1 Introduction

7.1.1 The problem of textsetting

Songs can be analysed as composite objects consisting of a tier of words set to a tune (Dell & Halle 2009). Evidence for the independence of these two levels of structure can be found in strophic songs, where the same tune is repeated several times with different lyrics set to it (like in Example 7.1). The alignment of these two tiers has been shown to be non-random in a number of languages, and its systematicity is captured by so-called *textsetting constraints* (see Chapter 6, and Proto 2015 for an overview).

In general, two broad types of constraints are proposed in the literature: (1) constituent alignment, and (2) feature alignment. To illustrate the first type of constraint, in Example 7.1 we can observe that the beginning of the musical phrase always coincides with the beginning of a linguistic phrase, rather than starting in the middle of a word, for example.

The second class of constraints usually describe a correlation between a prosodic feature (like stress or tone) and a musical feature (like metrical prominence or pitch contour). For instance, the opening lyrics of the song *Yesterday* (Example 7.1), musically, start on the beat, i.e. the first note has greater metrical prominence than the following two notes; and linguistically we observe a similar pattern, where the syllable *yes-* has greater stress than the following *-terday*. We can describe this pattern as a decreasing stress contour (`s–`), matched to a decreasing metrical prominence contour (`p–`).

This kind of alignment, summarised as `s–p–`, involves *parallel* contours, whereas a pattern like `s+p–` involves *opposing* contours. We would obtain the latter alignment by keeping the same melody but replacing the word *yésterday* with a word like *tomórrow*, which shows an increasing stress contour (`s+`) in its first two syllables. In languages employing linguistic stress (like Italian or English),

Example 7.1: The opening melody of the song *Yesterday* by The Beatles, with three different lyrics set to it.

parallel contours (s+p+, s–p–) are preferred over opposing contours (s+p–, s–p+) (Hayes 2009; Proto & Dell 2013).

From a methodological point of view, most analysis of textsetting rely on corpus data. Combinations of linguistic and musical features which are statistically rare or fully absent from a given corpus are considered ill-formed. Nevertheless, a well-known limitation of corpora is that they can only provide positive data; that is, absent or under-represented patterns are hard to interpret (Schütze 2011; 2016). This is particularly evident in smaller datasets, where it is likely to find accidental gaps with no statistical implications.

7.1.2 Experimental approaches

Despite these shortcomings, corpus analyses, combined with authors' judgements, provide precise hypotheses about which alignments may be perceived as (un)acceptable by native speakers. The current study and a few preceding ones attempt to substantiate and refine these hypotheses using experimental methods.

Hayes & Kaun (1996) developed a production task in order to address the textsetting intuitions of native speakers of English and compare them to corpus-derived rules. Participants were shown the words of 670 lines taken from a corpus of folk songs and asked to chant them as set to a binary template. The proposed settings showed a high degree of between-subject agreement, and further supported the textsetting patterns found in the original corpus.

More recently, Gordon, Magne & Large (2011) used a combination of behavioural and brain imaging data to test the sensitivity of participants to different textsetting patterns. The stimuli consisted of short sentences sung to newly-composed melodies paired with an isochronous beat. In some trials, the beat matched the stressed syllables; in others, the beat was displaced so that it matched unstressed syllables instead. Immediately after each trial, subjects performed a lexical decision task, which was executed more quickly and more accurately in the well-aligned trials (i.e. beats match stresses). Regarding the EEG

Table 7.1: Sample textsetting alignments with parallel and opposing contours. Each pattern is a combination of a stress (s) contour and a prominence (p) contour. The syllable with the greatest stress or prominence within the word is marked with an accent or an underline respectively.

	Contour	Pattern	Example
1	Opposing	s–p+	*méisje* 'girl'
2	Opposing	s+p–	*pastóor* 'priest'
3	Parallel	s–p–	*méisje* 'girl'
4	Parallel	s+p+	*pastóor* 'priest'

recordings, an analysis at the alpha (8–12 Hz), beta (13–29 Hz), and low-gamma (30–50 Hz) bands further showed significant differences between the aligned and misaligned settings.

Unlike corpus-based studies, the materials used in experiments can be designed so as to cover an exhaustive range of linguistic and musical patterns. This gives the researcher finer-grained control over the hypotheses to test. Hence, experiments can potentially provide positive *and* negative data, narrowing down the characterisation of textsetting intuitions.

7.1.3 The present study

The purpose of the chapter is twofold: (1) to further our understanding of Dutch textsetting, (2) to describe a simple yet effective methodology which can be employed to uncover the textsetting intuitions of a community of speakers of a given language.

To the best of our knowledge, the preceding chapter constitutes the first dedicated study of textsetting in Dutch. Based on a corpus of ca. 3,700 songs, we described a number of under-represented text-to-tune alignments. Among content words (i.e. nouns and verbs), the most avoided patterns involve opposing contours of stress and metrical prominence, as illustrated in the first two examples of Table 7.1. Patterns with parallel contours were, unsurprisingly, very common (rows 3 and 4 of Table 7.1). As a notational shorthand, we indicate linguistic stress with an acute accent, and underline the syllable which receives the highest metrical prominence in the word.

Based on the results from Chapter 6, the present study tests two hypotheses. First, we predict that native speakers of Dutch will be sensitive to the marked distributional contrast found in the corpus between parallel and opposing alignments. Second, we test whether they also show a difference in preference between the two opposing patterns, favouring s+p– cases like *pastóor* over s–p+ cases like *méisje* (see Table 6.2b and Example 6.5, Supplementary Information of Chapter 6).

The study also provides three methodological novelties. First, the musical dimension has been simplified by excluding pitch, yielding more controlled stimuli. It is known that the melodic contour in musical tunes induces variations in perceived prominence, e.g. via so-called *melodic accents* (Thomassen 1982; Huron & Royal 1996; Müllensiefen, Pfleiderer & Frieler 2009). Our stimuli, hence, employ speech-like pitch contours instead of musical melodies, thereby eliminating a potential confound which remains under-studied within the textsetting of non-tonal languages (cf. Särg & Ambrazevičius 2007).

Second, the differences between the trials from the different conditions is minimal: it affects the alignment of a single word, keeping a constant frame. By using these localised contrasts, we can pinpoint more accurately the variables driving the different judgements provided by the subjects, unlike other procedures where alignment differences affect all the words contained in a trial (Gordon, Magne & Large 2011:4).

Third, we assume that textsetting intuitions are gradual rather than binary, as it is claimed for metrics more generally (Ryan 2011). Hence, we apply a suitable methodology to derive a ranking from the most preferred to the least preferred patterns: a two-alternative forced-choice task (Thurstone 1927). Thus, instead of asking to rate the well-formedness of individual settings, subjects are asked to choose one out of two minimally-differing trials. Though mainly used in the field of psychophysics, this approach to uncovering gradual acceptability intuitions has been successfully applied in linguistic studies too (Cowart 1997; Stadthagen-González et al. 2017).

7.2 Method

7.2.1 Participants

We used the Meertens Panel database[1] to recruit 135 participants via the internet (74 females, 61 males; mean age = 60.43). All participants are native speakers of Dutch, and most (93 %) were born and currently live in The Netherlands.

[1] The database is managed by the Meertens Institute (Amsterdam, The Netherlands). More information can be found here: www.meertens.knaw.nl/meertenspanel/.

Table 7.2: Drum pattern to which the experimental sentences were aligned. The labels B1–4 provide a reference for the four beats contained in the pattern.

	B1		B2		B3		B4	
Bass drum	●				●	●		
Snare drum			●				●	
Closed hi-hat	●	●	●	●	●	●	●	●
	*							
Relative	*				*			
prominence	*		*		*		*	
	*	*	*	*	*	*	*	*

7.2.2 Procedure

The whole experimental procedure was conducted online, so each participant performed the task in the environment of their choice. Subjects were presented a screen with instructions, followed by 36 pages with the experimental task, i.e. a two-alternative forced-choice task per page. In each of these pages, they were asked to listen to two separate audio recordings. Each recording consisted of a sentence (4-5 words) and a simultaneous drum sequence in the background. Participants were asked to select one of the two recordings based on how well the words fitted the background rhythm. Both the order of the 36 comparisons and the order of the two recordings within each comparison were randomised for each participant.

7.2.3 Stimuli

Each stimulus had the same structure: a sentence of four or five words is spoken by a female native speaker of Dutch while a metrical drum sequence is heard in the background. The drum sequence was the same in every recording, but the string of words and the word-to-drum alignment varied.

The drum sequence (Table 7.2) is expected to provide strong metrical cues to Western listeners (Bouwer, Van Zuijen & Honing 2014). The sequence consists of a concatenation of eight sound events; the time points labelled as B1, B2, B3, B4 are equally spaced, with a time-span of 666 ms between adjacent positions. Listeners are likely to perceive a sense of beat at this rate (Honing 2013), which would yield a tempo of 90 beats per minute. The pattern in Table 7.2 is repeated four times in

Table 7.3: Complete set of base sentences used to create the 24 sound stimuli.

1	*willem* willem	*lévert / bestélt* supplies / orders	*mooie kleren* nice clothes
2	*marloes* marloes	*óefent / studéert* practises / studies	*een vreemde taal* a foreign language
3	*sandra* sandra	*ántwoordt / notéert* answers / notes down	*het laatste punt* the last point
4	*jeroen* jeroen	*tékent / verbéeldt* draws / imagines	*een bonte herfst* a colourful autumn
5	*femke* femke	*schíldert / verlícht* paints / illuminates	*onze kamer* our room
6	*matthijs* matthijs	*flúistert / verzínt* whispers / makes up	*lieve woordjes* sweet words

each stimulus, with a total duration of approximately eleven seconds. We used the open-source Hydrogen drum machine with the TR707 sample-set (Cominu, Wolkstein & Moors 2015) in order to generate the drum sounds. The lower part of Table 7.2 shows the relative metrical prominence which the drum pattern is expected to elicit (Bouwer, Van Zuijen & Honing 2014). Following the method described by Lerdahl & Jackendoff (1983), a larger amount of asterisks indicate greater metrical prominence.

In each recording, the participants were presented a sentence parallel to the drum pattern; these sentences, however, were recorded independently and then aligned with the drums in a controlled way. We used a total of six base sentences (Table 7.3), all following the same structure: subject - verb - object. For each base sentence, the subject and the object were kept constant, as well as their alignment with respect to the drum pattern. Hence, the critical section of the stimuli is the verb.

All the verbs here chosen are bisyllabic, and belong to two rhythmic categories: iambic or trochaic. Iambic verbs show an increasing stress contour (first syllable is unstressed, second is stressed), while trochaic verbs show a decreasing stress contour. In our manipulation, each verb contour can be aligned to an increasing or decreasing metrical context; this 2×2 design yields the four experimental conditions listed in Table 7.4. We refer to the conditions with the following acronyms,

Table 7.4: The four experimental conditions, produced by combining two contours of metrical prominence with two contours of linguistic stress. Underlined syllables indicate a position with greater metrical prominence than non-underlined ones.

	prominence−	prominence+
stress−	1. *lé-vert*	2. *lé-vert*
stress+	3. *be-stélt*	4. *be-stélt*

Table 7.5: The four experimental conditions using one of the six frame sentences. The critical section being manipulated (i.e. the verb) has been visually framed here. The sequence of starts depict the metrical structure of the background drum sequence, the number of starts being correlated with metrical prominence.

	B1		B2		B3		B4		B1		B2	
	*								*			
	*				*				*			
	*		*		*		*		*		*	
	*	*	*	*	*	*	*	*	*	*	*	*
1					*lé-*	*vert*						
2	*wil-*	*lem*		*lé-*	*vert*				*moo-*	*ie*	*kle-*	*ren*
3					*be-*	*stélt*						
4				*be-*	*stélt*							

where *s* stands for *stress*, and *p* for *metrical prominence*: (1) s−p−, (2) s−p+, (3) s+p−, (4) s+p+. Hereby, we highlight the fact that conditions 1 and 4 represent a match between the stress and the prominence contours, whereas conditions 2 and 3 combine divergent contours.

Table 7.5 exemplifies this by depicting the four stimuli produced by manipulating the first base sentence. The relative metrical prominence elicited by the drum pattern is represented with asterisks, as explained before. The first two events of the pattern, for instance, show a decrease in metrical prominence, going from 4 to 1 asterisk. In the framed section we highlight the critical part of the trials, where the verb displays the four possible alignments of linguistic and metrical prominence contours.

A native speaker of Dutch recorded the complete set of subjects, verbs and objects separately. She was instructed to pronounce the items as in speaking (not singing), but with an isochronous timing of the syllables, unlike in everyday speech. To ensure this, she produced the items while listening to a metronome

Table 7.6: Number of times (and proportion) the condition on the row is preferred over the condition on the column.

	1. s–p–	2. s–p+	3. s+p–	4. s+p+
1. s–p–	–	594 (.73)	605 (.75)	372 (.46)
2. s–p+	216 (.27)	–	465 (.57)	229 (.28)
3. s+p–	205 (.25)	345 (.43)	–	167 (.21)
4. s+p+	438 (.54)	581 (.72)	643 (.79)	–

track set at 180 beats per minute. Later, the relevant recordings were concatenated to form sentences, and aligned with the drum pattern according to one of the four conditions.

In order to test the textsetting intuitions of the participants, we presented them the four stimuli from each base sentence in a pairwise manner. This produces six comparisons per base sentence and a total of thirty-six comparisons using all the base sentences in Table 7.3.

7.2.4 Statistical analyses

We used two different models to assess which textsetting alignments were regarded as better-formed. Each judgement made by a subject yielded a winner (i.e. a preferred setting), and the relative number of wins and losses for each stimulus is used to obtain a ranking from the most preferred to the least preferred condition.

We first applied the Thurstone-Mosteller model (Thurstone's Law of Comparative Judgement, Case V, Thurstone 1927), and further refined the analysis with the Bradley-Terry model (Bradley & Terry 1952). Thurstone's model is far more widespread, but Bradley-Terry's offers a number of advantages, like the flexibility to incorporate additional (random) predictors (Handley 2001). Both models were implemented in `R` (R Core Team 2017) with the packages `psych` (Revelle 2017) and `BradleyTerry2` (Turner & Firth 2012).

Table 7.7: Results of the Thurstone-Mosteller test. Given the pairwise results in Table 7.6, the test produces values between 0 and 1 indicating the relative preference for each condition, the least favoured option given a 0 by convention.

Pattern	Example	Condition	Thurstone value
s+p+	*be-stélt*	4	0.79
s–p–	*lé-vert*	1	0.71
s–p+	*lé-vert*	2	0.17
s+p–	*be-stélt*	3	0.00

7.3 Results

Table 7.6 presents the number of times (and proportion) the conditions on the rows are preferred over the conditions on the columns. On the second column of row 1, for instance, we observe that the first condition (e.g. *lé-vert*) is favoured over the second (e.g. *lé-vert*) in 73% of the trials.

By applying the Thurstone-Mosteller model to these data, we obtain the relative ranking of preferences shown in Table 7.7. By convention, the least preferred option receives a value of 0, and the other options receive higher values, with a theoretical maximum of 1. As predicted from the literature, participants display a sharp division between parallel contours (conditions 1 and 4), and opposing contours (conditions 2 and 3).

In addition, an iambic verb set to a trochaic pattern (*be-stélt*) is more heavily dispreferred than its mirror setting (*lé-vert*). A smaller difference is found between the two parallel settings, where iambic verbs (*be-stélt*) are preferred over trochaic ones (*lé-vert*).

In order to assess the statistical significance of these preference differences between conditions, we fit the data to a Bradley-Terry model. The most dispreferred condition (3, *be-stélt*) is taken as a baseline, and the model tests what the relative increase in preference is provided by each of the other three conditions, and the robustness of the difference. The results are displayed in Table 7.8, where we verify that the preference shown by the remaining three conditions is statistically robust ($\alpha = 0.05$). By setting condition 1 (*lé-vert*) as a baseline, we can further confirm that the difference between the two parallel conditions is also statistically significant ($z = 2.43$, $p = 0.015$).

Additionally, we controlled for the potential effect of the position of a sentence on the screen. In each page, subjects were asked to play first one sentence, then

Table 7.8: Results of the Bradley-Terry test. Each row indicates the extent to which the predictor shown in the first column increases the likelihood of a trial being preferred over the baseline condition (3).

Predictor	Estimate	Std. Error	z value	Pr ($> z$)
condition1	1.17	0.056	20.9	4.22e-97
condition2	0.276	0.0535	5.16	2.51e-07
condition4	1.3	0.0568	22.9	9.43e-116
first.displayed	0.118	0.0316	3.73	0.000193

the other, and finally make a decision about the preferred one. It is likely that subjects listened first to the sentence displayed at the left side of the screen, and that they systematically showed a tendency to prefer or disprefer a sentence based on the display order of the stimuli. Hence, we added the predictor `first.displayed` to the model. Indeed, being shown first increases the likelihood of a sentence being preferred by a factor of 0.118 ($z = 3.73$, $p = 0.000193$, last row of Table 7.8). Still, the relative differences in preference ranking remain robust after correcting for order of display.

7.4 Discussion

Using a two-alternative forced-choice task we show that native speakers of Dutch prefer parallel rather than opposing contours. This preference for a congruent alignment is unsurprising and predicted by previous literature; yet, it provides support for the validity of the methodology employed. More interestingly, we also show that the two types of opposing pattern are not equally dispreferred. The reason why participants judged patterns like *be-stélt* as being worse than patterns like *lé-vert* may be grounded on phonological properties of Dutch, or stem from more general principles of how prominence is parsed.

First, there is a strong trochaic bias in the Dutch language (and, more generally, in Germanic), both synchronically and in acquisition (Fikkert 1994). Regarding the subset of the lexicon critical to our study, over 75 % of Dutch bisyllabic conjugated verb forms are trochaic (Baayen, Piepenbrock & van Rijn 1995). Even if the lexical items chosen as stimuli are comparable in terms of frequency of use, speakers could still be sensitive to the lexical prevalence of trochaic (i.e. `s–`) verb forms, and exhibit a `s–p+ > s+p–` preference. Nevertheless, if that general preference for trochaic forms was in place, we should expect it to affect parallel contours too. Still, subjects show a preference for iambic forms within the context of parallel contours (i.e. s+p+ is preferred over s–p–). The effect of a trochaic bias remains a possibility, but, given the limited number of lexical items here tested, the available evidence is inconclusive.

Second, an alternative account can be that an increase in stress within a decreasing prominence environment (s+p–) is more salient, and hence more readily dispreferred, than a decrease in stress in an increasing prominence environment (s–p+). This is somehow reflected in the Stress Maximum Constraint proposed for English poetry by Halle & Keyser (1971), although evidence from other unrelated (non-trochaic) languages would be required in order to generalise the principle. A stress maximum is defined as a stressed syllable between two unstressed syllables within a given constituent; according to the constraint, maxima are only allowed in metrically strong positions.

Under our analysis, prominence is always parsed left-to-right, in a strict temporal order, so that the relative prominence of a syllable n is defined only by the prominence of syllable $n - 1$, and not affected by the prominence of $n + 1$, as in the definition by Halle & Keyser (1971). Hence, among the stress contours of our critical bisyllabic verbs, only the second syllable of iambic verbs (*bestélt*) can be equated to a stress maximum.

According to Halle & Keyser (1971), stress maxima must occur in strong positions (p+ in our notation). That means that p– positions are more constrained, because they do not allow stress maxima (s+), but p+ positions in turn are free to receive any kind of syllable, whether s+ or s–. In our results we do observe a preference for p+ contexts to be aligned with s+ words (*be-stélt* > *lé-vert*), but this difference is smaller than the preference for s–p– (*lé-vert*) over s+p– (*be-stélt*). This suggests that, indeed, p– contours are more stringent, as argued by Halle & Keyser (1971).

Additional evidence for this view that s+p– is perceptually more salient than s–p+ comes from experiments testing the ease with which we process deviant tones in metrical contexts. Bouwer & Honing (2015) found that subjects were faster and more accurate in detecting unexpected amplitude *increments* in a metrical drum sequence, compared to unexpected *decrements* in amplitude.

In contrast with our study, this kind of approach provides higher temporal resolution, since subjects do not make judgements over a whole stimulus lasting several seconds, but react directly to deviant events. In order to address the issue of how unexpected a particular syllable is in a given metrical context, the offline preference judgements we have employed can be followed up with complementary online tasks.

On the behavioural side, phoneme monitoring can provide a measure of how disruptive different text-to-metre alignments are (Quené & Port 2005; Connine & Titone 1996; Finney, Protopapas & Eimas 1996). This bears resemblance with the lexical decision task used by Gordon, Magne & Large (2011), but with the advantage of recording the dependent variable closer to the required moment of the stimulus. Specifically, subjects can be asked to detect a target phoneme occurring at or just after the critical textsetting item (i.e. the verb in our case).

Finally, brain imaging techniques such as EEG can provide very fine-grained measures of how strongly different textsetting patterns violate the listener's expectations. To be sure, the most dispreferred patterns in our study are likely to produce noticeable mismatch negativities (Honing, Bouwer & Háden 2014; Winkler 2007). All in all, corpora remain invaluable sources to extract underrepresented patterns, and define the musical and linguistic features which can then be systematically manipulated to conduct experiments like the one here reported.

7.5 Conclusion

We have shown that participants are sensitive to four different alignments of linguistic stress and metrical prominence. This validates the two-alternative forced-choice task for textsetting purposes, and its use to complement corpus-based analyses; here, indeed, we have replicated the corpus findings described in Chapter 6. The extent to which the judgements depend on the phonological properties of the Dutch language or on general auditory cognition is to be determined by further research.

References

Baayen, R. H., R Piepenbrock & H van Rijn. 1995. *The CELEX lexical database (release 2)*. Web version (last consulted: 2016.06.24). Philadelphia: Linguistic Data Consortium.

Bouwer, F. L. & H. Honing. 2015. Temporal attending and prediction influence the perception of metrical rhythm: evidence from reaction times and ERPs. *Frontiers in psychology* 6.

Bouwer, F. L., T. L. Van Zuijen & H. Honing. 2014. Beat processing is pre-attentive for metrically simple rhythms with clear accents: an ERP study. *PloS one* 9(5). e97467.

Bradley, R. A. & M. E. Terry. 1952. Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika* 39(3/4). 324–345.

Cominu, A., M. Wolkstein & S. Moors. 2015. *Hydrogen drum machine. Version 0.9.6.1*. Computer program. URL: <www.hydrogen-music.org>.

Connine, C. M. & D. Titone. 1996. Phoneme monitoring. *Language and Cognitive Processes* 11(6). 635–645.

Cowart, W. 1997. *Experimental syntax: applying objective methods to sentence judgments*. California: Sage publications.

Dell, F. & J. Halle. 2009. Comparing musical textsetting in French and English songs. In J.-L. Aroui & A Arleo (eds.), *Towards a typology of poetic forms*, 63–78. Benjamins.

Fikkert, P. 1994. *On the acquisition of prosodic structure*. Holland Institute of Generative Linguistics.

Finney, S. A., A. Protopapas & P. D. Eimas. 1996. Attentional allocation to syllables in American English. *Journal of Memory and Language* 35(6). 893–909.

Gordon, R. L., C. L. Magne & E. W. Large. 2011. EEG correlates of song prosody: A new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology* 2(352).

Halle, M & S.-J. Keyser. 1971. A theory of meter. In *English stress: its form, its growth, and its role in verse*, 139–180. Harper & Row.

Handley, J. C. 2001. Comparative analysis of Bradley-Terry and Thurstone-Mosteller paired comparison models for image quality assessment. In *Pics*, vol. 1, 108–112.

Hayes, B. 2009. Textsetting as constraint conflict. In J.-L. Aroui & A Arleo (eds.), *Towards a typology of poetic forms*, 43–61. Benjamins.

Hayes, B. & A. Kaun. 1996. The role of phonological phrasing in sung and chanted verse. *The Linguistic Review* 13. 243–303.

Honing, H. 2013. Structure and interpretation of rhythm in music. In D Deutsch (ed.), *The psychology of music*, 369–404. Elsevier.

Honing, H., F. L. Bouwer & G. P. Háden. 2014. Perceiving temporal regularity in music: the role of auditory event-related potentials (ERPs) in probing beat perception. In H Merchant & V de Lafuente (eds.), *Neurobiology of interval timing*, 305–323. Springer.

Huron, D & M Royal. 1996. What is melodic accent? Converging evidence from musical practice. *Music Perception* 13(4). 489–516.

Lerdahl, F & R Jackendoff. 1983. *A generative theory of tonal music*. Massachusetts Institute of Technology.

Müllensiefen, D, M Pfleiderer & K Frieler. 2009. The perception of accents in pop music melodies. *Journal of New Music Research* 38(1). 19–44.

Proto, T & F Dell. 2013. The structure of metrical patterns in tunes and in literary verse. Evidence from discrepancies between musical and linguistic rhythm in Italian songs. *Probus* 25(1). 105–138.

Proto, T. 2015. Prosody, melody and rhythm in vocal music: the problem of textsetting in a linguistic perspective. *Linguistics in the Netherlands* 32(1). 116–129.

Quené, H. & R Port. 2005. Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62(1). 1–13.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Revelle, W. 2017. *psych: Procedures for Psychological, Psychometric, and Personality Research*. R package version 1.7.8. Northwestern University. Evanston, Illinois.

Ryan, K. M. 2011. Gradient syllable weight and weight universals in quantitative metrics. *Phonology* 28(03). 413–454.

Särg, T & R Ambrazevičius. 2007. Melodic accent in Estonian and Lithuanian folk songs. In K Maimets-Volt, R Parncutt, M Marin & J Ross (eds.), *Proceedings of the 3rd conference on interdisciplinary musicology*. Published at: http://www-gewi.uni-graz.at/cim07. Tallinn (Estonia): Estonian Academy of Music & Theatre.

Schütze, C. T. 2011. Linguistic evidence and grammatical theory. *Wiley Interdisciplinary Reviews: Cognitive Science* 2(2). 206–221.

Schütze, C. T. 2016. *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Berlin: Language Science Press.

Stadthagen-González, H., L. López, M. C. Parafita Couto & C. A. Párraga. 2017. Using two-alternative forced choice tasks and Thurstone's law of comparative judgments for code-switching research. *Linguistic Approaches to Bilingualism*.

Thomassen, J.-M. 1982. Melodic accent: experiments and a tentative model. *Journal of the Acoustical Society of America* 71. 1596–1605.

Thurstone, L. L. 1927. A law of comparative judgment. *Psychological review* 34(4). 273–286.

Turner, H. & D. Firth. 2012. Bradley-terry models in R: the BradleyTerry2 package. *Journal of Statistical Software* 48(9). 1–21.

Winkler, I. 2007. Interpreting the mismatch negativity. *Journal of Psychophysiology* 21(3-4). 147–163.

8 Conclusion

Verse is a cultural system inherent to every human society, and I have argued that shared aspects of human cognition may have shaped the diversity of verse systems in three specific ways. First, the numeric requirements imposed by verse templates do not demand counting beyond four; second, when creating new texts, poets and songwriters are more faithful to templatic features towards the end of lines; third, the musical properties of templates interact with the prosodic properties of texts in a non-random manner. To be sure, these three findings constitute hypotheses with varying empirical support.

In the even chapters of the dissertation (2, 4, 6), I have developed strategies to describe verse systems, focusing on methodologies which can be applied in a consistent way across traditions. These methods have been applied to a range of languages in this dissertation, but their main merit is that they have the potential to highlight core, invariant aspects of verse if employed on larger samples. The experiments presented in the odd chapters (3, 5, 7) use simplified verse-like material to narrow down, in a controlled fashion, hypotheses on the factors driving the three recurrent verse features listed above. By combining these two types of studies, I promote a research programme where the analysis of verse phenomena directly engages with the study of human cognition. This programme constitutes a two-way street. On the one hand, the diversity of human verse provides patterns (such as final strictness) which need to be accounted for by theories of cognition; on the other hand, these theories (e.g. core systems of number, Feigenson, Dehaene & Spelke 2004) bring forth coherent sets of principles which are expected to shape any human cultural system, including verse.

As stated in the introduction, the overarching goal of the dissertation is to develop ways of explaining features of verse, particularly those with a widespread distribution. I have focused on explaining *formal* aspects of verse from a *cognitive* perspective. Specifically, Part 1 covers how to explain features of templates, and Parts 2 and 3 cover how to explain the way words are set to templates. Still, the enterprise can be expanded (1) by addressing other dimensions, such as the meanings, social dynamics, or functions of verse (e.g. Savage et al. 2015), (2) by investigating causal effects of verse other than those derived from shared cognitive capacities.

Indeed, the kind of explanations I have offered involve biases present at the online processing level of individuals, which entails that they are unable, on their own, to explain the tendencies already present in the verse corpora (Enfield 2014:18). Cognitive hypotheses about synchronic facts generally suffer from a linkage problem (Clark & Malt 1984:201), that is, we need to provide an account of how some (but not all) online biases get transmitted across generations so that a verse system as a whole shows a tendency synchronically.

> The synchronic facts are an aggregate outcome of individual people's biases multiplied in a community and through time. The bias has a causal effect precisely in so far as it affects the likelihood that a pattern will spread throughout that community. (Enfield 2014:18)

How these processing biases come into place phylogenetically or ontogenetically is not a topic treated in this dissertation. Similarly, it is beyond its scope to investigate processes of interaction and diachrony by which verse systems evolve under the pressure of e.g. language contact or psychological factors. In this respect, only the iterated learning experiment in Chapter 3 tackles a larger temporal frame, by analysing the effect of cultural transmission using a simple model where each subject indirectly interacts with another subject by trying to imitate accurately their syllable sequences.

The themes which I have developed in the dissertation do provide, nonetheless, building blocks on which three specific strands of future studies can build readily.

First, I have highlighted the need for comparative concepts (such as instance-strictness, Section 4.2.2) and standardised descriptive measures (such as the under-representation metrics to describe textsetting, Section 6.2). Substantiating robust typological tendencies requires the systematic application of these kinds of methods to balanced samples of verse traditions. Furthermore, a successful typology of verse would benefit from two additional collaborative ingredients. On the one hand, it is critical to compile existing resources on versification systems, whether raw data, or analytic and descriptive publications. This facilitates the building of samples and, more pressingly, it highlights major gaps in the documentation of endangered verse traditions. On the other hand, effective cross-linguistic studies become more straightforward when the comparative concepts and metrics are aggregated into conventional data structures, be them under unified databases or distributed repositories.

Second, the experimental work here presented offers a benchmark for future studies with speakers from languages other than Dutch. Psychology-related fields such as psycholinguistics or music cognition generally suffer from an ethnocentric bias, sometimes assuming that the principles observed in Western undergraduate students directly generalise to humankind as a whole (Majid & Levinson 2010). Hence, it would prove informative to conduct our experiments with subjects from a variety of backgrounds. The chunking and final strictness effects shown in Chapters 3 and 5, even if based on non-linguistic stimuli (i.e. drum strokes and meaningless syllables), are prone to be affected by the rhythmic properties of Dutch. To be sure, comparable studies on the grouping of acoustic events have demonstrated marked differences between speakers of e.g. English and Japanese (Iversen, Patel & Ohgushi 2008). While the experiments in this dissertation do show parallels with real verse data from several independent languages, it is too early to establish the extent to which the effects are universal or language-specific. In this respect, follow-up studies may sample speakers from languages with different word prosody, phrasal prosody, and information structure.

Finally, brain imaging techniques can complement the behavioural findings from Chapters 5 and 7. Methods with high temporal granularity such as EEG can help pinpoint the contexts which violate expectations, while using a minimally conscious task from the listener's perspective (cf. the judgement task in Chapter 7). Additionally, the continuous nature of neural signals (e.g. mismatch negativity, Duncan et al. 2009) provide an adequate means to account for the gradual nature of words-to-template alignment rules in songs and poetry (Ryan 2011), as argued in the case of final strictness (Part 2) and textsetting (Part 3).

References

Clark, H. H. & B. C. Malt. 1984. Psychological constraints on language: a commentary on Bresnan and Kaplan and on Givón. In *Method and tactics in cognitive science*, 191–214. Hillsdale, NJ: Lawrence Erlbaum.

Duncan, C. C., R. J. Barry, J. F. Connolly, C. Fischer, P. T. Michie, R. Näätänen, J. Polich, I. Reinvang & C. Van Petten. 2009. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology* 120(11). 1883–1908.

Enfield, N. J. 2014. *Natural causes of language: frames, biases, and cultural transmission*. Language Science Press.

Feigenson, L., S. Dehaene & E. Spelke. 2004. Core systems of number. *Trends in cognitive sciences* 8(7). 307–314.

Iversen, J. R., A. D. Patel & K. Ohgushi. 2008. Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America* 124(4). 2263–2271.

Majid, A. & S. C. Levinson. 2010. WEIRD languages have misled us, too. *Behavioral and Brain Sciences* 33(2-3). 103.

Ryan, K. M. 2011. Gradient syllable weight and weight universals in quantitative metrics. *Phonology* 28(03). 413–454.

Savage, P. E., S. Brown, E. Sakai & T. E. Currie. 2015. Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences* 112(29). 8987–8992.

Samenvatting

Dit proefschrift gaat over versbouw. Enerzijds heb ik een aantal eigenschappen geanalyseerd die wijdverspreid zijn over de wereld; anderzijds heb ik cognitieve hypothesen ontwikkeld die de verspreiding van deze eigenschappen zouden kunnen verklaren.

Wat is versbouw? Vanuit een breed perspectief vallen er een aantal verbale fenomenen binnen de definitie: liederen, gedichten, aftelversjes of leuzen bij demonstraties. In vergelijking met alledaags taalgebruik, hebben deze fenomenen extra structuurniveaus, zoals de regelmaat van geaccentueerde lettergrepen of klemtonen, het herhalen van een melodie, of een vaststaand aantal lettergrepen per regel. Deze eigenschappen kunnen, samen beschouwd, worden begrepen als een model, sjabloon of mentale structuur; verzen worden gevormd door woorden in het sjabloon te plaatsen.

Alle taalgemeenschappen in de wereld houden zich bezig met versbouw, dat wil zeggen vers is een universeel kenmerk van menselijke culturen, net als taal. Tenminste, tot nu toe zijn er geen talen zonder versbouw gevonden. Dat is op zichzelf mysterieus want we weten niet welke basisfunctie vers evolutionair gezien vervult, of wat elke gemeenschap heeft aangemoedigd om vers te creëren. Er zijn ongeveer 6000 talen in de wereld en ze hebben heel verschillende manieren ontwikkeld om liedjes en gedichten te produceren. Alhoewel elke gemeenschap, van generatie op generatie, innovaties introduceert, lijken er ook constanten te bestaan. Dit tweede mysterie is het startpunt voor dit proefschrift.

Waarom zijn rijm en binaire ritmen zo wijdverspreid? Waarom creëren we niet regelmatig verzen met veertig lettergrepen? Vele factoren kunnen onze versproductie bepalen; de anatomie van ons lichaam (de hersenarchitectuur, de grenzen van de adem), de sociale omgeving (dagelijks werk en vrije tijd, historische gebeurtenissen), en misschien zelfs onze fysieke omgeving (zingen ze zachter in koudere gebieden?). In deze brede zee van deze potentiële verklaringen, heb ik een specifieke hypothese ontwikkeld in dit proefschrift: mensen hebben cognitieve overeenkomsten en die kunnen overeenkomsten in verstradities van de wereld determineren.

Het onderzoek bestaat uit drie delen die drie verschillende kenmerken van versbouw bestuderen. Elk deel bestaat uit twee hoofdstukken, het eerste deel is gebaseerd op verscorpora van verschillende talen, en het tweede deel op perceptuele experimenten.

Het eerste deel gaat over de onderdelen van verssjablonen. Als men een nieuw sonnet à la Shakespeare wil schrijven, moet men een tekst van ongeveer 150 lettergrepen produceren. Toch weten we dat de 150 lettergrepen niet als een geheel begrepen worden maar worden georganiseerd in componenten en subcomponenten. Een sonnet bestaat uit drie kwatrijnen en een tweeregelig couplet, elke versregel bestaat uit twee delen, die twee of drie versvoeten bevatten, die uit twee lettergrepen bestaan. Na analyse van tientallen verstradities uit alle continenten (hoofdstuk 2) heb ik twee conclusies getrokken. (1) Alle verssjablonen bestaan uit onderdelen die te vinden zijn door structurele kenmerken te analyseren, zoals bijvoorbeeld eindrijm of verlenging van de lettergreep aan het einde van een regel. (2) Ondanks dat versdelen vaak herhaald worden, bijvoorbeeld bij twee regels met hetzelfde rijmpatroon en dezelfde melodie, worden versdelen met dezelfde kenmerken nooit meer dan vier keer na elkaar herhaald. Het is mogelijk dat deze grens in tradities over de hele wereld het gevolg is van cognitieve eigenschappen, aangezien hoeveelheden van meer dan drie of vier elementen lastiger zijn om meteen voor de geest te halen of zich te herinneren. In hoofdstuk 3 presenteer ik een hieraan gerelateerd experiment. Deelnemers wordt gevraagd om naar nietszeggende lettergrepen te luisteren en die weer te geven. Als de reeksen lang worden (bijvoorbeeld 12 lettergrepen), blijken deelnemers de neiging te vertonen om de lettergrepen te groeperen zoals dit ook gebeurt in verssjablonen.

Het tweede deel gaat over 'fouten' van versmakers (i.e. dichters, songschrijvers). In vele talen (bijv. in het Nederlands of in het Berber), hebben verssjablonen een bepaald ritme van lettergrepen nodig, bijvoorbeeld door afwisseling van zware en lichte lettergrepen. Dichters voldoen echter niet altijd aan deze normen. Zo hebben onderzoekers beschreven dat het einde van versregels regelmatiger is dan het begin. Om te analyseren of deze neiging universeel is, heb ik een systematische methode ontwikkeld (hoofdstuk 4) en geïmplementeerd in verscorpora van vijf verschillende talen. Hoewel dit een kleine steekproef is zijn de resultaten robuust, met minder uitzonderingen later in de versregel. In hoofdstuk 5 presenteer ik een reeks experimenten om te toetsen of deze opwaartse regelmaat gerelateerd is aan aandacht. De deelnemers luisteren naar reeks drumslagen; van tijd tot tijd wordt er een verschillende slag gespeeld, en dan moeten de deelnemers zo snel mogelijk op een knop drukken. Ze identificeren deze uitzonderingen sneller als ze later in de volgorde verschijnen, vergelijkbaar met de patroon die dichters tonen in versregels. Het is daarom mogelijk dat onze aandacht zichzelf optimaliseert naarmate we de reeks geluiden verwerken, en dat zou de resultaten van de corpora en de experimenten kunnen verklaren.

In het laatste deel heb ik onderzocht hoe melodieën en woorden bij elkaar aansluiten. Bij het gebruik van een taal zijn sommige lettergrepen prominenter, hoger of langer dan andere; net als bij de melodieën van liedjes. In dit licht bezien, kunnen Engelstaligen een geaccentueerde lettergreep gepaard met een zwakke noot beschouwen als ongepast. Veel van dergelijke hypotheses zijn voorgesteld voor verschillende talen, maar we weten niet of die regels eigenaardigheden van elke taal zijn, of dat er generalisaties te vinden zijn over de hele wereld. Om dergelijke regels systematisch te ontdekken, heb ik een computationele methode ontwikkeld, en ter illustratie heb ik een groot Nederlands corpus (met meer dan 3000 liedjes) geanalyseerd (hoofdstuk 6). Tot slot, om te bevestigen dat de methode geschikt is, heb ik een perceptueel experiment uitgevoerd met tientallen deelnemers (hoofdstuk 7). Volgens de analyse van het corpus en het experiment, geven Nederlandstaligen de voorkeur aan een overeenkomst tussen het ritme van spraak en muziek. Daarenboven, als dat niet het geval is, hebben ze een voorkeur voor trocheïsche woorden gepaard met een jambisch muzikaal ritme (*lé-vert*), boven jambische woorden gepaard met een trocheïsch ritme (*be-stélt*).

Zoals ik in deze drie delen van het proefschrift beargumenteer, is het essentieel om de algemene kenmerken van de verstradities van de wereld systematisch te beschrijven. De methoden gepresenteerd in de even hoofdstukken (2, 4, 6) kunnen dit veeleisend werk verlichten. Op deze manier kunnen we, als we wijdverspreide kenmerken ontdekken, verklarende hypothesen testen. De oneven hoofdstukken (3, 5, 7) bieden een aantal dergelijke voorbeelden, met alleen Nederlandstalige deelnemers. Door soortgelijke experimenten in verschillende talen uit te voeren kunnen we, mogelijk, universele tendensen aan het licht brengen. Kortom, versbouw wordt, hoewel het een prototype is van onbelemmerde creativiteit, steeds beperkt door het menselijk cognitief systeem.

Laburpena

Tesi hau bertsogintzari buruzkoa da. Batetik, munduan zehar oso zabalduta dauden hainbat ezaugarri aztertu ditut; bestetik, ezaugarri horien ohikotasuna azaldu lezaketen hipotesi kognitiboak garatu ditut.

Zer da bertsoa? Ikuspegi zabala hartuz, hitzari lotutako hainbat jardun sar daitezke bertsoaren definizioan: abestiak, olerkiak, zotz egiteko formulak, bestelako jolas-kantak, manifestazioetako oihuak. Eguneroko hizketarekin alderatuz, egituraketa maila gehigarriak dauzkate jardun horiek, esate baterako, silaba azentudunen edo taupada baten erregulartasuna, errepikatzen den doinu bat, edo silaba kopuru zehatz bat esaldi bakoitzeko. Ezaugarri horiek, batera hartuz gero, eredu, molde edo egitura mental bat eskaintzen dutela uler daiteke; ereduoi hitzak ezarriz bertsoak osatzen dira.

Bertsogintza munduko hizkuntza-komunitate guztietan ematen da, hau da, eguneroko hizketa bezalaxe, giza-kulturen ezaugarri unibertsala da. Apalago esanda, orain arte behintzat ez da bertso-bako hizkuntzarik aurkitu. Hori, bere horretan, misteriotsua da, ez baitakigu zein oinarrizko funtzio bete lezakeen bertsoak ebolutiboki, edo zerk bultzatu duen komunitate oro bertsoa sortzera. Munduan 6000 bat hizkuntza daude, eta abestiak eta olerkiak sortzeko era oso ezberdinak garatu dituzte. Komunitate bakoitzak belaunaldiz belaunaldi berrikuntzak ezartzen ditu, eta, hala eta guztiz ere, badirudi antzekotasunak ere badauzkatela. Bigarren misterio hori tesi honen abiapuntua da.

Zergatik ote dago horren zabaldua errima edo binakako erritmoak? Zergatik ez ote ditugu berrogei silabatako bertso-lerroak sortzen? Faktore ugarik baldintzatzen dute bertsoen ekoizpena: gure gorputzen anatomiak (burmuinaren arkitekturak, arnasaren mugak), ingurune sozialaren dinamikek (eguneroko lanak eta aisiak, herrien bilakaera historikoak), eta baita, beharbada, ingurune fisikoak ere (apalago abesten al dute toki hotzetan?). Azalpen potentzial horien itsaso zabalean, hipotesi jakin bat garatu dut tesian, alegia, gizakiok antzekotasun kognitiboak dauzkagula eta horiek antzekotasunak sortzen dituztela munduko bertsogintzetan.

Ikerketak hiru atal ditu, bakoitzak bertsoen ezaugarri bat aztertzen duelarik. Atalek bina kapitulu dituzte, aurrenekoa hizkuntza ezberdinetako bertso-

korpusetan oinarrituta, eta bigarrena pertzepzio-esperimentuetan oinarrituta.

Lehenengo atala bertso-moldeen osagaiei buruzkoa da. Bertso berri bat sortu nahi izanez gero *Haizeak bidali du* doinua erabiliz (zortziko txikian, alegia), 52 silabatako testua osatu behar da. Dena dela, jakin badakigu 52 silaba horiek ez ditugula multzo bakar batean ulertzen, osagai eta azpi-osagaietan antolatuta baizik. Bertsoa lau puntuk osatzen dute, puntu bakoitzak bi lerro dauzka, lerrook zazpina eta seina silaba dauzkate, eta, azkenik, taupada musikalek banaka edo binaka antolatzen dituzte silabak. Kontinente guztietako hamarnaka bertso-tradizio aztertuz (2. kapitulua) bi ondorio nagusi atera ditut. (1) Bertso-molde guztiek osagai eta azpi-osagaiak dauzkate euren baitan, ezaugarri estrukturalak aztertuz aurki daitezkeenak (adib. puntuen arteko errima edo bukaerako silaben luzapena). (2) Maiz, bertso-moldearen osagai bat errepikaturik ematen da (adib. errima eta doinu berdina duten bi lerro), baina ezaugarri berdinak dauzkaten osagaiak ez dira sekula lau bider baino gehiago ematen jarraian. Litekeena da munduko tradizioetan aurkitzen den muga hori ezaugarri kognitiboen ondorio izatea, izan ere, hiruzpalau elementutik gora kosta egiten zaigu kopuruaz berehala jabetzea edo oroitzea. Hirugarren kapituluan, horri lotutako esperimentu bat aurkezten dut. Bertan, esanahirik gabeko silabak entzun eta errepikarazi egiten zaizkie partaideei. Silaba sekuentzia luzeak entzuten dituztenean (adib. 12 silaba) multzoka antolatzeko joera erakusten dute, bertso-moldeetan bezalaxe.

Bigarren atala bertsogileen ‘akatsak’ aztertu ditut. Hizkuntza askotan (adib. ingelesez, amazigeraz) bertso moldeek silaben erritmo jakin bat eskatzen dute, silaba azentudunak eta azentugabeak tartekatuz, esaterako. Hala ere, bertsogileek ez dituzte beti-beti arau horiek betetzen, eta, hainbat kasutan, bertso-lerroen bukaera erregularragoa dela deskribatu izan dute ikerlariek. Joera hori unibertsala ote den aztertzeko, metodo sistematiko bat garatu dut (4. kapitulua), eta baita bost hizkuntzetako korpusetan inplementatu ere. Lagin kopurua xumea izan arren, emaitzak sendoak dira, salbuespen gutxiago aurkitzen direlarik bertso-lerroak aurrera egin ahala. Bostgarren kapituluan, esperimentu sorta bat aurkezten dut ea goranzko erregulartasun hori arretarekin lotuta ote dagoen frogatzeko. Partaideek danbor-hotsez egindako sekuentziak entzuten dituzte; horietako batzuek bestelako hots bat daukate, eta berau identifikatu bezain azkar botoi bat sakatu behar dute. Bertso-lerroetako joeraren antzera, errazago identifikatzen dituzte salbuespenak sekuentzian zenbat eta aurrerago egon. Baliteke, beraz, gure arreta optimizatzea hots segidak prozesatu ahala, esperimentuetako eta korpusetako emaitzak azalduz.

Azken atalean, doinuak eta hitzak zelan uztartzen diren ikertu dut. Edozein hizkuntza erabiltzerakoan, silaba batzuk nabarmenagoak, altuagoak edo luzeagoak izango dira beste batzuk baino; abestien doinuetako notekin gertatzen den bezalaxe. Hori hala izanik, ingelesezko hiztunek, adibidez, txartzat jo lezakete silaba azentudun bat nota ahul batekin uztartzea. Horrelako hipotesi ugari proposatu izan dira hainbat hizkuntzarentzat, baina ez dakigu arauak hizkuntzen berezitasunak ote diren, ala orokortasunak ote dauden munduan zehar. Horrelako arauak era sistematiko batean aurkitzeko, metodo konputazional bat garatu dut eta, ilustrazio gisa, nederlanderazko korpus handi bat (hiru mila abestitik gora) aztertu dut (6. kapitulua). Azkenik, metodoa egokia dela berresteko, pertzepzio esperimentu bat burutu dut nederlanderazko hiztunekin (7. kapitulua).

Tesiaren hiru atal hauetan argudiatzen dudanez, munduko bertsogintzen ezaugarri orokorrak sistematikoki deskribatzea funtsezkoa da. Kapitulu bikoitietan (2, 4, 6) aurkeztutako metodoek lan mardul hori arin lezakete. Bestetik, hedadura zabaleko ezaugarriak aurkitu ahala, zergatiak argitzeko hipotesiak frogatu ahal izango ditugu. Kapitulu bakoitiek (3, 5, 7) hainbat adibide eskaintzen dituzte, soilik nederlanderazko hiztunekin ordea. Antzeko esperimentuak hizkuntza ezberdinetan burutuz, beharbada, unibertsalak izan litezkeen joerak azaleratuko ditugu. Azken batean, bertsogintza, sormen askearen fruitu prototipikoa izan arren, gizakion sistema kognitiboaren eremuan sortzen da.

Curriculum Vitae

Varuṇ deCastro-Arrazola was born in Canberra in 1988. His preparatory instruction for higher education was carried out at the Zubiri Manteo High School (Donostia, 2006). He then completed a degree in Ethnomusicology at the Escola Superior de Música de Catalunya (Barcelona, 2012), and a degree in Language Sciences at Université Paris 8 (2012). From 2013 to 2017 he has been employed by Leiden University in order to perform doctoral research within the project *Knowledge and culture*, on the role of core knowledge systems in shaping the versification systems of the world. This work has been conducted primarily at the Meertens Instituut (Amsterdam), under the supervision of Marc van Oostendorp and Johan Rooryck, with a visiting period at the Australian National University (Canberra). While working on his thesis, he has performed teaching duties at Leiden University on phonetics, phonology and syntax.