Available online at www.sciencedirect.com

ScienceDirect

Lingua 171 (2016) 74–91

Lingua

www.elsevier.com/locate/lingua

Is speech processing influenced by abstract or detailed phonotactic representations? The case of the Obligatory Contour Principle

Natalie Boll-Avetisyan [a,b,*], René Kager [a]

[a] Utrecht University, Trans 10, 3512 BL Utrecht, The Netherlands
[b] University of Potsdam, Department of Linguistics, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

Received 25 October 2014; received in revised form 7 November 2015; accepted 27 November 2015
Available online 14 January 2016

Abstract

Many languages restrict their lexicons by OCP-Place, a phonotactic constraint against co-occurrences of consonants with shared [place] (e.g., McCarthy, 1986). While many previous studies have suggested that listeners have knowledge of OCP-Place and use this for speech processing, it is less clear whether they make reference to an abstract representation of this constraint. In Dutch, OCP-Place gradiently restricts non-adjacent consonant co-occurrences in the lexicon. Focusing on labial-vowel-labial co-occurrences, we found that there are, however, exceptions from the general effect of OCP-Labial: (A) co-occurrences of identical labials are systematically less restricted than co-occurrences of homorganic labials, and (B) some specific pairs (e.g., /pVp/, /bVv/) occur more often than expected. Setting out to study whether exceptions such as (A) and (B) had an effect on processing, the current study presents an artificial language learning experiment and a reanalysis of Boll-Avetisyan and Kager's (2014) speech segmentation data. Results indicate that Dutch listeners can use both knowledge of phonotactic detail and an abstract constraint OCP-Labial as a cue for speech segmentation. We suggest that whether detailed or abstract representations are drawn on depends on the complexity of processing demands.

Keywords: OCP; Artificial language learning; Speech segmentation; Phonotactics; Identity; Phonotactic probability

1. Introduction

Languages' lexicons are restricted by phonological constraints that determine which adjacent and non-adjacent within-word combinations of sounds are likely, unlikely or even illegal. There is ample evidence that humans have knowledge of phonotactic probabilities in their language input, which influences their wellformedness intuitions about nonwords (e.g., Bailey and Hahn, 2001; Frisch et al., 2000; Scholes, 1966) and lexical access (e.g., Berent et al., 2001b; Vitevitch and Luce, 1998, 1999). Moreover, listeners can use phonotactic cues for speech segmentation (e.g., Boll-Avetisyan and Kager, 2014; McQueen, 1998; Suomi et al., 1997; Vroomen et al., 1998) and benefit from phonotactic knowledge when learning new words (e.g., Boll-Avetisyan, 2012a; Gathercole et al., 1999; Storkel, 2001). In spite of this vast interest in the

* Corresponding author at: University of Potsdam, Department of Linguistics, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany. Tel.: +49 0331 9772374.
E-mail addresses: nboll@uni-potsdam.de (N. Boll-Avetisyan), r.w.j.kager@uu.nl (R. Kager).

http://dx.doi.org/10.1016/j.lingua.2015.11.008
0024-3841/

use of phonotactic knowledge in speech processing and language acquisition in the literature, much less is known of how phonotactic knowledge is represented.

The current study sets out to contribute to the discussion of how exactly phonotactic knowledge is represented in the speakers' minds. In principle, there are three possibilities that have been suggested in the literature: first, phonotactic representations could be detailed. That is, listeners might represent each possible phoneme combination associated with knowledge of its frequency or probability of occurrence — a view classically taken by connectionists (e.g., Plaut et al., 1996). Second, listeners might represent phonotactics by means of abstract constraints in the sense that they abstract over phonological features. For example, if co-occurrences of consonants that are part of a natural class (e.g., all labial consonants) are systematically unattested in a language, listeners would only represent a single abstract constraint, namely *[labial][labial]. This abstract representation would be more economic rather than having many separate specific representations such as */pb/, *pm/, */fm/ et cetera, as it subsumes all specific instances enumerated. This latter view is classically taken by generative phonologists (e.g., Chomsky and Halle, 1968). A third possibility is that knowledge is both abstract and detailed (e.g., Adriaans and Kager, 2010; Albright and Hayes, 2003; Hayes, 1999; Hayes and Wilson, 2008).

The current study investigates whether abstract or detailed knowledge of phonotactics influences speech segmentation. The abstract constraint under test is OCP-Labial, a constraint restricting the co-occurrence of consonants sharing the feature [labial]. In Dutch, effects of OCP-Labial are not absolutely systematic. We focus on two issues: first, we will investigate whether Dutch listeners attribute a special role to co-occurrences of identical labials. Second, we will investigate whether specific pairs of labials that are likely to co-occur in a language, and hence are exceptions from OCP-Labial, will still fall under the effect of the constraint.

1.1. OCP-Place

The *Obligatory Contour Principle* (OCP) is a typologically well-attested constraint on within-word phonological structure requiring co-occurring segments to be featurally non-identical (i.e., to disharmonize). Originally, the OCP was formulated to account for similarity avoidance between tones in West-African languages (Leben, 1980). Later, the OCP was extended to OCP-Place to formally account for co-occurrence constraints against consonants with shared [place] (McCarthy, 1986), as, for example, in Arabic (Greenberg, 1950), where verbs (e.g., *katab-a* 'he wrote', *kutib-a* 'it was written', and *kuttib-a* 'he was made to write') are derived from root morphemes consisting of three consonants (e.g., /k t b/), which, with very rare exceptions, are not homorganic (e.g., */f b m/), or, if root-initially, identical (e.g., */b b k/).

Effects of OCP-Place have first been found in the lexicons of the Semitic language family (e.g., Frisch et al., 2004; Greenberg, 1950; McCarthy, 1985), and were later discovered in many genetically and geographically unrelated languages (e.g., English: Berkley, 1994; Muna: Coetzee and Pater, 2008; Niger-Congo languages: Pozdniakov and Segerer, 2007; Dutch: Kager and Shatzman, 2007). In fact, recent large-scale quantitative typological studies (Graff, 2012; Mayer et al., 2010; Pozdniakov and Segerer, 2007) indicate that virtually every language is restricted by similar place avoidance (c.f. Boll-Avetisyan, 2012b; Graff, 2012: OCP-effects could not be found for the languages Mandarin Chinese, Guarani, Wantoat and Tatar). In light of this very strong typological preference, OCP-Place has been proposed to be a (statistical) universal (e.g., Frisch, 2004; McCarthy, 1986; Pozdniakov and Segerer, 2007). Classically, it is assumed that OCP-Place is a constraint that abstracts over features, i.e., it is represented as *[place][place] (McCarthy, 1986). If sub-classes of consonants with shared place are affected to different degrees, then OCP-Place can be further broken down into constraints such as OCP-Labial (i.e., *[labial][labial]) or OCP-Dorsal (i.e., *[dorsal][dorsal]) (Coetzee and Pater, 2008; McCarthy, 1988).

The majority of languages does not display categorical effects of OCP-Place targeting all homorganic consonant combinations. Rather, most languages display gradient effects of this constraint. This means that pairs of consonants with shared place features can be found in the lexicon; however, they are systematically underattested. That is, pairs such as /pVm/ in English *spam* or /mVb/ in *mob* do occur, albeit significantly less often than expected if non-adjacent consonants co-occurred at random. Frisch et al. (2004) observed that even in the case of Arabic, OCP-effects are graded: forms with roots with a higher overlap in featural identity are less frequent than forms in which roots are less similar. Consequently, they propose a single constraint of similarity avoidance that is quantitatively sensitive to the degree of violation as defined by phonetic feature distance. For Arabic, this is an elegant way to account for the co-occurrence dependencies between consonants that captures both the gradient and abstract aspects of OCP-Place in a single analysis. Other studies, however, have raised doubts that Frisch, Pierrehumbert and Broe's model can be applied to many other languages (e.g., Berent and Shimron, 2003; Coetzee and Pater, 2008; Graff and Jaeger, 2009). Coetzee and Pater, for example, used a model of weighted constraints referring to more specific feature-based sub-classes of consonants (as inspired by, a.o., Padgett, 1995). This approach was more successful in modeling gradient consonant co-occurrence restrictions in Arabic and Muna than Frisch et al.'s similarity metric (Coetzee and Pater, 2008).

In the current study, we address the question of whether the assumption of an abstract representation of OCP-Labial is justified, if the effects of the constraint are gradient. The classical arguments for abstractness are that productivity can best