One phonotactic restriction for speaking, listening and reading: The case of the *no geminate* constraint in German

Silke Hamann, University of Amsterdam

Abstract: This article illustrates the cross-modal application of the phonotactic *no geminate* constraint that disallows geminate consonants within a prosodic word in German. In phonological production, forms like |ha**t**+**t**ə| 'have-3SG.PST' surface as /ha**t**ə/ (instead of */ha**tː**ə/) due to this constraint. In speech perception, phonetic forms with geminates like [bʁoː**tː**aɪk] 'bread dough' are perceived as consisting of two prosodic words because of this constraint. And in the reading process, orthographic forms with double consonantal graphemes like <Wa**ll**> and <Te**ll**er> are read as /wa**l**/ and /tɛ**l**ɐ/, not */wa**lː**/ and */tɛ**lː**ɐ/ due to this constraint. I show that this observation can be formalized in Optimality Theory by applying the same phonotactic constraint in phonological production, in the perception grammar (Boersma 2007) and in the reading grammar (Hamann & Colombo 2017). In both the perception and the reading grammar, this constraint interacts with constraints that handle the arbitrary mapping between the sensory input (auditory or written) and a surface phonological form. This approach is shown to be preferable to previous analyses of reading double consonantal graphemes in German, which either lack an explicit formalization of the phonological knowledge or reduplicate this in their phoneme-to-grapheme mappings.

1 Introduction

The standard variety of New High German (henceforth: German) has no geminates in its phoneme system, i.e. it does not distinguish consonants by their duration only as e.g. Italian does. Due to inflection, geminates could occur within prosodic words (also called phonological words; henceforth: p-words) but are avoided via a phonological degemination process (Wiese 2000: 41, 230), see the examples in (1).[1]

(1)	*Morphemes*	*realization*	*meaning*
	hat + tə	[hatə]	have 3SG.PST
	lad + t	[lɛːt]	load 3SG.PRS
	ʁaɪs + st	[ʁaɪst]	tear 2SG.PRS
	leːz + st	[liːst]	read 2SG.PRS
	zɪt͡s + st	[zɪt͡st]	sit 2SG.PRS
	t͡sa͡ɪçən + n	[t͡sa͡ɪçən]	sign DAT.PL

Derived, so-called fake geminates can occur across p-word boundaries in German. Degemination is optional in this context; see the examples in (2), but is more likely in

[1] The umlaut and ablaut processes occuring in some of these examples are not further discussed in this article. The interested reader is referred to Wiese (2000), who also discusses the degemination of only the second part of the affricate /t͡s/, see the second to last example in (1).

fast speech (Wiese 2000: 231) and when the two syllables containing the fake geminate are unstressed (Kohler 2001).

(2) *Morphemes*	*realizations*	*meaning*
a) ʃʁɪft + tuːm	[ʃʁɪftːuːm]~[ʃʁɪftˑuːm]~[ʃʁɪftuːm]	‘professional literature’
ʃɪf + faːʁt	[ʃɪfːaɐt]~[ʃɪfˑaɐt]~[ʃɪfaɐt]	‘shipping’
bʁoːt + taɪg	[bʁoːtːaɪk]~[bʁoːtˑaɪk]~[bʁoːtaɪk]	‘bread dough’
b) ɪl + legaːl	[ɪlːegaːl]~[ɪlˑegaːl]~[ɪlegaːl]	‘illegal’
an + neːmən	[anːeːmən]~[anˑeːmən]~[aneːmən]	‘to assume’
ʊm + moːdəln	[ʊmːoːdəln]~[ʊmˑoːdəln]~[ʊmoːdəln]	‘ro remodel’

This second degemination process is considered in this article to be phonetic, due to its different context, optional application and speech-rate dependence (see Hamann 2016 for a detailed discussion and a comparison to similar phonological and phonetic degemination processes in Dutch). Such a strict distinction between phonetic and phonological processes is only possible in a modular generative grammar theory (following Chomsky & Halle 1968) with a phonological module that maps an underlying form onto a phonological surface form and that is different from the phonetic module, which maps the phonological surface form onto a phonetic form.

While traditional generative models, whether rule-based or Optimality-theoretic (Prince and Smolensky 1993 [2004]; henceforth: OT), restrict themselves to the description of the production process (starting with a lexical form and ending in a phonetic output), the present article includes the perception/comprehension process (starting with a phonetic form, the auditory input, and ending in a lexical form), following the model of *Bidirectional Phonetics and Phonology* by Boersma (henceforth: BiPhon; within OT: Boersma 2007; within neural networks: Boersma, Benders & Seinhorst 2018).

In the process of speech perception, the knowledge that fake geminates are only possible across p-word boundaries will be shown to guide the German listener in parsing phonetic forms like [ʃʁɪftːuːm] into two separate p-words, and forms like [hatə] into one.[2] The application of the same *no geminate* restriction both in phonological production and in speech perception is formalized within BiPhon-OT. The *no geminate* restriction is further shown to apply in the reading process, formalized in Hamann & Colombo’s (2017) *reading grammar*, thereby forming the first cross-modal application of the same structural constraint in perception, production and reading.

The application of *no geminate* in reading is illustrated with so-called orthographic ‘sharpening’ (German ‘Schärfung’, e.g. Maas 1992; Neef 2002): German uses graphemes consisting of two identical consonantal letters to indicate the shortness and usually laxness[3] of the preceding stressed vowel within p-words. This is necessary

2 There are, of course, further cues such as vowel duration and intensity (indicating stress) that guide the parsing process but these will not be dealt with in the present article.

3 Exceptions to the correlation of long with tense in German are the low short vowel /a/, that only differs from /aː/ in length, and the long vowel /ɛː/, that only differs from /ɛ/ in length.

as single vowel letters can be used in German to stand for a short or a long vowel, e.g. <a> can express /a/ and /aː/, and <i> both /ɪ/ and /iː/. Examples of 'sharpened' words are given in (3), their long counterparts in (4).

(3) *Orthography* *realization* *meaning*

<Ratte>	[ʁatə]	'rat'
<dann>	[dan]	'then'
<Wall>	[wal]	'rampart'
<wirr>	[wɪɐ]	'confused'
<offen>	[ɔfən]	'open'

(4) *Orthography* *realization* *meaning*

<Rate>	[ʁaːtə]	'rate'
<kam>	[kaːm]	'come' 1/3SG.PST
<Wal>	[waːl]	'whale'
<wir>	[wiːɐ]	'we'
<Ofen>	[oːfən]	'oven'

In the reading process, the double consonant grapheme in each word in (3), all of them single p-words, cannot be interpreted as a fake geminate, because of the *no geminate* constraint. Together with other phonological restrictions, this constraint thus restrains the German-specific mappings from graphemes to phonemes (just as it restricts phonological production and speech perception).

This article is structured as follows. Section briefly 2 discusses earlier accounts of the use of double consonantal graphemes in German. Section 3 provides the present account with a cross-modal *no geminate* constraint. It illustrates how this applies in phonological production and perception of German words with single and double consonants, and in the reading of German words written with double consonantal graphemes. Section 4 provides a conclusion.

2 Earlier accounts of double consonantal graphemes in German

The use of double consonantal graphemes, henceforth: <$\beta_i\beta_i$>, in German has been the topic of many studies. Much of the discussion centered on the question whether <$\beta_i\beta_i$> expresses the quality and quantity of the preceding vowel (see e.g. Wiese 1987 and Ramers 1999), or the ambisyllabicity of the consonant itself (see e.g. Eisenberg 1999; Sternefeld 2000). The second interpretation neglects the fact that sharpening is also employed in word-final position, e.g. in the monosyllabic words in (3) such as <dann> [dan] 'then', where <$\beta_i\beta_i$> cannot represent an ambisyllabic consonant, and requires additional machinery to account for such cases.

Most of the previous studies on sharpening focus on the *writing* direction, and if providing a formalization, employ grapheme-phoneme correspondence rules as introduced by Bierwisch (1972). These rules are analogous to generative phonological rules (Chomsky & Halle 1968), and often include phonological contexts, thereby duplicating phonological knowledge.

Wiese (2004) provides the first OT account of German writing. In his analysis, underlying phonological input is mapped onto a written form. Despite their different nature, the mapping between these two forms is evaluated by correspondence constraints such as MAX, DEP and IDENT, which in OT phonology are used to compare two abstract phonological forms (Prince and Smolensky 1993 [2004]). The exact workings of these constraints in Wiese's account is therefore unclear. Though his account is restricted to writing, Wiese correctly points out that the "bidirectional nature of correspondence relations [in OT] allows for constraints looking into both directions" (p. 316).

Neef (2002, 2012) describes and partially formalizes the reading process, i.e. the transformation of a written form into a 'surface form'[4] (calling this process 'recoding', as is tradition in psycholinguistic literature). Since his account includes sharpening, it is described here in more detail. Neef uses correspondence rules such as <m>→[m] and additional graphematic constraints that capture general properties of the writing system. All outputs generated by correspondence rules and graphematic constraints are checked for their phonological well-formedness via a phonological filter. The two graphematic constraints from Neef (2012: 211) responsible for orthographic sharpening are given in (5).

(5) a) In a sequence of identical letters, all non-initial ones may be recoded as zero.
b) A vowel letter does not correspond to a tense vowel if it is immediately followed by […] a sequence of identical consonantal letters.

For the written input <Lamm> 'lamb', for instance, the correspondence rules provides the forms [la(ː)mm], and the graphematic constraint (5a) the additional alternative outputs [la(ː)m]. Constraint (5b) restricts this set to [lam] and [lamm], and the phonological filter discards the ungrammatical output [lamm], leaving [lam] as the output of this recoding process. Neef (2012) claims that it is not necessary to assume a particular ordering in the application of these constraints and rules, and that the "phonological filter applies every time when necessary" (p. 222). An unordered application of constraints that can enlarge (as in (5a)) or restrict (as in (5b)) the candidate set, and the multiple application of phonological restrictions leave this analysis rather unrestricted. Furthermore, the phonological filter is not made explicit, nor is its continuous influence on the recoding process.

Another drawback of Neef's proposal is that for lexically stored irregular orthographic forms, which "have priority over the set of regular correspondences" (2002: 172), it is not elaborated how such lexical knowledge can interfere with the correspondence relations that bypass the lexicon.

We will see in the following how an explicit formalization of German phonotactic restrictions and the phonological process of degemination both in speech production and recognition make a full-fledged model of the reading process (including lexical access) possible which avoid the shortcomings of Neef's proposal.

[4] Neef makes no distinction between phonological surface and phonetic form.

3 *No geminate* constraint in production, perception and reading

The restriction that geminates are only allowed across p-words in German, exemplified in (1) and (2), is captured in the present proposal with the OT constraint in (6).

(6) *GEM: Assign a violation mark to every geminate that is not spanning a lower p-word boundary.

Constraint (6) prohibits two identical adjacent elements within a p-word, thus follows from the *Obligatory Contour Principle* (OCP; McCarthy 1986), and is based on Rose's (2000) *no geminate* constraint. While Rose's constraint banned all long consonants, the constraint in (6) is restricted to true geminates, i.e. those that do not span p-word boundaries. It is further restricted to geminates spanning lower p-word boundaries, and does not apply to recursive p-words of compounds (Itô & Mester 2008) or of certain prefixed verbs illustrated in (2b) (see Raffelsiefen 2000 for a discussion).

This constraint is proposed to apply cross-modally: in phonological production and speech perception for spoken language, and in reading for written language. For the formalization thereof, I employ the BiPhon model by Boersma (2007), as shown in Figure 1. BiPhon is a linguistic model that provides an explicit formalization of phonology and phonetics in both processing directions, i.e. production and comprehension.

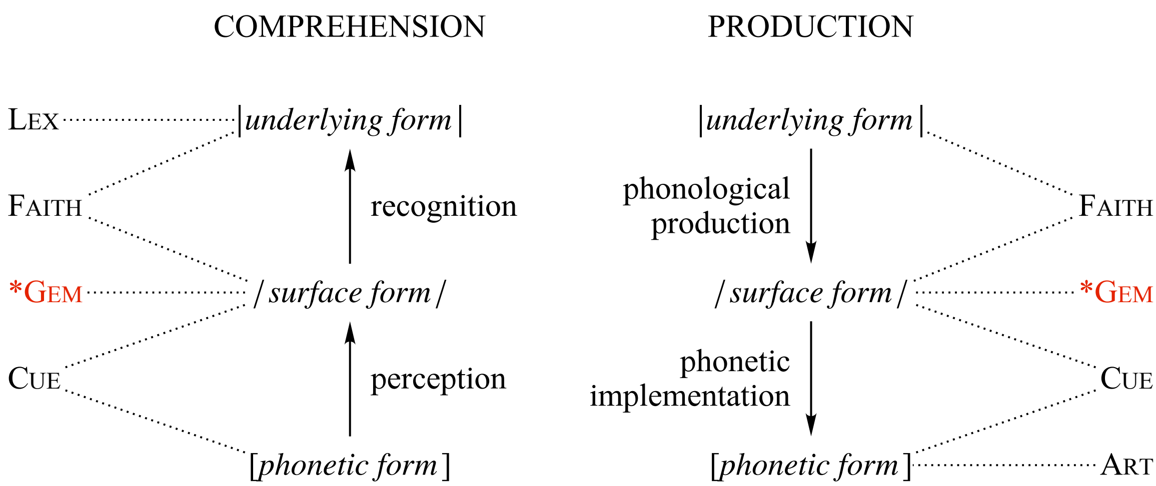

Fig. 1. The BiPhon grammar (Boersma 2007) with the structural constraint *GEM in red restricting the phonological surface form, thus applying both to the output of perception (lower left) and the output of phonological production (upper right).

The right side of Figure 1 shows the production direction, starting from an underlying form, which is mapped onto a surface form in phonological production (traditionally called 'phonology') via faithfulness and structural constraints. In phonetic implementation, the surface form is then transformed into a phonetic form via cue and articulatory constraints. On the left side of Figure 1, we see the comprehension process: there, the starting point is the phonetic form (lower left), which is mapped onto a surface form. This is performed by the so-called 'perception grammar' (Boersma 2007, Boersma & Hamann 2009). The surface form is then mapped onto an underlying, lexical form in speech recognition.[5]

The same constraints that are employed in the production direction also apply in the comprehension direction. In both processing directions, structural constraints like *GEM restrict the phonological surface form. In production, they apply to the output of the phonological production, where they interact with faithfulness constraints. This is illustrated in section 3.1 below with phonological degemination in German. In listening, they restrict the output of the perception process. Here they interact with cue constraints, as illustrated in section 3.2 for the perception of auditory forms with long and short consonants in German.

To account for the reading process, Hamann & Colombo (2017) introduced a *reading grammar*, which works very similar to the BiPhon perception grammar. Both involve the mapping of a sensory input onto a phonological surface form. In speech perception, the sensory input is the auditory form, in reading the written form. Instead of cue constraints, a reading grammar uses orthographic constraints (ORTH), mapping graphemes onto phonemes. The output of both perception and reading grammar is restricted by the same structural constraints, e.g. *GEM. This reading grammar is represented in Figure 2.

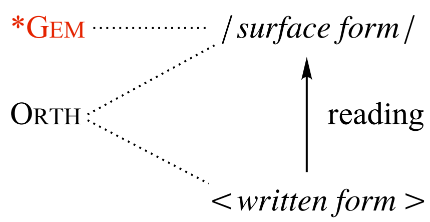

Fig. 2. Reading grammar (Hamann & Colombo 2017): mapping of a written form onto a phonological surface form via ORTH(OGRAPHIC) constraints, which interact with structural restrictions on the surface form, here represented by *GEM.

The surface form that results from this mapping undergoes the same recognition process as the surface form resulting from the perception process (left upper part of Fig. 1), and is thus influenced by the existence or non-existence of a corresponding underlying form in the lexicon (cf. FN 5). In section 3.3 below the workings of the reading grammar is illustrated with the case of orthographic sharpening in German.

Note that the reading grammar as depicted in Fig. 2 describes silent reading. The process of reading aloud needs the additional component of mapping the surface form onto a phonetic form, i.e. the phonetic implementation part of the BiPhon grammar in Fig. 1.

---

[5] Though production and comprehension are described in the present article as each consisting of two serial steps, these two steps can be performed in parallel. Parallel evaluation of surface and underlying form in comprehension, for instance, is necessary to account for the Ganong effect, i.e. the influence of lexical information on speech perception (Ganong 1980), see the formalization in Boersma (2012).

3.1. Phonological production and phonetic implementation

In phonological production, the obligatory phonological process of degemination applies whenever two identical consonants are adjacent to each other within a p-word as a result of morphophonological concatenation. Phonological degemination is illustrated in the following with the two words in (7). In line with the conventions of the BiPhon model, underlying forms are given in straight lines, surface forms in slashes, and phonetic forms in square brackets. Transcriptions of the surface form include brackets for lower p-word boundaries and dots for syllable boundaries. Consonants between a stressed short, lax vowel and an unstressed vowel (i.e. in foot-internal position) as in (7a) are assumed to be ambisyllabic, following e.g. Becker (1996) and Wiese (2000).

(7) a) |hat+tə| /(haṭə)/ 'had'
b) |ʃʁɪft+tuːm| /(ʃʁɪft)(tuːm)/ 'professional literature'

In (7a) degemination applies to avoid two identical consonants adjacent to each other within a p-word, due to the structural constraint *GEM. In (7b), on the other hand, the two identical consonants are separated by a lower p-word boundary, *GEM is satisfied, and degemination is blocked.

For a complete modeling of phonological degemination, the structural constraint *GEM is not sufficient. We also need a restriction on forming incorrect p-words, because their boundaries could potentially block degemination. For this, we employ Hall's (1999: 114) *full vowel constraint* (FVC), given in (8).

(8) *FVC: Assign a violation mark to every p-word that does not contain at least one full vowel.

The high ranking of this constraint in German avoids that e.g. for the input form |hat+tə| in (7a) the output candidate */(hat)(tə)/ with two separate p-words and a fake geminate wins, see the first candidate in tableau (9). The actual occurring degeminated form, candidate two, violates the faithfulness constraint DEP-C since one underlying /t/ does not surface. Candidate three retains the double consonant, but violates the high-ranked structural constraint *GEM. Please note that the assignment of syllable boundaries and stress are not formalized in this and the following tableaux as they would go beyond the scope of the present paper.[6]

[6] I follow Wiese (2000) in assuming default stress is penultimate and assigned via a right-aligned trochaic foot; cases of final or antepenultimate stress are then irregular and lexically stored. For an alternative account where German is assumed to be quantity-sensitive, see e.g. Féry (1998).

(9) *Phonological production of* |hat+tə|

|hat+tə|	*GEM	FVC	DEP-C
/(hat)(tə)/		*!	
☞ /(hat̩ə)/			*
/(hat.tə)/	*!		

The non-application of degemination across p-word boundaries is formalized for the example (7b) in tableau (10). The same ranking of structural and faithfulness constraints as in (9) applies.

(10) *Phonological production of* |ʃʁɪft+tuːm|

|ʃʁɪft+tuːm|	*GEM	FVC	DEP-C
☞ /(ʃʁɪft)(tuːm)/			
/(ʃʁɪft̩uːm)/			*!
/(ʃʁɪft.tuːm)/	*!		

In this case, candidate one with two p-words and a fake geminate wins, because two separate p-words with two full vowels can be formed from the input (without violating FVC).

As we saw in the overview of the BiPhon model above, phonological production is complemented by phonetic implementation, where discrete phonological surface forms are mapped onto a continuous, non-discrete auditory form consisting of several cues. This mapping is handled by cue constraints (Escudero & Boersma 2004). Two cue constraints relevant for the present data are given in (11).

(11) */tt/[t]: Assign a violation mark for every two alveolar plosives in the surface form that are mapped onto a short closure phase in the phonetic form.
*/t/[tː]: Assign a violation mark for every single alveolar plosive in the surface form that is mapped onto a long closure phase in the phonetic form.

A set of cue constraints all referring to the same surface form, e.g. /tt/, and the same auditory cue dimension, such as closure duration, is inherently ranked according to how well the values on the dimension cue the surface form. This is illustrated in (12) with two sets of constraints, one referring to double consonants in the surface form (12a), the other to single consonants (12b), both dealing with closure duration. Note that the continuous auditory dimension of closure duration is restricted here to three realizations, short [t], middle [tˑ], and long [tː].

(12) a) */tt/[t] >> */tt/[tˑ] >> */tt/[tː]
b) */t/[tː] >> */t/[tˑ] >> */t/[t]

As evidence to the contrary is lacking, I assume that the two hierarchies in (12) with three strata each can be collapsed into one hierarchy with three strata.

In phonetic implementation, the process of phonetic degemination across p-word boundaries takes place, cf. example (13) with its output forms that depend on speech style and rate.

(13) /(ʃʁɪft)(tuːm)/ [ʃʁɪftːuːm]~[ʃʁɪftˑuːm]~[ʃʁɪftuːm] ‘professional literature’

For the phonetic implementation of a word like /(ʃʁɪft)(tuːm)/, not only cue constraints but also articulatory effort plays a role, more precisely how effortful the phonetic realization of a closure phase is. This is formalized as gradient *EFFORT constraint, which is violated more often in realizations with a long closure duration than in those with a short one. Tableau (14), with cue constraints (12a), shows that if the cue constraints */tt/[t] and */tt/[tˑ] are ranked above *EFFORT, this results in a careful pronunciation of /(ʃʁɪft)(tuːm)/ with a long closure phase:

(14) *Phonetic implementation of* /(ʃʁɪft)(tuːm)/ *in slow, careful speech*

/(ʃʁɪft)(tuːm)/	*/tt/[t]	*/tt/[tˑ]	*EFFORT	*/tt/[tː]
☞ [ʃʁɪftːuːm]			***	*
[ʃʁɪftˑuːm]		*!	**	
[ʃʁɪftuːm]	*!		*	

High-ranked *EFFORT, on the other hand, results in a realization with a short closure phase, cf. (15), with the same cue constraints and their ranking as in (14).

(15) *Phonetic implementation of* /(ʃʁɪft)(tuːm)/ *in quick speech*

/(ʃʁɪft)(tuːm)/	*EFFORT	*/tt/[t]	*/tt/[tˑ]	*/tt/[tː]
[ʃʁɪftːuːm]	**!*			*
[ʃʁɪftˑuːm]	**!		*	
☞ [ʃʁɪftuːm]	*	*		

For the phonetic implementation of /(haṭə)/, with a single surface alveolar plosive, the cue constraints in (12b) are relevant, see the formalization in tableau (16). The exact ranking of the *EFFORT constraint in this tableau is irrelevant, since both the ranking of the cue constraints and *EFFORT favor the same, third candidate. This means that the phonetic realization of this form is not dependent on speech rate.

(16) *Phonetic implementation of* /(haṭə)/

/(haṭə)/	*/t/[tː]	*/t/[tˑ]	*EFFORT	*/t/[t]
[hatːə]	*!		***	
[hatˑə]		*!	**	
☞ [hatə]			*	*

This section showed how obligatory phonological degemination within p-words is formalized in the phonology module with the high-ranked structural constraint *GEM, while optional phonetic degemination across p-words is formalized in the phonetics module with an articulatory *EFFORT constraint whose ranking depends on speech rate. In the following section, we will see that the same high-ranked phonological *GEM plays a role in speech perception.

3.2. Perception and recognition

The process of speech comprehension consists of perception and recognition. In speech perception, an auditory input has to be mapped onto a phonological surface representation, recall Figure 1. Listeners use auditory durational cues to perceive a consonant with short closure duration as singleton and a consonant with long closure duration as geminate. This is formalized with the same cue constraints as given for phonetic implementation in (11), though now they are interpreted in the reverse direction, e.g. */tt/[t] as "Assign a violation mark for every short closure phase that is mapped onto two alveolar plosives". Articulatory constraints like *EFFORT play no role in perception. Instead, the output of this mapping is restricted by the structural constraints *GEM and FVC that we know already from phonological production.

The interaction of cue and structural constraints is illustrated in tableaux (17) for the perception of [ʃʁɪftːuːm], with a long closure duration, and in (18) for the perception of [hatə], with a short closure duration. Only those cue constraints from (12) are included that refer to the auditory cues in the respective inputs.

A listener faced with the auditory input [ʃʁɪftːuːm] "perceives" this as containing two separate p-words, see the winning candidate in (17), because the long closure phase has to be mapped onto two surface /tt/ (otherwise the cue constraint */t/[tː] is violated, see candidate two), and interpreting these two surface /tt/ as belonging to one p-word is not allowed (as it would violate *GEM, see candidate three).

(17) *Perception of* [ʃʁɪftːuːm]

[ʃʁɪftːuːm]	*GEM	FVC	*/t/[tː]	*/tt/[tː]
☞ /(ʃʁɪft)(tuːm)/				*
/(ʃʁɪfṭuːm)/			*!	
/(ʃʁɪft.tuːm)/	*!			*

For the input [hatə] in (18), on the other hand, the short closure phase cannot be mapped onto two surface /tt/ (this would violate the cue constraint */tt/[t], see candidates one and three). Candidate one, with two p-words, additionally violates FVC because the second p-word has no full vowel, and candidate three has a geminate within a p-word and violates *GEM.

(18) *Perception of* [hatə]

[hatə]	*GEM	FVC	*/tt/[t]	*/t/[t]
/(hat)(tə)/		*(!)	*(!)	
☞ /(haṭə)/				*
/(hat.tə)/	*(!)		*(!)	

Winner is the second candidate, and a listener perceives the phonetic form [hatə] as surface /(haṭə)/. However, this form has to be retrieved in the lexicon as consisting of the stem |hat| and the inflectional suffix |tə|. A reconstruction of the correct underlying form, that is the undoing of phonological degemination, happens in the recognition process with the help of lexical information: the lexicon does not contain a monomorphemic entry |hatə|. This is formalized with a lexical restriction LEX: "only employ words that exist in your lexicon" (based on Boersma 2001), see tableau (19). Candidates one and two violate LEX because they result in non-existent underlying forms. Candidate three, the winning candidate, violates the faithfulness constraint DEP-C, which militates against insertion of consonants in the recognition process: The winning candidate has one plosive in the surface form, but two in the underlying form.

(19) *Recognition of the surface form* /(haṭə)/

/(haṭə)/	LEX	DEP-C
|hatə|	*!	
|hat+ə|	*!	
☞ |hat+tə|		*

For [hatə], the perception and recognition together result in |hat+tə|, the only possible parse in German. For homonyms with different morphological structure, e.g. German [matə] for monomorphemic |matə| 'mat' and bimorphemic |mat+tə| 'dull-F', however, lexical information alone would always result in |matə| as winning candidate, because this candidate does not violate DEP-C, while |mat+tə| does. In such cases, syntactic and semantic context are necessary to retrieve the correct meaning. This could be formalized with separate semantic and syntactic representations and corresponding constraints, or with LEX constraints that include semantic context (for an example of the latter, see Boersma 2001). Due to lack of space I do not provide a formalization of this.

A similar reconstruction of the correct underlying form as we saw in (19) is also necessary for the extremely shortened [ʃʁɪftuːm] in very quick speech. [ʃʁɪftuːm] is perceived with the perception grammar employed before as surface /(ʃʁɪf)(tuːm)/, cf. the winning third candidate in tableau (20).[7]

(20) *Perception of* [ʃʁɪftuːm]

[ʃʁɪftuːm]	*GEM	FVC	*/tt/[t]	*/t/[t]
/(ʃʁɪft)(tuːm)/			*!	
/(ʃʁɪft.tuːm)/	*(!)		*(!)	
☞ /(ʃʁɪf)(tuːm)/				*

Perceiving the short closure phase as /tt/, as in the first and second candidate, violates the cue constraint */tt/[t]. This "misperception" of a phonetically shortened fake geminate as surface /(ʃʁɪf)(tuːm)/ can only be corrected via lexical information: the lexicon does not contain an entry with the form |ʃʁɪf|. This recognition phase for /(ʃʁɪf)(tuːm)/ is formalized in tableau (21), with the same constraints and ranking as in recognition tableau (19):

(21) *Recognition of the surface form* /(ʃʁɪf)(tuːm)/

/(ʃʁɪf)(tuːm)/	LEX	DEP-C
☞ |ʃʁɪft+tuːm|		*
|ʃʁɪf+tuːm|	*!	

3.3. Reading

In section 3.2, we saw how the high-ranked structural constraint *GEM influences both phonological production and speech perception of German. In this section, I illustrate how the same constraint is relevant in the process of reading German, and will formalize this process with a reading grammar (Hamann & Colombo 2017). Figure 2 above showed that a reading grammar maps a written form onto a surface phonological form with the help of orthographic constraints. Two orthographic constraints necessary for the present analysis, and their ranking, are given in (22): (22a) maps a single-letter grapheme onto a phoneme, and constraint (22b) a vowel grapheme in a specific orthographic context onto a phonological feature. The latter accounts for German sharpening, i.e. the fact that a stressed vowel has to be phonologically short/tense if it is written with a vowel grapheme that is followed by two identical consonantal letters.[8]

---

[7] A candidate /(ʃʁɪfṭuːm)/ with an ambisyllabic coronal plosive and only one p-word is not included in this tableau (or the earlier production tableaux (9) and (10) and the perception tableaux (17) and
)). It would violate a constraint requiring that every (primarily or secondarily) stressed full vowel forms the head of its own p-word (see e.g. Raffelsiefen 2000 for a similar argument), which I left out for reasons of clarity and space.

[8] The native reader needs to make a distinction between vowel and consonantal letters in order to be able to interpret this constraint, see Neef (2002: 171) for a similar assumption.

(22) a) <t>/t/: Assign a violation mark for every grapheme <t> that is not mapped onto a surface form /t/, and vice versa.

b) <$\alpha(\beta_i\beta_i)$>/−long/: Assign a violation mark for every vowel grapheme followed by two identical consonantal letters that is mapped onto a surface long vowel.

c) <$\alpha(\beta_i\beta_i)$>/−long/ >> <t>/t/

In the German written form <Matte> for phonological /(maṭə)/ 'mat', one of the two graphemes <t> is thus used solely to express the shortness of the preceding vowel, and does not map onto its own surface /t/. This shows that a constraint regulating the mapping between single consonantal graphemes and single consonantal phonemes as in (22a) has to be ranked below the shortening constraint (22b), with the resulting ranking given in (22c), in order to yield the correct reading of German.

Tableau (23) formalizes the reading of <Matte>, where the two orthographic constraints in (22) interact with the by now familiar structural constraints *GEM and FVC, to render the correct output. Recall that in German, the same vowel graphemes are used for both the short and the long low vowel of the same/similar quality, thus <a> can stand for /a/ or /aː/, and that orthographic sharpening is used exactly to disambiguate between these two possibilities.[9]

(23) *Reading of native* <Matte>

<Matte>	*GEM	FVC	<$\alpha(\beta_i\beta_i)$>/−long/	<t>/t/
/(maː.tə)/			*!	*
/(mat.tə)/	*!			
/(mat)(tə)/		*!		
☞ /(maṭə)/				*

The first candidate maps the vowel grapheme <a> followed by two identical consonantal graphemes onto a [+long] vowel, thereby violating the sharpening constraint (22b). Candidate one furthermore violates the orthographic constraint <t>/t/, and so does candidate four, because one of the two graphemes <t> in the input form is not mapped onto a surface /t/. Candidate two and three, on the other hand, have two corresponding consonants (/tt/) in the surface form. While in candidate two, these surface consonants occur in the same p-word and thus violate *GEM, they are assigned to different p-words in candidate three, the second without a full vowel and therefore violating FVC. The winner is candidate four with a single surface /t/ preceded by a short vowel.

Parallel to the evaluation in (23), the bimorphemic <hatte> 'had' results in the phonological surface /(haṭə)/. In order to arrive at the correct underlying form |hat+tə|, the recognition process needs to take into account that |hatə| does not exist in the

[9] Other ways to orthographically disembiguate phonological vowel length in German are the use of double vowel letters or a single vowel letter followed by <h> for phonologically long/tense vowels. A full formal account of these mechanisms and their exceptions would go beyond the scope of this paper and are left for future studies.

lexicon. This step has been formalized already with Lex in the recognition grammar, see tableau (19) above.

The word <matt> /(mat)/ ‘dull-m/n’ shows that German orthographic sharpening also applies in p-word final position, because the two final identical graphemes are again interpreted as one single surface consonant. For a formalization, see tableau (24).

(24) *Reading of* <matt>

<matt>	*Gem	FVC	<α($\beta_i\beta_i$)>/–long/	<t>/t/
/(maːt)/			*!	*
/(matt)/	*!			
☞ /(mat)/				*
/(maːtt)/	*!			

Fake geminates preceded by a short vowel are also read correctly with the present reading grammar. This is illustrated with the word <Schifffahrt> ‘shipping’ in (25), where <f>/f/ is used instead of <t>/t/.

(25) *Reading of* <Schifffahrt> ‘shipping’

<Schifffahrt>	*Gem	FVC	<α($\beta_i\beta_i$)>/–long/	<f>/f/
/(ʃɪff)(faːɐt)/	*!			
☞ /(ʃɪf)(faːɐt)/				*
/(ʃɪf̣aːɐt)/				**!
/(ʃiːf)(faːɐt)/			*!	*

The first candidate maps all three <f> graphemes onto their own surface /f/, and by doing so violates *Gem. In the second, winning candidate one <f> is not mapped onto a surface /f/, but this candidate suffices the structural constraints and the sharpening constraint. Candidate four maps the grapheme <i> onto a long vowel, and thereby violates the sharpening constraint.

Words such as <Brotteig> |bʁoːt+taɪg| ‘bread dough’, with a vowel that is not orthographically marked as long (by a double vowel grapheme or a following <h>) and is followed by a fake geminate, are read by the present reading grammar as having a short vowel because of the sharpening constraint, see tableau (26).

(26) *Reading of* <Brotteig>

<Brotteig>	*GEM	FVC	<$\alpha(\beta_i\beta_i)$>/–long/	<t>/t/
/(bʁoːṭaɪg)/			*!	*
/(bʁoːt.taɪg)/	*(!)		*(!)	
☞ /(bʁɔt)(taɪg)/				
/(bʁoːt)(taɪg)/			*!	

The first two candidates illustrate that the orthographic word has to be interpreted as consisting of two p-words with a fake geminate. Candidate three, the winning candidate, interprets the grapheme sequence <ott> as having a short vowel, and thus follows the sharpening constraint, as opposed to candidate four. We assume that the winning surface form /(bʁɔt)(taɪg)/ is correctly interpreted in the recognition process as |bʁoːt+taɪg| with the help of the lexical knowledge that |bʁɔt| is not an existing German word form (formalized with LEX).

For words like <Schrifttum> |ʃʁɪft+tuːm| 'professional literature', the reading grammar results in two winning candidates, see tableau (27). The two identical consonant graphemes are preceded by another consonant grapheme, and therefore the sharpening constraint does not apply, and phonotactically, both short /ɪ/ and long /iː/ are possible in this context.

(27) *Reading of* <Schrifttum>

<Schrifttum>	*GEM	FVC	<$\alpha(\beta_i\beta_i)$>/–long/	<t>/t/
/(ʃʁɪfṭuːm)/				*!
/(ʃʁɪft.tuːm)/	*!			
☞ /(ʃʁɪft)(tuːm)/				
☞ /(ʃʁiːft)(tuːm)/				

For cases like these, where the quantity of the vowel is orthographically not specified, the recognition process will provide again the correct underlying form.[10]

4 Discussion and conclusion

In this article, I illustrated that the same high-ranked *no geminate* constraint that restricts the phonological production in German also applies in the processes of speech perception and reading. This cross-modal application was formalized in the OT BiPhon grammar (Boersma 2007) with the extension of a reading grammar (Hamann &

[10] Ramers (1999: 57) observes that vowel graphemes followed by two (or more) consonantal graphemes/letters in German are usually short, which could be formalized as an orthographic constraint <$\alpha(\beta_i\beta_{ii})$>/–long/ in the reading process (see Neef 2002 for a similar proposal). However, due to the considerable number of exceptions both in monomorphemic words and in inflected verbs, this constraint could only be low ranked, and the exceptions would still need to be "repaired" in the recognition process.

Columbo 2017). As opposed to earlier accounts of reading and writing, where phonological knowledge such as phonotactic constraints and phonological processes is either reduplicated in the orthographic mapping or not formalized at all, the present model employs the independently needed phonological module to restrict the reading process.

Several issues have not or only very superficially been dealt with in the present article. To illustrate the workings of a reading grammar, I used German orthographic sharpening. Other orthographic encodings of German vowel duration, such as the use of double vowel graphemes, the grapheme <ie> or one vowel followed by <h> for phonologically long/tense vowels were not discussed. A full reading grammar of German would of course need to incorporate all of these mappings.

Secondly, reading of an alphabetic script was consistently viewed as proceeding via the so-called sub-lexical route, i.e. the mapping of an orthographic form via a phonological surface onto a lexical form. Psycho- and neurolinguistic studies, however, provide evidence for the simultaneous existence of a direct lexical route, especially in proficient readers, where the visual input of the written form is mapped directly onto a lexical form (the dual route model; see e.g. Coltheart et al. 1993, Coltheart et al. 2001, Grainger et al. 2012). This lexical route can be included in the present OT model by constraints mapping whole orthographic word forms onto pairs of meaning and underlying forms. These constraints can interact in a parallel evaluation with the sub-lexical route constraints, allowing for a competition between the two routes. The elaboration of this topic is left for future research.

And lastly, this article did not deal with the writing process. As illustrated for cue and faithfulness constraints, all mapping constraints in the BiPhon model can be used bidirectionally, i.e. for the speaking as well as the listening direction. This bidirectionality also holds for the orthographic constraints in the reading grammar. Hamann & Colombo (2017) show with examples from Italian, a language with a very transparent orthography, how orthographic constraints can be used in the reverse direction to formalize the writing process. How the orthographic sharpening constraint introduced in this article is applied in the writing process in German, and how it interacts with possible constraints on the orthographic output (such as "double graphemes like <schsch> or <cc> are not allowed") needs to be elaborated in future work.

Acknowledgements

The content of this manuscript was presented at the workshop "Geschriebene und gesprochene Sprache als Modalitäten eines Sprachsystems" as part of the 38th *Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft* in Konstanz. Parts of it were also presented at the 24th *Manchester Phonology Meeting*. I would like to thank especially the audience of the latter for very productive discussions. I also thank an anonymous reviewer for helpful comments.

References

Becker, Thomas (1996). Zur Repräsentation der Vokallänge in der deutschen Standardsprache.

Zeitschrift für Sprachwissenschaft 15: 3–21.
Bierwisch, Manfred (1972). Schriftstruktur und Phonologie. *Probleme und Ergebnisse der Psychologie* 43: 21–44. Reprint in: F. Kiefer (ed.) *Phonologie und Generative Grammatik*. Frankfurt: Athenaion (1975). First volume; 11–51.
Boersma, Paul (2001). Phonology-semantics interaction in OT, and its acquisition. In: R. Kirchner, Wikeley & J. Pater (eds.) *University of Alberta Papers in Experimental and Theoretical Linguistics* 6: 24–35.
Boersma, Paul (2007). Some listener-oriented accounts of h-aspiré in French. *Lingua* 117: 1989–2054.
Boersma, Paul (2012). A constraint-based explanation of the McGurk effect. In R. Noske and B. Botma (eds.) *Phonological Architecture: Empirical, Theoretical and Conceptual Issues*. Berlin/New York: Mouton de Gruyter, 299–312.
Boersma, Paul, Titia Benders, and Klaas Seinhorst (2018). *Neural Networks for Phonology and Phonetics*. Manuscript, University of Amsterdam. Downloadable from: http://www.fon.hum.uva.nl/paul/
Boersma, Paul and Silke Hamann (2009). Loanword adaptation as first-language phonological perception. In A. Calabrese and L. Wetzels (eds.) *Loan Phonology*. Berlin: Mouton de Gruyter, 11–58.
Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English*. New York: Harper & Row.
Coltheart, Max, Brent Curtis, Paul Atkins and Michael Haller (1993). Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review* 4: 589–608.
Coltheart, Max, Kathleen Rastle, Conrad Perry, Robyn Langdon and Johannes Ziegler (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* 108: 204–256.
Eisenberg, Peter (1999). Vokallängenbezeichnung als Problem. *Linguistische Berichte* 179: 343–349.
Escudero, Paola and Paul Boersma (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–585.
Féry, Caroline (1998). German word stress in optimality theory. *Journal of Comparative Germanic Linguistics* 2: 101–142.
Ganong, William F. (1980) Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6: 110–125
Grainger, Jonathan, Bernard Lété, Daisy Bertand, Stéphane Dufau and Johannes C. Ziegler (2012). Evidence for multiple routes in learning to read. *Cognition* 123: 280–292.
Hall, T. A. (1999). Phonotactics and the prosodic structure of German function words. In: T.A. Hall & U. Kleinhenz (eds.) *Studies on the Phonological Word*. Amsterdam: John Benjamins, 99–131.
Hamann, Silke and Ilaria Colombo (2017). A formal account of the interaction of orthography and perception: English intervocalic consonants borrowed into Italian. *Natural Language & Linguistic Theory* 35,3: 683–714.
Hamann, Silke (2016). Phonetic, phonological and orthographic degemination in Dutch and German – a unified formal account. Poster presentation at the *24th Manchester Phonology Meeting*, Manchester.
Itô, Junko and Armin Mester (2008). The extended prosodic word. In: J. Grijzenhout and B. Kabak (eds.) *Phonological domains: universals and deviations*. Berlin: Mouton de Gruyter, 105–194.
Kohler, Klaus (2001). Variability of opening and closing gestures in speech communication. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel*

35: 33–96.
Maas, Utz (1992). *Grundzüge der deutschen Orthographie*. Tübingen: Stauffenburg.
McCarthy, John (1986). OCP effects: Gemination and Antigemination. *Linguistic Inquiry* 17: 207–63.
Neef, Martin (2002). The Reader's View: Sharpening in German. In: M. Neef, A. Neijt and R. Sproat (eds.)*The Relation of Writing to Spoken Language.* Tübingen: Niemeyer; 162–192.
Neef, Martin (2012). Graphematics as part of a modular theory of phonographic writing systems. *Writing Systems Research* 4: 214–228.
Prince, Alan and Paul Smolensky (1993 [2004]). *Optimality Theory: constraint interaction in Generative Grammar*. London: Blackwell.
Raffelsiefen, Renate (2000). Evidence for word-internal phonological words in German. In: R. Thieroff, M. Tamrat, N. Fuhrhop and O. Teuber (eds.) *Deutsche Grammatik in Theorie und Praxis.* Tübingen: Niemeyer, 43–56.
Ramers, Karl-Heinz (1999). Vokalquantität als orthographisches Problem: Zur Funktion der Doppelkonsonanzschreibung im Deutschen*. Linguistische Berichte* 177: 54–64.
Rose, Sharon (2000). Rethinking geminates, long-distance geminates and the OCP. *Linguistic Inquiry* 31: 85–122.
Sternefeld, Wolfgang (2000). Schreibgeminaten im Deutschen: Ein Fall für die Optimalitäts-theorie. *Linguistische Berichte* 181: 35–54.
Wiese, Richard (1987). Laut, Schrift und das Lexikon. *Deutsche Sprache* 15: 318–335.
Wiese, Richard (2000). *The Phonology of German.* Oxford: Oxford University Press. 2nd edition.
Wiese, Richard (2004). How to optimize orthography. *Written Language & Literacy* 7: 305–331.