Recursion in Phonology - Stuttering as a possible language

Gertjan Postma
Meertens Institute

Abstract
In this study, we carry out a phonological analysis of a language fragment of an adult person who stutters. We replicate the finding in Nippold (2002) and Marshall (2005) that stutter does not correlate with the featural or the structural complexity of the onset of the stuttered word. We argue, on the other hand, that enhanced levels of stuttering are found in function of the complexity of the codas of the immediately preceding word. Complex codas make use of the prosodic space opened up by the subsequent word, more specifically Lowenstamm's CV structure that precedes each word. Stuttered language has stricter licensing conditions on (empty) moras and syllables in general, which becomes directly visible in this leading CV. Various types of stutter, such as consonant prolongation, CV-reduplication, as well as dummy prolonged schwas, follow from this stricter licensing. The fact that this parameter setting has communicative disadvantages, such that it is called a disorder by the WHO, shows that parameter setting in natural language does not necessarily result in possible languages that are functionally equal. We recommend not to speak about "persons who stutter" (PWS) and "persons without stutter" (PNS), but about language variants with stutter (LWS) and language variants with no stutter (LNS). The fact that languages variants vary with respect to licensing conditions, locates stuttering at the level of ordinary linguistic micro-variation. Finally, we predict more advance levels of stuttering where recursion is involved. It does not situate the difference between LWS and LNS in the licensing conditions on empty mora alone but also in a difference in complexity of Lowenstamm's structure (monomoraic and/or bimoraic words generate it). The monomoraic licensing causes recursion in this phonological structure being monomoraic itself, while bimoraic licensing blocks recursion. This phonological recursion is functionally disadvantageous. We suggest that while recursion in syntax is functionally advantageous (because of its interface with the semantics), recursion in phonology is disadvantageous, but a normal option of UG.

Introduction

Stuttering is a speech disorder in which the flow of speech is disrupted by involuntary repetitions and prolongations of sounds, syllables, words or phrases as well as involuntary silent pauses or blocks in which the person who stutters is unable to produce sounds (WHO definition). Stuttering varies a lot in its realization, from virtually unnoticeable to completely blocking communication. Though the causes are not really known, various favoring factors have been identified, such as sex (it has higher incidence in males than in females, up to a factor 4), genetic disposition (Bloodstein & Bernstein Ratner 2008; Rautakoski *et al* 2012, Raza *et al* 2012), other speech disorders (Blood et al 2003, Ntourou et al 2013), and cultural background: low incidence of this disorder has been reported in South-Amerindian and South-East Asian areas (Reese & Jordania 2001, Sayer 2010). Apart from these language-external factors, language-internal factors have been studied: there is a laic believe that some sounds (e.g. /p/ and /h/) favor stuttering more than others (cf. Klok 2009). Various studies have been carried out on the phonology (Marshall 2005, Greg & Yari 2012) but a truly linguistic analysis is missing thus far.[1] Moreover, stuttering is exclusively studied from a developmental perspective, not from a stable adult perspective -- from a perspective of speech pathology, not from a possible language perspective. Whatever the causes, the social reactivity and social attitude to stuttering greatly enhances the levels of stuttering and word duration (Choi *et al.* 2013). If there is a linguistic core, we should study the phenomenon in circumstances without influences of these social factors. We, therefore, define *core stuttering* as the language of an adult PWS who has perfect knowledge of his language and speaks in a completely relaxed mood, is perfectly certain of himself, without any conversational or sociological inhibitions, or any other negative force that may influence his speech. In this paper, we study core stuttering in this sense. To approximate this ideal PWS, we analyze an interview fragment of a PWS, an adult Dutch native speaker who is often on radio and television, and is not inhibited in any way of his stuttering. In this fragment, the stuttering nature of the speech is hardly noticeable for the general hearer, but on careful scrutiny the speech contains ample consonant prolongations, reduplications, repetitions and other disfluencies.

1. Empirical data

We used a 9 minute sound fragment (8:59 sec) with an interviewer (1:15s) a person without stutter (PNS) and an interviewee (6:44 sec), who is the person with stutter (PWS). This test person had 99 disfluencies in this fragment of 1393 words (roughly 1 disfluency every 5 to 6 sec, 7 disfluency per 100 words). The testing person was male, 52 years old, who has had speech lessons in his childhood. In this fragment, the PWS speaks completely relaxed about his stuttering, in an inhibited way. This fragment was visualized in Praat, annotated, corrected[2], and analyzed. Five types of disfluencies could be identified: consonant lengthening (c:), illustrated in (1abc), onset reduplications (cv2), as in (1de). Combinations of these two are possible (1f). Furthermore, word and phrasal repetitions were observed, given in (1ghi) respectively. Finally, prolongation of non-lexical material was found (not illustrated).

(1) a. Hoewel dat mij **b:ijna** nooit overkomt (4:44)
although that me almost never occurs

b. want ik had het als **k:leine** knaap **m:ee**gemaakt (2:08)
because I had it as small kid experienced
'because it happed to me as asmall kid'

c. ja, **w:as** een soort zelfoverwinning (2:36)
yes, (it) was a kind of self-conquer

d en toen-toen **he-heb**, toen heb ik daarmee een soort eh drempel overwonnen
and then have, then have I therewith a kind tresholt conquered
'and then, doing so, I made a big step'

e. Jeetje, je begint meestal **zo-zo'n** jaar of (acht)... (3:27)
jesus, you start usually such year or eight
'Jesus, one usually starts at an age of eight or so'

f. Toen kwam mijn **moe-m:oeder** te overlijden (2:00)
then came my mother to die
'then my mother happened to die'

g. Bijvoorbeeld **voor-voor** een optreden op de radio (4:00)
for example for a performance on the radio

[1] Various studies use the term 'phonological' (e.g. Marchall 2005, Arnold 2005) in a broader sense, and do not make use of phonological theories.

[2] I thank Etske Ooijevaar for correcting the transcription and analysis in Praat. The intersubjective disagreement was less than 2%.

h. Ja, **ik heb-ik heb** er nooit lak aan (4:26)
yes, I have-I have there never gum at
'yes, I never ignore it'

i. iemand die ik **hele-hele**maal[3] niet kende (7:12)
somebody that I at all not know
'somebody that I did not know at all'

We will call (1a-f) phonological stutter, and (1g-i) prosodic stutter. We will concentrate in this paper on the phonological stutter.

2. Results

The following distribution of the various types of disfluencies was found.

(2)

Stutters	occurrence	type	code
1a. consonant stutters ("m:annetje"):	44	phonological	c:
1b. CV-stutters ("ma-mannetje"):	20	phonological	cv2
2a. word stutter ("van-van"):	16	prosodic	w2
2b. phrasal stutters ("manne-manne-tje"):	19	prosodic	p2
3. non-lexical disfluencies	2		
Total	99		

The full dataset is given in Appendix 1. In the next sections we will analyze this data set quantitatively and theoretically.

2.1 Onset articulatory features

One obvious trigger of stutter that comes to mind are the phonological features of the onset in the stuttered word or phrase. The interviewee explicitly mentions [h] and [p] in this fragment to be notoriously difficult. In the next table, we give the distribution of the disfluencies over the onsets together with the token frequencies in the language (measured with CELEX database - Dutch word forms)[4]. Stutter seem to occur with all onsets and no specific bias seems to be present.

(3)

onset	CELEX in millions	Total Stutter Incidence	c:+cv	c:	cv	p2+w2
[ʔ-]	7.2 [5]	24	8	3	5	16
[d-]	5.6	12	8	6	2	5
[f/v-]	4.2	12	6	6	-	6
[h-]	3.5	4	1	0	1	3
[m-]	2.1	7	7	5	2	1

[3] It is possible to make this repetitive degree sequence in the target language without it being a stutter. It then has a upstep intonation, typical for degree phrases. We may assume that it is then supported by the syntax. In a stutter, on the other hand, the two prosodic patterns over *hele* are equal.

[4] Cf. Baayen *et al.* (1993). To be consulted on line at: http://celex.mpi.nl.

[5] This is the number of vocalic onsets (11.6 million) in Celex (Query: (PhonCV =~ /^V/)) minus the function words (*en*/*een*/*is*, 2.5 million) and the light prepositions (1.9 million), of which we assume they are without underlying glottal stop. Cf. Jongenburger & van Heuven (1991) for an interpretation in terms of stress.

[w-]	1.9	8	6	5	1	3
[x/ɣ]	1.8	3	1	-	1	2
[b-]	1.7	5	5	4	1	-
[t-]	1.7	4	4	3	1	-
[n-]	1.4	1	1	0	1	1
[s-]	1.3	5	5	5	-	-
[k-]	1.3	2	2	2	-	-
[l-]	0.8	1	1	1	-	-
[p-]	0.7	1	1	1	-	-
[j-]	0.5	1	-	-	-	1
[r-]	0.5	0	-	-	-	-

The three most frequent onsets in Dutch (ʔ, d and f/v) are also most frequent in stutter. The three less frequent onsets are also the less frequent in stutter. Moreover, no articulatory feature, neither place features [labial], [dental], nor manner features, such as [voice] and [nasal], can be identified that favors or disfavors phonological stutter.[6] The laryngeal features [h] and [ʔ] predominantly trigger prosodic stutter rather than phonological stutter (c: and cv2), though the latter do occur in this small dataset with glottal stutter, illustrated in (4).

(4) a. Pa is over een '**echt** onderwerp aan het woord [ʔ:ɛxt]
dad is about a real subject on the word
'Dad is discussing a real subject'
b. Da's gewoon een soort '**ellendige** mug [ʔ:ɛlɛndəɣə]
that-is simply a kind-of terrible mosquito
c. Maar eh **i-ik** kan be-bijvoorbeeld voor-voor een optreden ... [ʔiʔik]
but uhm I can for instance before a performance
d. het is een rare kluwen van **oorzaak** en (gevolg) [ʔ:o:rza:k]
it is a strange tangle of cause and effect

We cannot conclude that any articulatory feature of the onset favors phonological stuttering. Stutter is more or less random distributed over the various onsets, as to token frequency on the basis of CELEX. In the next section we will study the *complexity* of the onset as a potential trigger.

2.2. Onset complexity

Let us now look at the segmental complexity of the onset. Does a more complex onset favor or disfavor stuttering? In this section, we confine ourselves to phonological stutter, i.e. consonant prolongation ("gemination", c:) and reduplicative onsets (cv2). In (5) we listed the incidence of these disfluencies in this corpus.

[6] The reader can easily check this using the data above. I here give the data for [±voice]:

	segments	CELEX	Stutter	Expected using CELEX
[+voice]	/dbw/	9.2 m	25	24.5
[-voice]	/tpf/	6.6 m	17	17.5
	Total	15.8 m	42	42

(5) *Distribution of consonant lengthening (c:) in comparison with the token frequencies in CELEX*

onset type	CELEX	incidence	expected using CELEX
c	2.6 m	25 (86%)	26.5
cc	2.6 m	4 (14%)	2.5
ccc	0.15 m	0 (0%)	0
Total	5.35 m	29	29

There is no significant effect of favoring of disfavoring (p-value 0.3). It seems that vocalic onsets count as consonantal in Dutch, such as *echt* [ʔεχt], as this [ʔ] cannot be dropped in bound speech in Dutch and participates in stutter, cf. above in (4). If we consider them consonantal as well, we add this single example to the c-onset and we get the table in (6).

(6) *Distribution of consonant lengthening (c:) in comparison with the token frequencies in CELEX (vocalic onset are taken as [ʔ])*

onset type	CELEX	incidence	expected using CELEX
v=ʔ	7.5 m		
c	2.6 m	25+1 (86%)	24
cc	2.6 m	4 (14%)	6
ccc	0.15 m	0 (0%)	0
Total	12.85	30	30

Once again, there is no significant effect of favoring of disfavoring (p-value=0.3). It does not deviate in a significant way from the expected stutter upon a random distribution. We conclude that stutter is not sensitive to the onset complexity. This replicates the findings in (Marshall 2005). If phonology is involved at all, it might only be in a dimension that is not sensitive for the onset complexity. One such dimension that is well-known from the phonological literature, is moraic weight. In moraic theory, onsets are considered weightless (Hayes 1989), while coda material has moraic weight. In the next, section we study the influence of the coda of the preceding word on stuttering.[7]

2.3. Relation with preceding word

In the previous sections we showed that the onset of the stuttered word is not a trigger for the stuttering, neither with respect to feature content nor complexity of the onset: all onset are more or less equally affected by stuttering. Our next question is whether the preceding word is a favoring factor. In segmental terms, we get the following distribution of stutter in function of the preceding segment.

[7] Marshal (2005) studies the influence of the *coda of the stuttered word* on the word itself, i.e. a word internal influence of coda on onset. No effect was found. To my knowledge, no phonological theory predicts such interaction. Marshall is testing a processing hypothesis.

(7)

coda consonant	CELEX in millions	word count in text	phonological stutter (c: + cv) stand.dev.=1.8 [8]	expected based on CELEX	expected based on text	effect up/down
ə +ən	10.1	107 + 62 =169	**3**	**14.3**	7	↓↓
t	7.5	304	**3**	**10.6**	12	↓↓
V minus ə	3.4	143	4	4.8	6	-
m	0.7	19	0	0.9	1	-
f	0.5	23	0	0.7	1	-
r	3.2	120	5	4.5	5	-
l	1.0	75	3	1.4	3	-
x	1.0	42	4	1.4	2	-
p	0.7	20	3	0.9	1	-
ŋ	0.6	8	0	1	0	-
n minus ən	6.7	273 - 62 =201	**14**	**9.5**	8	↑↑
s	2.5	93	**8**	**3.5**	4	↑↑
k	1.5	143	**8**	**2.1**	6	↑↑
Total	39.4	1370	55	55	55	

In contrast to the onset of the stuttered word, the coda of the word that precedes the stutter does play a role. If we order the codas again with respect to their frequency in CELEX, we see the two most frequent codas, responsible for almost 50% of words, co-occur with less than 15% of the stutters. Conversely, the three consonants {rsk} are responsible for almost 50% of the stutters while they contribute only for 20% of the number of words. A strongly disfavoring effect can be observed with preceding codas in /-ə/ and to a lesser extent with codas in /-t/. With full vowels and the consonant /-n/, there is very little effect. All other consonants have a positive effect on stutter. These are the coronal and velar segments /rskxl/. Because of the low numbers obtained in this corpus, no conclusions can be drawn for preceding words in the labials /-p/, /-m/ and /-f/. We conclude that it is not so much the articulatory features on the coda, but rather the prosodic nature of the coda that influences stuttering in the word that follows. Relevant conditions, therefore, seem to be: Does the previous word ends in a schwa? (disfavoring effect). Does it ends in a high sonorous coda, e.g. full vowels and /n/? (very little effect). Does it end in a consonant (favoring effect)? This suggest a classification in prosodic terms.

In moraic terms, there are 4 possible leads: pause, 1 mora which typically occurs in reduced schwa syllables (Cə, CəC, CəCC, indicated by σ1), 2 moras (VV or VC, indicated with σ2), and 2 moras to extra consonantal material is attached (VVC and VCC), indicated with σ2+. The distribution is given in the table in (8).

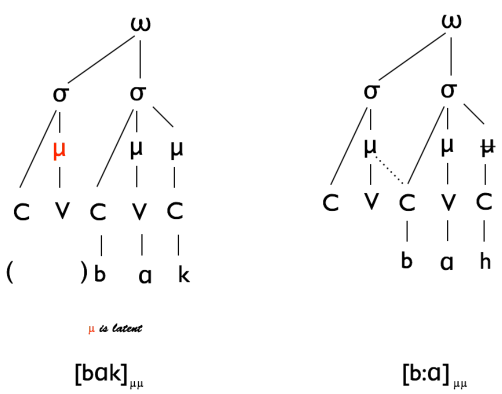
(8)	CELEX in 10^6 words	stutter (c:)	expected	effect
leading pause	-	(2)		
leading σ1 (Cə,CəC,CəCC)	13.0 (10.8+1.6+0.6)	5	12	down 60%
leading σ2 (CVV,CVC)	9.3 (2.7+6.6)	10	8	up 25%
leading σ2+ (VVC,VCC)	10.6 (7.1+3.5)	14	9	up 50%
Total	32.9	29	29	

[8] The standard deviation is 1/4 •√n=1/4• √55 = 1.85 ≈ 2. We here assume values that are outside the range of two standard deviations (i.e. 4) to be unequal. The error is CELEX is taken to be zero.

We observe a disfavoring effect of 60% on c:-stutter in the context of a weak preceding syllable σ1. In the context of bimoraic preceding material (σ2), we see little effect. Here we see more or less the same distribution as one would expect of stutter were randomly distributed. In the context of preceding bimoraic material to which additional segmental material is adjoined, we see a relatively strong favoring effect of more than 50% upward with respect to random. If we lump the bimoraic together and confront it to the light schwa material, we get a significant weight effect on the stutter with a p-value of 0.04. If we lump the leading mono- and bimoraic together and confront is with the superheavy codas we can a significant favoring effect of superheavy material with a p-value between 0.1 (one-sided) and 0.2 (two sided). Although the data set is somewhat limited, we conclude that moraic weight of the preceding syllable has an favoring effect on stutter of the gemination type (c:). This suggest that there might be a moraic interaction between the moraic nature of the coda and the onset. In the next section, we will briefly review a recent proposal for onset geminates in Dutch. This analysis will turn out to be useful to capture the phonological favoring effect on stuttering in section 4.

3. Onset gemination

In Postma & Ooijevaar (2014), onset geminates were studied in Dutch interjections, such as *bah!* (interjection of disgust) and *goh!* (interjection of awe). These interjections have lax vowels in open syllables, which is otherwise forbidden in Dutch. The coda is, therefore, monomoraic, which is compensated by onset gemination. This gemination is obligatorily present. This suggests a potential moraic weight of the onset in Dutch. The problem is that most theories of phonological weight do not assign weight to onsets (Hayes 1989). Instead of *ad hoc* allowing for moraic onsets, for a small set of languages (Topintzi 2004, 2008) or for interjections, Postma & Ooijevaar assign the structure in (9b), which complies to the standard representation of ambisyllabic geminates (Hyman 1985, Hayes 1989), now applied to onsets, (cf. also Hayes 1989:302). This is done elegantly using Lowenstamm's onset clitic hypothesis (Lowenstamm 1999), recaptured in moraic theory.

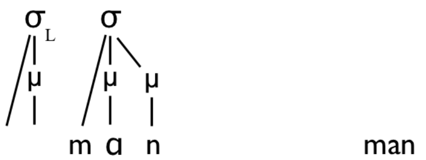


(9) a. b.

In (9a), featuring an ordinary word *bak* 'container', there is an leading CV structure, which remains unexpressed as it is outside the parsing domain. Lowenstamm's hypothetical leading CV allows us to subsume onset geminates to standard cases of coda-onset geminates, as is the standard account in phonology (cf. Hayes 1989; Morén 2001). Consider (9b). The interjection /bah/ has an illicit /h/ coda, closing off the syllable necessary to license the lax vowel. This /h/ is illicit as a coda consonant in Dutch. By remerging this illicit /h/ to the onset of Lowenstamm's onset clitic position, a moraic onset is created, as the related mora cannot remain unparsed. Since the clitic position is without vowel, the /b/ of /bah/ obligatorily spreads to this mora. The geminate [b:] is therefore not a result of a geminate onset, but a consonant tied to a preceding mora. The precise structure of P&O's argument is irrelevant now, but it is important to note that direct gemination of the onset is excluded in Dutch because there are OCP-effects in onsets, e.g. [tl-] and [dl-] are, for instance excluded, while [tr-]and [dr-] do occur in Dutch. Similarly *[pw-] and *[bw-] are excluded in Dutch, while [kw-] and [tw-]/[dw-] do occur (Van Oostendorp *pers. comm*, Linke 2014). Onset geminates must, therefore, come about by another mechanism than doubling the onset. They realize as a ambisyllabic consonant of which the preceding syllable is a relaization Lowenstamm's pre-word structure. In the next section we will apply it on stuttering.

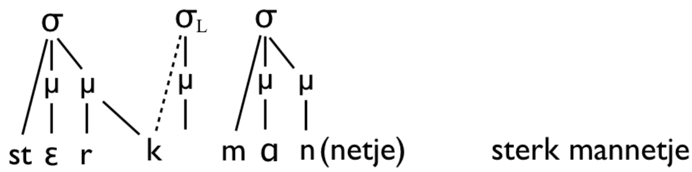
4. Theory

As we have seen in the previous section, the left-edge of a word is marked by a leading CV. This is the CV-clitic hypothesis by Lowenstamm (1999). We choose for a moraic implementation as in (9), along the lines of Postma & Ooijevaar. The Lowenstamm syllable is indicated with σ_L.



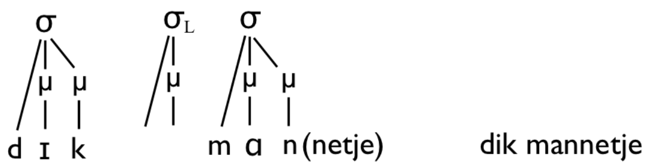
(10)

In many languages of the world that have a predominant CVCV structure, this leading σ_L is not standardly parsed. In languages like Dutch, on the other hand, this leading structure is used to realize complex codas, as illustrated in (10).



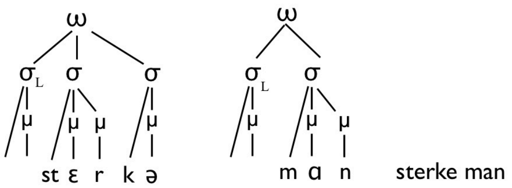
(11) a.

Words with complex codas are, therefore, parasitic on the Lowenstamm structure of the word that follows. Languages differ in allowing empty moras in otherwise parsed prosodic structure. Italian does not allow empty moras, while Dutch and Germanic language in general, do allow for it. Words without complex codas leave this Lowenstamm structure unused, i.e. unparsed, as in (11).

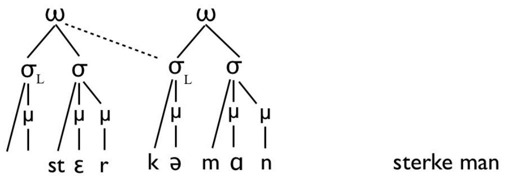


(11) b.

In Dutch, on the other hand, the Lowenstamm CV is sometimes parsed and segmentally licensed, sometimes it is left unparsed, and remains without segmental material as in (12a). It is possible that word final open CV structures occupies Lowenstamm's CV in fluent speech as in (12b), where the suffixal syllable /ke/ occupies Lowenstamm's structure. If so, this might be the Dutch counterpart of crasis in Portuguese and liaison in French.

(12) a. 

(staccato speech)

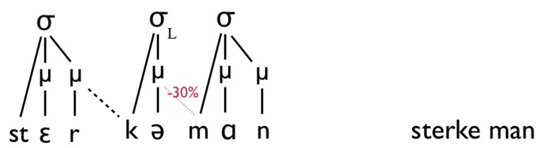
b. 

(bound speech)

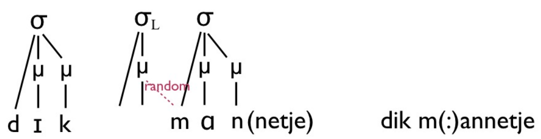
In the next section, we show how Lowenstamm's Hypothesis can explain how codas of preceding words favor and disfavor stuttering.

5. Analysis

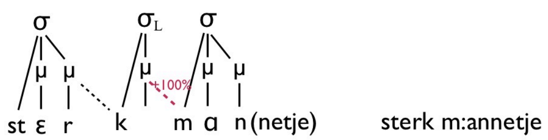
In the previous section, we have seen that there is a Lowenstamm syllable σ_L that must be segmentally licensed in fluent speech, but is left unparsed in staccato speech. Let us now consider the relevant structures taken in fluent speech in PWSs. Let us assume that PWSs have stricter licensing conditions on moras. This strict parameter setting is the default option in natural language, but a Dutch PWS has to pronounce Dutch, which requires a more marked parameter setting. The relevant structures are given in (13a-c). Consider first (13c), where the preceding word *sterk* 'strong' has a complex coda. The /r/ is licensed moraically under weight by position (Hayes 1989). The /k/ could make it superheavy but can instead link up with Lowenstamm's syllable. If so, the mora of this syllable is not segmentally licensed, unless the onset consonant of *man* links up with the mora. As we have seen in section 3, this realizes as a geminate. In this perspective, stutter is a licensing strategy of an empty mora.

(13) a. 

'strong man'

b. 

'thick man'

c. 

'strong small-man'

In the structure in (13a) where the preceding word ends in a CV, no such moraic linking of /m/ is possible. This is the stutter disfavoring structure. In (13b) we have two possibilities: we can leave Lowenstamm's syllable unparsed or parse it. In the former case, no onset-coda linking is possible, in the latter case onset-coda linking is obligatory. In (13c) spreading of m is needed to license the empty mora. σ_L cannot be left unparsed because of the ambi-morphematic /k/. From this perspective, core stuttering is a result of enhanced parsing of prosodic material. For further reference we summarize this finding in (14).

(14) [- stutter]: a parsed σ must be licensed by onset **or** coda
[+stutter]: a parsed σ must be licensed by onset **and** coda

In the ideal case, this is the only dimension in which stuttering differs from non-stuttering language.

6. CV stutter
The second type of stutter that we would like to discuss is the reduplicative stutter of the cv-type. An example is given under (15).

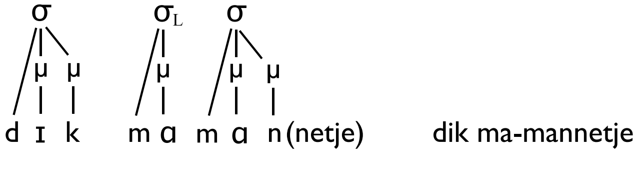
(15) ik kan **be-bijvoorbeeld** voor-voor een optreden op de radio... (3:59)
I can for instance before an performance on the radio ...

We see that *bijvoorbeeld* [bəʋo:rbe:lt] 'for example' is subject to stutter of a different type: the first CV [bə] is reduplicated. This type of stutter is especially interesting as it shows that it is not the pronunciation of the onset per se that is problematic. It is even pronounced twice in (15). From our moraic perspective, however, where words have a Lowenstamm clitic position marking their left-edge, which in certain circumstances must be segmentally licensed, the doubling becomes understandable. Instead of linking the onset to the preceding mora, the onset spreads to the other onset while the mora is lexicalized by a schwa or by copying the vowel, probably in an autosegmental way. This is a standard reduplication strategy in natural language. In the table below, I give all occurrences of this type of stutter in the fragment. In most of the cases, it is triggered by a leading (v)vc.

(16)

leading string	code1	stutter string	code2	tape in sec
omdat je (pause)	-	i-ik zie [ʔɪʔɪk]	cv2	5:34
erom heen	vvc	toetoen [tutun]	cv2	2:21
ja, het	vc	wawas [ʋɑʋɑs]	cv2	2:36
ik kan	vc	be-bijvoorbeeld [bəbəforbelt]	cv2	3:59
opladen	vc	da-dat [dɑdɑt]	cv2	4:06
ik	vc	zi-zie het [zizi]	cv2	5:18
in	vc	a-andere [ʔɑʔandərə]	cv2	5:23
dat ik	vc	i-ineens [ʔɪʔɪnens]	cv2	6:01
nee, nou	vv	zo-zorg [zɔzɔrx]	cv2	8:08
heb ik	vc	na na het [nanaət]	cv2 or w2	2:22
heus wel	vc	da-d:at [dɑd:ɑt]	cv2 , c:	8:17
nou	vv	dadat-dadat [dɑdɑtdɑdɑt]	cv2; foot2	4:24
alleen maar	vvc	me-me-m'n [məməmən]	cv3	7:28
pause	-	da-de-dat [dɑdədɑt]	cv3	

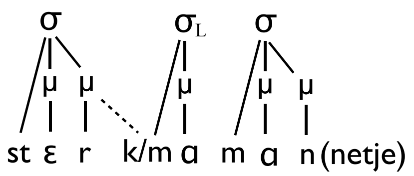
In (17) I give the structure in terms of our "mannetje"-example. The CV reduplication occupies the onset and the nucleus of the Lowenstamm syllable.



(17)

If this is correct, we make a prediction. We predict that CV-stutter will not occur after complex codas, as the complex coda and the reduplication compete for the same slot. This is illustrated in (18).

(18)



*sterk ma-mannetje

This is a correct prediction, at least in the fragment we have used in this study. No cases of CV stutter after cc-codas are found. The data are given under (19). The cases indicated with cv all concern a long v: or a diphthong. We do not know if this is accidental or not.

(19) *CV stutter (type* ma-mannetje)

coda	CELEX	stutter	expected	effect
cv	28%	0	3.6	100% down
vc	57%	11	7.4	50% up
cc	11%	0	1.4	100% down
vv	7%	2	3.5	40% down

Although v-codas occur 3 times less than c-codas, the occurrence v-codas in the set of cv-stutters is more than 5 times more. This means that a leading v-coda strongly disfavors reduplicative stutter in the next word (p value=0.000001).

7. Syntactic stutter

There is an interesting context where stutter occurs after pauses. This is problematic from the account given thus far. The theory that the coda of the previous word is the cueing factor, is correct, does not predict stutters after pauses. We list the relevant instances in the fragment.

(20) a. In mijn eh jonge jaren -- jeetje, je begint meestal zo-zo'n jaar of op je-op je achste jaar of zoiets mee... zevende eh --
in my young years – jesus, you usually begin on your eighth year or so, seventh –
d:eed ik het wel 's meer dan nu
(then) did I it PRT PRT more than now

b. Ik (pause) **zi-zie** het ook aan anderen, hè?
I, (pause) (I) see it also in others, don't I
'I see it in others, too'

c. (pause) **d:enk** toch een-een vrees dat het erg is
(I) think really a-a fear that it horrible is
'I do think it is a fear that it (i.e. stutter) is horrible'

d. Voor een optreden voor de radio of wat dan ook, (pause) **k:an** ik ook weer...opladen
before a performace at the radio or whatever, pause, (then) can I also recharge..

e. (After saying that he had a speech at a funeral)
Ja, **w:as** een soort zelfoverwinning
yes (that) was a kind-of selfconquer

f. ik (pause) zi-zie het ook aan anderen, hè?

I (pause) (I) see it also in others, TAG
'I also see it in other people'
g. **w:as** dat het maar want ik heb-ik heb er nooit lak aan
were that it PRT because I always care
were it like that because I always care

In all these cases the stutter occurs in the V2-position, while its specifier is empty. This comes as a surprise if our claim is correct that the preceding word causes stuttering. Now, Dutch has a set of topic pronouns, such as *die* 'he/she/they', *dat* 'that/it', *dan* 'then', *daarom* 'therefore', *toen* 'then', etc. which resume an earlier mentioned topic, as in (21a). This topic pronoun can be dropped if and only if it has fronted to the sentence peripheral position, as in (21b).

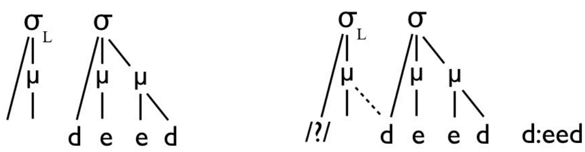
(21) (topic: John got very angry)
a. Ik heb **dat/*ø** helemaal niet gemerkt (topic drop is blocked)
I have that at-all not noticed
b **Dat/ø** heb ik helemaal niet gemerkt (topic drop is optional)
that/ø have I at all not noticed
'I did not notice it at all'

Topic drop in Dutch is only possible if the topic has been mentioned immediately before by the speaker himself in a left-dislocation, or by his interlocutor in the preceding utterance.[9] Despite the name, topic drop is usually analyzed as an empty variant (*pro*) of a lexical pronoun (Van Garderen & Postma 1981, Weerman 1992). Apparently, such a phonetically null pronoun (pro) must be licensed under adjacency ("proper government" (Chomsky 1981)). Any pause (apart from the pause during speaker change), or any other material such as a coordination makes the choice of the lexical topic pronoun obligatory.

(22) (topic: John got very angry)
c. En dat/*ø heb ik helemaal niet gemerkt

From this perspective, most of the empty topic pronouns under (20) are not well-formed in target Dutch.[10] In (20a), for instance, the topic (*in mijn jonge jaren*, 'in my young years') and the empty pronoun has been interrupted by a parenthetic thought marked by -- ... -- in (20a). Nevertheless, the speaker does not resume with a full lexical pronoun *dan* 'then'. Apparently, some other strategy is pursued by this speaker. Lowenstamm's clitic hypothesis sheds light on this issue. Notice first that the speaker stutters in these clauses. Let us assume that the speaker of the sentences under (20) does not consider *pro* completely phonetically null.[11] This would immediately account for the fact that these pronouns escape the adjacency requirement of *pro*. But if the pronoun is not completely null, it is a clitic, which links up with Lowenstamm's structure immediately preceding the verb *deed* in (20a). We have illustrated the two structures of a fluent speaker and a stutterer under (23).

[9] Another pronoun that participates in this topic-drop pattern is the first person pronoun (*ik* 'I').
[10] The optative case in (19g) is of slightly different nature. There is an empty operator in specCP rather than a topic pronoun.
[11] A wild hunch is that it is an element that carries the feature [spread glottis], i.e. [h], as in (8) above.



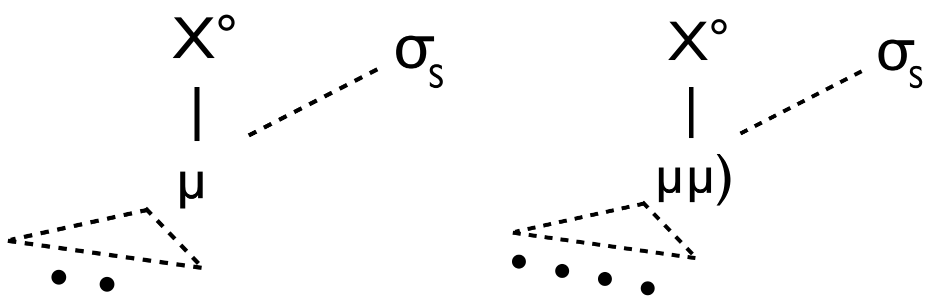
(23) a. b.

In target speech, Lowenstamm's structure σ_L is simply not parsed segmentally. In this long-distance topic-drop speech, however, the subject is not completely empty, though without weight. It then cliticizes to the Lowenstamm structure σ_L. In the configuration under (23b), the Lowenstamm structure σ_L cannot be left unparsed and, hence, must license its moraic weight. As a rescue strategy, the onset /d/ links up with the previous mora and geminates. We see that general requirements on Dutch topic pronouns, together with a deviant organization of the topic resumption causes stutter in this speaker.[12]

7. On the nature of σ_L

In this section we develop an interface account of the Lowenstamm structure σ_L. Where does it come from? When is it inserted? In our analysis of stuttering we have made use of σ_L in an essential way. More knowledge about its status within the organization of the grammar is, therefore, essential to make theoretical and empirical predictions. The original idea of the Lowenstamm's clitic CV is to mark the left-edge of phonological words. In this way, the use of diacritics such as # can be avoided. Moreover, the CV-structure predicts the interaction of the word's CVs with the left-edge by general laws that rule the interaction of CV. Finally, this CV can host clitics that remain under the 2μ word length. So we may say that the original proposal claims a left-edge Lowenstamm CV to phonological words, while elements below the phonological word level are parasitic to σ_L's of other words. This implies that clitics do not have a σ_L at their left edge, if they have independent prosodic space at all. So, we cannot say that all syntactic terminals have a CV edge: simple clitics do not.

Apart from empty categories, we have the representations of lexical material upon insertion at the syntax-phonology interface, as in (24).



(24) a. b.

[12] It is possible that problematic cases with stutter but without obvious phonological trigger for it, as in (i), subsumes under this explanation.

(i) toen we de eerste keer m:et eh met elkaar praatten, ø **v:iel** ik (...) op je. [pratə f:il]
when we the first time with each other talked, then fell I (...) on you
'When we talked with each other the first time, I fell in love'

In (24) we have a syntactic terminal symbol X into which a phonological stucture is inserted. This is drawn in the vertical plane. Insertion of phonological material is drawn in the horizontal plane, drawn in perpective using dashed lines. The interface is a prosodic layer: here the moraic layer has been chosen. We have illustrated two structures: in (24a) a monomoraic clitic, in (23b) as one-syllable structure above the minimal phonological word level of two or more morae. The half-open bracket[13] μμ) indicates that a lexeme is at least bimoraic, but can be bigger, e.g. (μμμ), which complies with μμ). Further on, as a shorthand, we will write both phonological and syntacic relations vertically, as in (25a).

(25a)
$$\begin{array}{c} X \\ | \\ \mu\mu) \\ | \\ \sigma_S \end{array}$$

b.
$$\begin{array}{c} X \\ | \\ (\mu\mu \\ | \\ \sigma_S \end{array}$$

Notice that the structure in (25b) is identical to the one in (25a) as no linear ordering is defined yet. The "mobile" in (24) can rotate freely.

Thus far we represented phonological material that realizes syntactic structure. Lowenstamm's σ_L is of a different nature. Its insertion is purely phonological and cannot see the syntactic structure: it is merely dependent on the phonological module. We assume a post syntactic insertion rule as in (26).

(26)
$$\begin{array}{cc} \mu\mu) & (\mu\mu \\ | & | \\ \sigma_S & \sigma_S \end{array} \longrightarrow \begin{array}{ccc} \mu\mu) & (\mu) & (\mu\mu \\ | & | & | \\ \sigma_S & \sigma_L & \sigma_S \end{array}$$
(Lowenstamm insertion)

Combined with the structures under (25), the Lowenstammian context is realized as in (27).

(27)
$$\begin{array}{cc} X & Y \\ | & | \\ \mu\mu) & (\mu\mu \\ | & | \\ \sigma_S & \sigma_S \end{array} \longrightarrow \begin{array}{ccc} X & & Y \\ | & & | \\ \mu\mu) & (\mu) & (\mu\mu \\ | & | & | \\ \sigma_S & \sigma_L & \sigma_S \end{array}$$

Notice that Lowenstamm's insertion of (26) is exclusively phonological: when applied in (27), the syntactic terminals X and Y are invisible to it. Finally, it is important to notice that σ_L insertion only occurs once. After insertion, the context of the insertion rule is not present anymore. The rule is not recursive. We will return to this. Notice that at the phonological side of the interface, σ_S and σ_L are defined not by their realizing a syntactic node or not, but by the moraic representations (mono-moraic versus poly-moraic). In the next section, we will show that enhanced levels of stuttering comes about by a minimal variation on this insertion pattern.

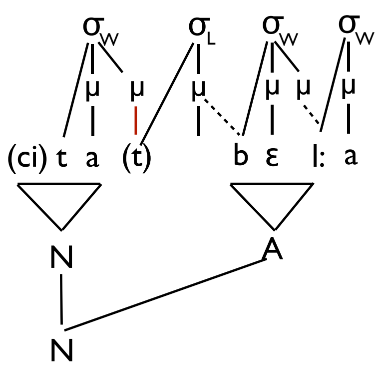
[13] The bracket are of course closed in the representation. We take a condition on a rule or a context of a rule as a kind of 'string search'. So μμ) 'finds' (μμ) and (μμμ), but not (μ).

8. Beyond core stuttering

In the previous sections, we analyzed core stuttering as a more restricted parameter sertting on the licensing of empty codas. As we saw, the typical c:-stutter comes about when a partial occupation of σ_L occurs by coda material of the previous word. In that case, σ_L may not be left unparsed. The empty mora in σ_L must find a licensing, e.g. by linking the onset to it (gemination). Why, then, does this happen in persons who stutter (PWSs) and not in persons who do not stutter (PNSs)? Various causes come to mind: it is possible that a PWS has difficulty to sometimes parse σ_L and sometimes not parse σ_L. He chooses for always parsing it, which results in stuttering in those cases when σ_L is empty, but not if weak codas have spread to it. This would put its cause outside language proper, for instance in the processing difficulties in PWSs. If so, it must be proved that a PWS has more limited processing capacities than PNS. It classifies stuttering as a mental deficiency. This is the standard view on stuttering, though the literature has difficulty to pinpoint where this deficiency is located. The strictly language-internal, phonological approach pursued in this paper, on the other hand, suggests a less dramatic cause. If a PWS does not allow for heavy codas in word syllables σ_W (which is a very comon restriction in many languages in the world), but is acquiring a lexicon that contains many heavy coda lexemes, the PWS must use σ_L of the next word to realize such heavy codas. Put differently, a PWS has a deviant parameter setting for the target language he is acquiring. Non target-like parameter setting is very common in the acquisition of natural language. In most of the cases and as long as the acquirer is below the maturation age, non-target parameter settings can be reset. Incidentally, however, the deviant parameter setting remains and causes a change in the language. This is how language change comes about (Lightfoot 2006, Kroch 1989).

A change thus described is necessary to generate stuttering but not sufficient. It only induces geminiation in onsets of the type of Italian *raddoppiamento fonosintáttico*.

(28) a. cittá bela (with underlying citta(t))



b.

c. all consonants must be linked to an onset (*raddoppiamento fonosintáttico*)

If (28c) holds in languges with consonantal codas, the consonants after tonic positions cannot be realized and link up with the σ_L of the subsequent word (Postma & Ooijevaar 2014). So what is stutter in addition to (28b)?

The additional property can be traced if we consider (1f) repeated here as (29).

(29) toen kwam mijn **m:oe-moeder** te overlijden
then came my mother to die
'then my mother happened to die'

In these cases, (i.e. [d:ɑdɑt] and [m:umudər]), **both** a reduplication of the initial consonant **and** a CV stutter is realized. This means that σ_L can be iterated by a person who stutters. In fact, it is a well-known property of stuttering that it has multiple realization of onset material, up to 4 or 5 times, and even with the possibility that the PWS "remains in it". Where does this recursion in σ_L come from? Now, in contrast to say Italian speakers, PWS may have gemination in light syllables, as illustrated in (30abc), as well as gemination after a light preceding syllable, as illustrated in (30d), although the latter occurs with low frequency. In (30ef) we even see stutter in clitic-like or inflectional material.

(30) a. dat het dan **v:er**dwijnt [dɑn-ʋ:ər'dwɛⁱnt]
that it then disappears
b. i-ik kan **be-bij**voorbeeld optreden [kɑn-bəbə'ʋorbe:lt]
I can for instance perform
c. ik **b:e**steed er ook altijd aandacht aan [ɪk-b:ə'ste:t]
I pay there PRT always attention to
'I always pay attention to it'
d. ... jij staat dat meisje daar in te pakken **m:et** dat gehakkel van je [pɑkə-m:ɛt]
you are that girl there flirting with that stutter by you
'you are flirting that girl with your stuttering'
e. als ik bijvoorbeeld op de radio geweest ben of op **d:e** teevee [opt:əteve]
when I for instance on the radio been am or on the television
'when I have been on the radio or on television'
f. die voor een groep d:ors**t:en** te staan[14] [d:ost:ətə]
who in-front-of a group dared to stand
'(people) who dared to stand in front of a group'

This indicates that the insertion of σ_L not only occurs before morphemes with the minimal word length of two moras, but also before weak syllables, such as weak prefixes and articles. This means that weak sylables must have a Lowenstamm σ_L preceding. Apparently, PWSs deviate from PNSs that they have (31) instead of (27).

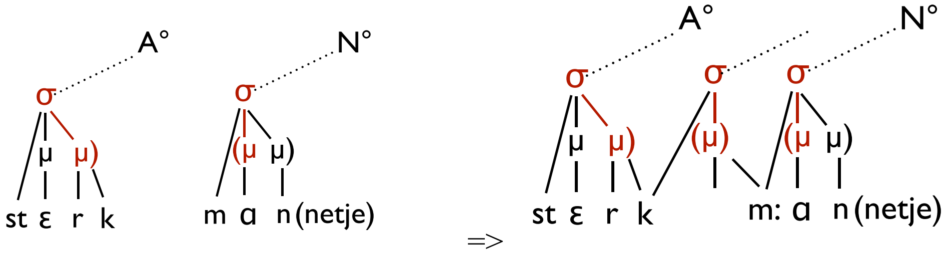
(31)
$$\begin{array}{cc} X & Y \\ | & | \\ \mu) & (\mu \\ | & | \\ \sigma_S & \sigma_S \end{array} \longrightarrow \begin{array}{ccc} X & & Y \\ | & & | \\ \mu) & (\mu) & (\mu \\ | & | & | \\ \sigma_S & \sigma_L & \sigma_S \end{array}$$

Notice that the sublabel in σ_S is just mnemonic for the reader that these are linked to a syntactic node X and Y, as in (27), but the phonological insertion rule of Lowenstamm's syllable, being phonological, is not sensitive to this distinction. We should think them away in the rule application.

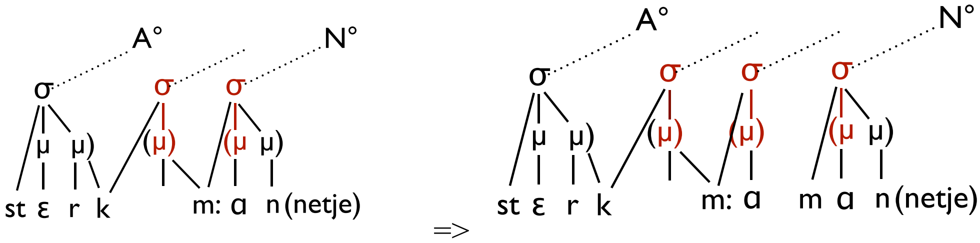
Look now at a special property of (31). In contrast to (27), the insertion in (31) has a symmetry in σ_L and σ_S. The insertion is, therefore, *recursive*, since the rule can apply to the output again. It generates structures with recursive phonological material that is not syntactically licensed. We have illustrated two recursion steps in (32a) and (32b). The input of (32b) is identical to the output of (32a). The context of the rule application is marked in red.

[14] (28ef) are parallel in that the stutter resides within the unstressed syllable of a word, while the next unstressed syllable is one that have a repetition in the target speech.

(32) a.



b.



The first step in (32a) is present in the syntax of PNSs and PWSs alike, though the conditioning context is deviant in PWSs. The second step is exclusively possible in the syntax of PWSs. The output of (32b) is again a possible context to insert an extra syllable, and so forth. The *insertion* of phonological material that is not syntactically (and semantically) licensed, has the advantage of linearizing a string with respect to the left-right ordering. The *recursion* of this syntactically unlicensed phonological material, however, has only communicative disadvantages for PWSs. But this does not imply (though it may be true) that PWSs necessarily have any mental, psychological, or processing defects. PWSs have only ended up with a parameter setting that is functionally non-optimal, but it is fully a possibility of natural language, not a pathological deviation from it, at least from a formal point of view. The analysis given clearly shows that natural language is not designed for optimal communication, in principle. Its communicative property is an irrelevant aspect within the formal discription of natural language. Its *use,* on the other hand, is a strong force not to opt for (31) instead of (27), but this bias is functional and not part of the core grammar.

8. Consequences

The idea pursued in this paper, that stuttering finds its cause in a parameter setting of natural language has far reaching consequences. It has consequences for our thinking about stuttering as well as consequences for the theory of grammar. To begin with the first, the theory presented situates stuttering in the normal domain of linguistic variation. This has consequences for the strategies stutterers might pursue to handle their language. If a paremeter setting is involved, stutterers have equal chance to change their language (reset their linguistic parameters) as that an Italian, whose parameter setting requires vocalic licensing of final syllables, will speak a Germanic language with a full set of final consonants: he may adequately suppress epenthetic vowels, but he wil never completely get rid of the tendency to insert them. Before the maturation persiod, on the other hand, parameter resetting is possible, and new strategies might get developed to facilitate resetting effectively. Another consequence might be that stuttering is limited to the language the PWS is a native speaker of. This is not an immediate consequence, however, as some languages, like Dutch and English might be similar or identical on point of the parameter setting involved.

There are also consequences for our reflecting on the theory of grammar. If stuttering is part of the normal language variation, our language faculty apparently allows for disfunctional paremeter settings. Apparently not all realizations of universal grammar are functionally equal. Recursion in phonology is as much an option of our language faculty as recursion in syntax, but phonological recursion is disfunctional because phonology does not have an interface with the semantics. Consequently, the set of natural languages cannot be emerged as a consequence of just functional pressure: the language faculty is not shaped functionally but rather has formal properties that allows for realizations, functional and disfunctional alike.

9. Conclusions

In this empirical study, we have shown quantitatively that stutter is phonological: stutter is significantly favored by consonantal preceding codas. Stutter is triggered by unlicensed coda material, which spreads to the onset the word preceding CV, postulated in Lowenstamm 1999, for which we created a moraic implementation, σ_L. This σ_L may be left unparsed, but whenever parsed, it must be fully parsed: this gives rise to consonant reduplication. Lowenstamm's structure inserts phonological material that is not licensed by syntactic structure, but by an insertion rule that mark word boundaries. A paramenter setting that conditions this insertion (between bimoraic syllables) is relaxed in the grammar of PWSs, which causes potential recursion in this Lowenstamm material. The parameter setting that defines this specific realization of UG should rather be described as a *language* with stutter (LWS) rather than attributig it to the *speech* of a person who stutters (PWS). We conclude that different parameter settings do not result in languages that are functionally equivalent, although they are formally equivalent realizations of UG.

References

Baayen, R. Harald; Richard Piepenbrock; and Léon Gulikers. (1993). The CELEX lexical database [CD-ROM]. Philadelphia, PA: Linguistics Data Consortium, Universityof Pennsylvania. Online: http://celex.mpi.nl

Boersma, Paul, and David Weenink. (2011). *Praat: Doing phonetics by computer.* Version 5.2.32. Online: http://www.praat.org

Choi, Dahye (2013). Behavioral inhibition and childhood stuttering. *Journal of Fluency Disorders*. 38, 171–183.

Garderen, René & Gertjan Postma (1982). *Het verzwegen eerste zinsdeel in het Nederlands - een geval van kleine pro? - werkstuk lege plaatsen onder begeleiding van Hans Bennis.* Ms Free University.

Hayes, Bruce. (1989). "Compensatory lengthening in moraic phonology". Linguistic Inquiry 20. 253-306.

Huinck WJ, van Lieshout PH, Peters HF, Hulstijn W. (2004). Gestural overlap in consonant clusters: effects on the fluent speech of stuttering and non-stuttering subjects. *Journal of Fluency Disorders* 29:3-25.

Hyman, L. (1985). A theory of phonological weight. *Publications in Language Sciences* 19. Dordrecht: Foris. [Reissued 2003, University of Chicago Press].

Jongenburger, Willy & Vincent van Heuven, (1991), 'The Distribution of (Word-Initial) Glottal Stops in Dutch'. In: F. Drijkoningen and A. van Kemenade (eds.), *Linguistics in the Netherlands* 1991 (Amsterdam: Benjamins), 101-110.

Kok, Auke (2009). NOS interview.

Linke, Katrin (2014). Onset: sequences of two consonants. In: *Taalportaal - the linguistics of Ducth and Frsian online*. http://www.taalportaal.org.

Lowenstamm, J., 1999. The beginning of the word. In: Rennison, J., Kühnhammer, K. (Eds.), *Phonologica 1996.* Holland Academic Graphics, La Hague. 153–166.
Marshall, Chloe. (2005). The impact of word-end phonology and morphology on stuttering. *Stammering Res.* 2005. 375–391.
Morén, Bruce. (2001). *Distinctiveness, coercion and sonority: A unified theory of weight.* New York and London: Routledge.
Nevins, Andrew & Nina Topintzi (2014). *Moraic Onsets and Cross-Anchoring in Arrernte.* Ms. University Tessaloniki.
Nippold, Marilyn. 2002. Stuttering and phonology: Is there an interaction? American Journal of Speech - Language Pathology. 11. 99-110.
Viswanath Nagalapura S. & Amy T. Neel (1995). Part-Word Repetitions by Persons Who Stutter Fragment Types and Their Articulatory Processes. Journal of Speech, Language, and Hearing Research 38, 740-750. doi:10.1044/jshr.3804.740
Ntourou, K., Conture, E. G., & Lipsey, M. W. (2011). Language Abilities of Children Who Stutter: A Meta- Analytical Review. *American Journal of Speech-Language Pathology*, 163-179.
Postma, Gertjan & Ooijevaar, Etske. 2014. Geminate onsets in Dutch interjections. Ms Meertens Institute.
Postma, Gertjan & Tobias Scheer. (2014). Hèh? [hɛ↗], and hé! [he:↘] in Dutch - Synchronic Linguistic Competence or Convergent Evolution? - A reply to Dingemanse et al. 2014. Ms Meertens Institute.
Reese, Sheree & Joseph Jordania (2001). *Stuttering in the Chinese population in some South-East Asian countries: A Preliminary investigation on attitude and incidence.* ms. Minnesota State University.
Sayer, Jean. (2010). *By the Numbers: Disfluency Analysis for Preschool Children who Stutter.* Ms. Illinois State University.
Topintzi, Nina. (2006). Moraic onsets. Dissertation, University College London.
Topintzi, Nina. (2008). On the existence of moraic onset geminates. *Natural Language & Linguistic Theory* 26. 147–184.