A CVX theory of Dutch Syllable Structure

This paper investigates the extent to which the syllable structure of Dutch conforms to CVX theory as stated by Duanmu (2002) and Borowksy (1989). One challenging aspect of the project stems from the difference in quality between the Dutch non-high long vowels and short vowels. Two theories are considered regarding this issue: Working under the assumption that the difference in quality is based on a distinction of tenseness and not length, I will be considering whether standard Dutch conforms to CVX syllable structure for non-final syllabic positions, as has been shown for English. This will be accomplished through a restriction of the data set which treats the vowel distinction as tense lax -variants of the same vowel, and through the exclusion of potential violators on independently motivated grounds. Diphthongs remain a contentious issue.

1. Introduction

The goal of this investigation is to provide evidence towards a universal theory of syllable structure as formulated by Duanmu (2002) following Borowksy (1989).
The proposal that medial position syllabic rime is restricted to two slots maximally was demonstrated to be a tenable view in Borowsky (1989). In a departure from the more traditional viewpoint that onset clusters are constrained by principles of sonority, Dunamu (2002) proposes that complex onset clusters in English and Chinese are composed of one slot. Together, the two theories posit a CVX structure for syllable structure. The question remains as to whether this claim can be shown to hold for languages other than English and Chinese. This paper investigates the extent to which standard Dutch con be shown to conform to CVX theory. In section 2, I introduce VX theory as formulated by Borowsky (1989). In section 3, I introduce Dunamu (2002) and the extension of the theory to CVX.
In section 4, I present both sides of the tense/lax versus length vowel debate for Dutch, a pertinent issue in the consideration of our theories application. In section 5, I detail the methodology used in this study. Finally in section 6, I present concluding remarks.

2. V(X) theory

Borowsky (1989) considers the composition of the syllable rime in English, noting that there exists an asymetrical distribution between rimes consisting of two slots and rimes consisting of more than two slots. This asymmetry can be explained, following Siegel (1974), by positing the existence of two levels of morphological affixation, level 1 and level 2. Under the Level Ordering Hypothesis, class 1 affixes, which trigger and undergo phonological processes, undergo affixation at level 1and can attach to free as well as bound morphemes. This type of affixation occurs first. Class 2 affixes, which are phonologically inert, undergo affixation at level 2 and thus can only attach to free morphemes (derived or un-derived words). Below are examples of English affixes classified accordingly (Spencer 1991):

1) Examples of English Affix Classification:
   Class I suffixes: +ion, +ity, +y, +al, +ic, +ate, +ous, +ive
   Class I prefixes: re+, con+, de+, sub+, pre+, in+, en+, be+
   Class II suffixes: #ness, #less, #hood, #ful, #ly, #y, #like
   Class II prefixes: re#, sub#, un#, non#, de#, semi#, anti#

Medial syllables of the form VVC and VCC in un-derived mono-morphemic words and before the attachment of level 1 affixes are very rare. Borowsky proposes that those with more than two slots are limited to word edges, indicating a strict constraint on the syllable rime structure in English at level 1. The strongest evidence in support of this theory is that in English, vowels in medial positions shorten at level 1:

2) /min/ +/t/ → /mɛnt/
mean meant

This observation assumes that long vowels take up two slots in English, which, as is discussed below, will not be the assumption made in this investigation of Dutch. The important principle we are testing is that, regardless of what parameter a language takes in defining what constitutes a filled slot, the number of available slots remains universal and constantly V(X), with V being a vowel and X being a variable consonant or vowel.

Borowsky (1989) states the Principle of Structure Preservation:

3) Structure Preservation:
language-particular structural constraints holding for underlying representation hold also for derived representation, and vice versa.

This enforces conformity with the Coda Condition. After Level 1 Structure preservation is inactive and syllable structure less restrictive, allowing longer codas and making vowel shortening unnecessary. Thus, only the phonology of level 1 is to be taken as structure preserving.

3. Extending V(X) to CVX

(Dunamu 2002) extends the Borosky (1989) analysis of medial rime constraints though an investigation of the possible constituents of onset clusters. Traditional accounts of onset clusters are reliant upon the sonority scale. Duanmu notes the following problems with such an account (for English): Clusters which appear to violate the MSD are allowed in English i.e. [lj], English allows clusters composed of sounds in the same place of articulation i.e. [tr] [dr], and many sounds that satisfy the MSD don't appear in English.

Duanmu posits a single slot analysis of onset clusters in which occurring clusters are complex single sounds. Necessary to this are two assumptions: first that words around a syllable need not be a proper part of it (extra-metricality), and second that if a cluster is a single sound it is a good sound, and single sounds must adhere to the No Contour Principle (Duanmu 1994) stated in (4):

4) *N — [αF] [-αF]

N = any node
α = any value
F = any feature

5) schraalhans [sxra:l][hAns] (meaning unknown) from the Dutch lemma

90908	schraalhans	Y	4	[CCCVVC][CVCC]	[sxra:l][hAns]	'sxral-hAns

represented as: [schr][aa][l].hanns or [s][chr][aa][l].hanns

Example (4) demonstrates that for a complex sound to be a good sound no single articulator can simultaneously produce opposing values of the same feature. In (5), [schr], or [chr] with extrasyllabic [s], is treated as a complex onset occupying one slot only, [aa] is (crucially, for this study) treated as a single slot in the nucleus, and [l] is the coda.

Building on Borowsky's V(X) model, Duanmu thus proposes an articulator based feature theory of syllable structure in which non-final syllables are maximally of the form CVX. The question remains as to whether such a theory can be proven to hold universally, which is the underlying goal of this study.

4. The representation of vowels: a debate

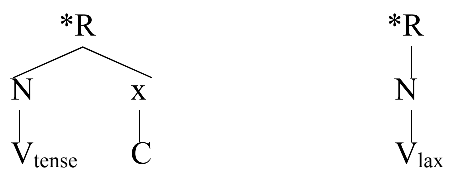
As noted above, in order for CVX theory to hold for Dutch, we must have some account of how vowels are to be represented moraically. Discussion of this matter has polarized in two directions. The first supports the view that they are to be represented in terms of tenseness and the second that they are to be represented in terms of length. Both views are discussed below, but for the purpose of this investigation, the tense/lax view is assumed to be correct.

3.1 The tense/lax approach

In Standard Dutch, phonetic length coincides with phonetic tenseness to a large degree. It has been a issue of debate as to whether the distinction that both seemingly describe is to be attributed to the feature [+tense] or to length as represented by a geminate moraic structure. Van Oostendorp (2000) and Swets (2004) claim that laxness is the phonological feature underlying the contrast between phonetically long tense and short lax vowels. The most persuasive argument for this account relies upon the distribution of stress in the Dutch stress system. a length based analysis incorrectly allows CVV syllables to remain light while CVC are counted as heavy, which is in contrast to a widely held belief that when there exists a heavy light contrast, CVV counts as heavy.

Van Oostendorp (2000) concludes that a branching rime must be headed by a lax vowel. Lax vowels do not occur syllable finally, as the vowel would then occur in a non-branching rime. A branching rime can never be headed b a tense vowel, thus we never see a tense vowel in a CVC - closed rime:

6)



*lan.tarn/lån.tarn *hˆ.at

3.2 The length approach

Booij (1995, 2002) holds that systematic opposition between short and long vowels in Dutch (except for high vowels) can be represented in structural terms rather than via binary feature (such as [+tense]). In a syllable, short vowels can be followed by at most 2 Cs i.e. V(short)CC, and after long vowels only one C can occur i.e. V(long)C (VVC).

Syllable structure is thus held to be more consonant with a theory of phonological length. Under this account, the binary nature of the rime, which consists of exactly two positions, accounts for the distribution of vowels in Dutch.

7)	$CV_{shortlax}$	$*CV_{longtense}CC$
	lˆnt	*li...nt/ *le...nt

d⍰mp [d⍰mp] possible syllable
d⍰⍰mp [damp] not possible

Example (7) demonstrates that in accordance with an exactly two-slot restriction in the rime a long/tense vowel cannot be followed by CC in its coda as this would necessarily result in three slots. Booij proposes that vowel length is not a purely phonetic property and that long vowels are represented as sequences of two identical [-consonant] segments/phonological units/moras. A long vowel, under this analysis, is a single feature bundle linked to two similar abstract representations.

With strong supporting evidence for both theories, we may conclude that at worst, the issue leans favorably towards both sides and is unresolved, and at best it is clearly in favor of there being one vowel representation, which occupies one slot and which varies in the [+tense] feature. On these grounds we can take the important step of justifying the removal of all VV occurrences in the lemma which are non-diphthongal and do not precede a CC+. Both of these qualifications can be taken up for consideration in the final stages of the study.

5. Methodology

5.1 The Dutch lemma

The corpus used in this investigation into the conformity of Dutch to CVX theory was a lemma obtained from the CELEX online database of standard Dutch. The lemma represents headwords in non-reducible forms. As stated in the release notes:

> The Dutch database, version N3.1, was released in March 1990 and contains information on 381,292 present-day Dutch word forms, corresponding to 124,136 lemmata. Apart from orthographic features, the CELEX database comprises representations of the phonological, morphological, syntactic and frequency properties of lemmata. For Dutch and English lemma homographs, frequencies have been disambiguated on the basis of the 42.4 m. Dutch INL and the 17.9 m. English Collins/COBUILD text corpora. (CELEX 1990)

The following categories (Table 1) were utilized in informing the reduction of apparent CVX violations:

Table 5.1.1 Starting categories by column

Fields										
A	B	C	D	E	F	G	H	I	J	K
IdNum	Head	Inl	PhonStrsDISC (word phones)	PhonCVBr (word CV)	PhonSylBCLX (word syllables)	PhonStrsStDISC (stem phones)	PhonStCVBr (stem CV)	PhonSylStBCLX (stem syllables)	PhonolCLX	PhonolCPA

The original 124,136 items in the lemma were then broken down into workable files and lines missing information were identified and labeled as "no sound."

Table 5.1.2 Starting statistics

Total	124136		
	No sound	Sound	All
1-40k	1696	38304	40000
40-80k	731	39269	40000
80k-end	507	43629	44136
All	2934	121202	124136
	2.36%	97.64%	100.00%

5.2 Removal of compounds and no sounds

Compound words may be excluded from analysis. The composition of the lemma is such that it lists, in theory, every non-productive word in the Dutch language. Thus, we may expect that if a word is identified as a compound its compositional parts have been accounted for already. CELEX identifies compounds though the use of both hyphens (-) and word boundaries (#) in their transcriptions. Compound words that somehow escaped CELEX identification require manual identification and removal, with the assumption that their composite parts have also been accounted for in a the lemma.

In addition, there were 2,934 items for which no description existed in the lemma. These were excluded from analysis as no-sounds items, though they deserve further examination.

Working with the entire unreduced lemma, all no sounds and CELEX identified compounds were removed; compounds being identified as an occurrence of + or # in the phonological transcription. After removal of these items 63,152 items remain in the lemma.

Table 5.2.1 Compound and no sound items removed

	start count	"#" and "+"	No sound	Remaining
1-40000	40000	16082	1696	22222
40001-80000	40000	18713	731	20556
80001-124136	44136	23255	507	20374
Totals	124136	58050	2934	63152

5.3 The identification of apparent counterexamples/CVX violations

The next important step is to identify all potential violators of CVX theory, manifested as all possible combinations of VXX super-heavy syllables in column: PhonStCVBr (stem CV). All instances of VXX are indicated in table 5.3.1, however many of these occurred within the same word. Total lines or items containing at least one CVX violation are listed in Table 5.3.2.

Table 5.3.1 Based on H, identify VXX occurences

	1-40k instance	40-80k instance	80k-end instance		
VVCCCC][	0	0	0		
VVCCC][	100	21	15		
VVCC][	1126	773	739		
VVC][	3448	3482	4061	total instances Y	13765
VCCCCC][	0	0	0		
VCCCC][	4	0	0		
VCCC][	277	369	275		
VCC][	2450	2099	2465	total instances X	7939
All	7405	6744	7555		
total instances of VXX (this is not equal to the number of lines containing VXX)					21704

Table 5.3.2 Identify VXX by item

1-40000	6361
40001-80000	5770
80001-124136	6340
Total VXX	18471

5.4 The determination of true CVX violations

Within the 18, 471 potential violators of CVX theory we will be able to exclude many on certain grounds. For native speaker judgments, I enlisted the assistance of two native Dutch Speakers, currently residing in Amsterdam. The speakers were David Lingerak, age 35 and Martine Brinksma, age 33. Both were raised in West-Friesland (Noord-Holland) and are unaware of their particular dialect. the reduction proceeded as follows:

Hyphenated words are removed, ass shown in Table 5.4.1, as they represent non-CELEX identified compounds. 17, 936 words remain.

Table 5.4.1 Remove Hyphenated words

worksheets	start count	instances	lines	Stems Rem.
1-40000	6361	222	194	6167
40001-80000	5770	196	163	5607
80001-124136	6340	203	178	6162
Totals	18471	621	535	17936

Level 2 affixes may be excluded, as they are phonologically inert and do not represent free morphemes. A potential confound may be the inappropriate removal of words with level one affixes. The classification of levels is certainly different in Dutch than in English and has not been examined in this essay. However, I believe the majority will be appropriate by the same reasoning as was applied to compounds. In my first reduction attempt I utilized a list of affixes derived from Booij (2002). This proved useful only in assessing a hypothetical number as I had to rely on statistics due to the error rate during my removal process. The method was abandoned in favor of one that yielded more concrete and tenable results. Searching re-occurring strings of five and six letters at the word edges we are able to reduce the lemma in a conservative yet productive way. Table 5.4.2 shows that after affix removal, only 1,151 words remain as potential violations of CVX theory.

Table 5.4.2 Affix removal

Affixes	number	total removed	total remaining
last 6	13954	13954	3982
last 5	1378	15332	2604
first6	1179	16511	1425
first5	274	16785	**1151**

By the same reasoning detailed above, compounds that were not CELEX identified, but were identified by the native speakers are removed.

Foreign words borrowed by Dutch may, in many cases be excluded, as their composition is purportedly not reducible in Dutch in the same way as in the original source language. for example, item 4008 /airconditioning/ is not reducible to /air/ and /conditioning/ in Dutch. I must acknowledge some degree of error in this removal as it is dangerous to assume that all borrowings are of such a nature and that there may well be overlap in simplest morphologically-free forms.

Borowksy attributes counterexamples found in proper names to the fact that they constitute a deviant subsystem with respect to many phonological phenomena. Assuming that most of them appear to be compounds or foreign, we can deem them derivations of the level 2 variety and exclude them from our list of CVX violators.

Table 5.4 3 Removed speaker identified compounds, names, and borrowed words

Initial amount	1151
removed	738
remaining	413

Of the remaining 413 words in the lemma, 278 were identified as containing long vowels or diphthongs.

Table 5.4.4 Instances of Diphthongs (identification, not removal)

EI AU UI	49 14 20	
total	83	in 82 words

These were marked for closer scrutiny, turning first to the case of long vowels. Assuming that differences in length can be explained by a tense/lax account, we extract all VV that are non-Diphthongal and that do not constitute any further violation of the rime (Non VVCC][or VVCCC][(marked with an X)). VV that meet these two criteria are removable under a theory which assumes long vowels to actually be identical to short with the exception of the feature [+tense].

Table 5.4.5 Remove non-diphthongal and non-VVCC+][VVs

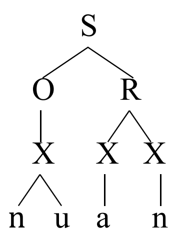
removed VVC][	275
remaining VXX	138

Table 5.4.6 Instances of Diphthongs (all left in lemma)

EI / ´i	49
AU / au / øu	14
UI / √y / œy	20
total	83

Under Dunamu (2002) we can eliminate non-word initial diphthongs, yet word onset diphthongs may still present a problem as there is no way of attributing the initial vowel sound to a complex onset. An example from Chinese follows below:

8)



[n^wan] warm (Chinese)

Following this theory, all non-violating diphthongs are now removed (where a violation consists of non-initial diphthongs and VVCC+ violations) , as shown in Table 5.4.7.

Table 5.4.7 Remove non-violating (non initial diphthongs and YY marked (double VV and VV without VVCC+ violations)

initial VXX	138
removed VXX	62
remaining VXX	76

For the remaining 76 words, a manual lookup was conducted to determine if any further reduction to the lemma could be made. From this 30 words were determined to be compounds, 13 had homorganic nasals which could be taken to constitute a long V and thus together with the vowel only one slot. Finally, 18 were excluded for various reasons examples of which are given below:

9)	70798	ontkleed	un-clad	[VCC][CCVVC]	[Ont][kle:t]
10)	27891	exclameren	undefined	[VCC][CCVV][CVV][CV]	[Eks][kla:][me:][r@]
11)	15617	breeduit	undefined	[CCVVC][VVC]	[bre:d][UIt]

Exclusions similar to example (9) are excluded on the grounds that they contains homo-organic nasals: a nasal followed by a consonant at the same place of articulation, in this case alveolar. In this case, the nasal can thus be assimilated into the vowel sound leaving the second slot in the rime for /t/. Example (10) illustrates a group of words that were extracted on the basis that /ex-/ constitutes a prefix. This may be a slightly contentious point, however it is assumed to hold here. Example (11) represents a class of words that we can exclude if we re-syllabify so that the offending C is relocated to the onset of the final syllable i.e. [bre:d][UIt] →[bre:][dUIt], thus removing the problem of a diphthongal rime (VV). Fifteen violating words now remain, presented in table 5.4.8.

Table 5.4.8 Remaining VXX

6494	astma	problem	[VCC][CVV]	[Ast][ma:]
6686	auctie	problem diph	[VVC][CVV]	[AUk][si:]
6712	Aufklarung	problem diph	[VVC][CCVV][CVC]	[AUf][kle:][rUN]
6715	augment	problem diph	[VVC][CVCC]	[AUG][mEnt]
6745	ausputzer	problem diph	[VVC][CVVC][CVC]	[AUs][pu:t][s@r]
26421	einsel	problem diph	[VVC][CVC]	[EIn][s@l]
26424	einze	problem diph	[VVC][CV]	[EIn][z@]
32072	geijsbeer	problem diph	[CV][VVC][CVVC]	[x@][EIz][be:r]
45326	istmus	problem	[VCC][CVC]	[Ist][mUs]
46265	juncto	problem Non-HomoOrganicNasal	[CVCC][CVV]	[jUNk][to:]
52334	kortjan	problem derived	[CVCC][CVC]	[kOrt][jAn]
73372	ordner	problem	[VCC][CVC]	[Ord][n@r]
82084	punctum	problem Non-HomoOrganicNasal	[CVCC][CVC]	[pUNk][tUm]
87596	rucksichtslos	problem	[CVVC][CVVCCC][CVVC]	[ry:k][si:xts][lo:s]
88468	sanctie	problem Non-HomoOrganicNasal	[CVCC][CVV]	[sANk][si:]

Of the remaining 15 some [VNC] rhymes that could be analyzed as [V~C]. 'einsel' and punctum' may be homorganic, though the /n/ in punctum is not velar and einsel would continue to have a diphthong in its rime with no complex onset to attach to. 'Rucksichtslos' is probably 'reckless', but it has not been confirmed by my speakers.

5. Conclusion

Aside from a few unresolved exceptions, this study has shown that, if we assume a tense/lax and not length distinction, Dutch conforms to CVX theory. Many of these exceptions could most likely be shown to be legitimate exclusions though given current resources I have left them in as violations. The results would also be drastically different, and could be said to not hold true under a different account of vowel length. there would in fact be an additional 275 words to account for. However, given the unresolved state of the data, we have been just as justified in choosing the tense/lax account for the investigation. What remains to be proved is that the same holds true for other, less related languages than those for which similar studies have been conducted.

6. References

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

Booij , G. 2002. The Morphology of Dutch. Oxford: Oxford University Press.

Borowsky, Toni. 1989. Structure preservation and the syllable coda in English. *Natural Language and Linguistic Theory*. 7: 145-166.

Cauberghe, J (ed.) 1949. Kramer's Woordenboek Engels. Turnhout: Brepols

CELEX Dutch database (Release N31) [On-line]. (1990). Available: Nijmegen: Centre for Lexical Information [Producer and Distributor].

Donaldson, B. 1997. Dutch, a Comprehensive Grammar. London and NewYork: Routledge.

Duanmu, San. 2002. Two Theories of Onset Clusters. *Journal of Chinese Phonology* 11: 97-120.

Kager, R. 1989. A Metrical Theory of Stress and Destressing in Dutch and English. Dordrecht: Foris Publications.

Swets, F. 2004. The Phonological Word in Tilburg Dutch Government Phonology and a city dialect of Dutch. Utrecht: LOT 3512 JK

Spencer, A. 1991. Morphological Theory. Cambridge, Mass. and Oxford, UK: Blackwell dictionarys.

Veen, P. A. F. van (ed.) 1990. Etymologisch Woordenboek. Utrecht/Antwerpen: Van Dale Lexicografie.