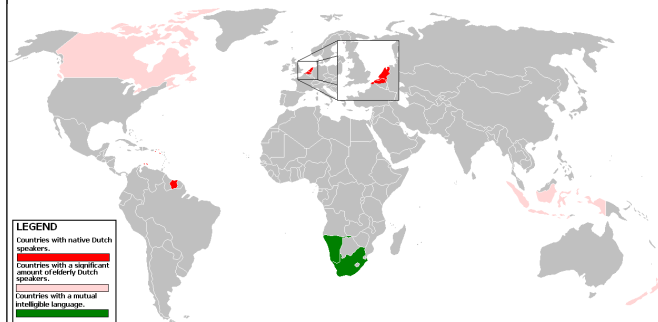


Finding Syntactic Characteristics of Surinamese Dutch

Erik Tjong Kim Sang
Meertens Institute

12 June 2014



Woordenboek van het Nederlands in Suriname van 1667 tot 1876

J. van Donselaar



Finding Syntactic Characteristics of Surinamese Dutch

12 June 2014

Goal

Evaluating automatic methods for finding **syntactic** differences between Surinamese Dutch and standard Dutch

3



Finding Syntactic Characteristics of Surinamese Dutch

12 June 2014

Approach

1. Select two texts: one written in Surinamese Dutch and one written in standard Dutch
2. Process the two texts with a syntactic parser for Dutch
3. Compare the two parse results: can we find frequent syntactic constructions in Surinamese Dutch which are infrequent in standard Dutch?

4



Finding Syntactic Characteristics of Surinamese Dutch

12 June 2014

Technical details

We used the syntactic parser Alpino: a state-of-the-art parser for Dutch

We collected the frequencies of the syntactic constructions and compared these with the t-test:

$$(f_1 - f_2) / \sqrt{f_1 + f_2}$$

where f_1 is the frequency of a construction in Surinamese Dutch and f_2 its frequency in standard Dutch

The higher the t-score, the more unusual the construction is

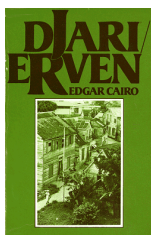
5



Finding Syntactic Characteristics of Surinamese Dutch

12 June 2014

Chosen text for Surinamese Dutch (from dbnl.nl)



...Als je 'em zo zijn straat zag binnen komen fietstrappen! Krio, krio, krio..., niet om naar te luisteren! Volgens zijn stuur ging hij naar links! Volgens zijn wiel reed hij kaarsrecht rechtdoor. Maar Bo's schuinse pet wees weer een andere kant op.
Aaj, ik zag al sins hoelang! Meneertje slaat damp, stinkender dan koeiekak! Hij is dronken als een wát! Na sakasaka ellendeling dati!
Mamsi joeg die kindertjes uit haar weg. Ze had voorop 't erf onder die amandelboom urenlang staan schoonmaken. Eerst zogenaamd erf bezemen. Daarna die kindertjes gras laten trekken...

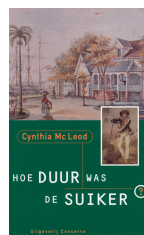
6



Finding Syntactic Characteristics of Surinamese Dutch

12 June 2014

Chosen reference text (from dbnl.nl)



...Aan de voorkant van het prachtige grote huis waren echter niet alle vensters meer gesloten. Op de eerste verdieping was er een geopend raam en daar stond de 17-jarige Elza en keek uit over het groene gazon dat zich voor het huis uitstrekte tot aan de waterkant waarlangs de brede Surinamerivier langzaam stroomde. Een heerlijke ochtend, het begin van een fijne dag. Vandaag 11 oktober 1765 ging de familie naar Joden-Savanna voor de 65e verjaardag van grootmama. Dat zou de volgende dag zijn, de 12e oktober en tegelijk ook de verjaardag van de synagoge op Joden-Savanna...

7



Foreseen pitfalls

Differences between texts may have other causes than language differences:

- author style
- story genre
- story topic
- story time setting
- ...

The automatic parser may not detect interesting syntactic aspects because it is developed to handle contemporary standard Dutch

8

Vocabulary results

t-score	f ₁	f ₂	token	t-score	f ₁	f ₂	token
0.99944	1796	0	z'n	0.99600	249	0	Willy
0.99925	1338	0	nie	0.99507	202	0	Laila
0.99921	1270	0	d'r	0.99502	200	0	...!
0.99863	728	0	Bo	0.99444	179	0	dinges
0.99839	621	0	fo	0.99401	166	0	Couplet
0.99830	586	0	em	0.99390	163	0	Aaj
0.99778	450	0	Mamsi	0.99296	141	0	baja
0.99714	349	0	Gusta	0.99275	137	0	Faader
0.99653	287	0	Baas	0.99180	121	0	Coola
0.99620	262	0	wou	0.99153	117	0	néks

9

Syntactic rules

We analyse sentences with dependency rules:

Steven sings sad songs

generates the syntactic analysis:

name	Steven	is a	subject	of	verb	sings
adjective	sad	is a	object	of	noun	songs
noun	songs	is an	object	of	verb	sings

We also lookup the lemma of each word: the basic word form:
Steven → Steven; *sings* → sing; *sad* → sad; *songs* → song;

10

Syntactic results: Lemma Relation Lemma

t-score	Lemma Relation Lemma	t-score	Lem Rel Lem
0.98780	ding hd/det dat	0.96875	van hd/obj1 erf
0.98529	soort hd/mod van	0.96774	oog hd/det je
0.98113	ma tag/nucl ben	0.96774	kijk hd/mod daar
0.98077	erf hd/det zijn	0.96774	jongen hd/det die
0.97872	erf hd/det dat	0.96774	hoofd hd/det je
0.97561	ander hd/det die	0.96552	ben hd/su erf
0.97500	al cmp/mod ook	0.96000	zeg hd/mod zo
0.97436	kind hd/det die	0.96000	in hd/obj1 me
0.97297	verkoop hd/obj1 erf	0.96000	broek hd/det zijn
0.97059	boom hd/det die	0.95946	ga hd/vc kom

11

Example sentences

Pattern: **ga hd/vc kom**

hij heb vermoeden dat die Bo ga kom
he has suspicion that that Bo goes come

Pattern: **zeg hd/mod zo**

Droomboek zeg zo, dus Vrouw Couplet ook.
Droomboek says so, so Mrs Couplet too.

12

Syntactic results: PoS Relation Lemma

t-score	PoS Relation Lemma	t-score	PoS Relation Lemma
0.98361	verb hd/su hond	0.96429	prep hd/obj1 hoed
0.98361	verb hd/su erf	0.96296	comp dlink/nucl dan
0.97561	noun hd/det jullie	0.96296	prep hd/obj1 broek
0.97500	prep hd/obj1 soort	0.96000	verb hd/prec baas
0.97222	noun hd/mod schoon	0.96000	comp dp/dp ga
0.97143	verb hd/vc breek	0.96000	comp dlink/nucl laat
0.96774	verb hd/su boom	0.96000	comp dlink/nucl dat
0.96698	noun hd/mod daar	0.95652	adv dp/dp met
0.96667	comp dlink/nucl met	0.95455	prep hd/obj1 dood
0.96552	verb nucl/tag vind	0.95455	comp dlink/nucl te

13

Example sentence:

Pattern: **comp dp/dp ga**

want iemand van me familie ga kom!
because someone of my family goes coming!

14

Syntactic results: PoS Relation PoS

t-score	PoS Relation PoS	t-score	PoS Relation PoS
0.97378	comp dlink/nucl noun	0.92308	pron dp/dp prep
0.96875	adj hd/ld prep	0.91667	tag tag/nucl det
0.96689	det hd/mod noun	0.91667	adv rhd/body comp
0.95775	comp dlink/nucl prep	0.91096	comp dlink/nucl comp
0.95652	comp dp/dp det	0.90909	pron dp/dp adv
0.95455	prep nucl/tag tag	0.90909	adj hd/obj2 noun
0.95288	det hd/mod name	0.90625	prep hd/prec comp
0.94330	comp dlink/nucl adv	0.90566	comp nucl/tag tag
0.93333	noun tag/nucl noun	0.89552	adj dp/dp adv
0.92857	pron hd/mod noun	0.89157	adv dp/dp noun

15

Example sentence

Pattern: **adv rhd/body comp**

waar dat *ze staande loerde*
where that *she standingly peeked*

16

Syntactic results: Lemma Relation PoS

t-score	Lemma Relation PoS	t-score	Lemma Relation PoS
0.99495	ma tag/nucl verb	0.96667	ma tag/nucl adv
0.99020	dan dp/dp noun	0.96667	dan hd/mod comp
0.99000	zijn hd/mod noun	0.96429	zie dp/dp noun
0.98947	dan dp/dp verb	0.96429	hoor dp/dp noun
0.98592	soort hd/mod prep	0.96000	kijk dp/dp noun
0.98214	ma dp/dp verb	0.95833	want dlink/nucl noun
0.97778	baas hd/app name	0.95783	en dlink/nucl noun
0.97561	maar dlink/nucl noun	0.95652	Dan dp/dp verb
0.97297	ma tag/nucl noun	0.95652	ma tag/nucl adj
0.96774	ma dp/dp noun	0.95652	zijn hd/mod name

17

Example sentence

Pattern: **dan dp/dp verb**

dan kijk *hoe ze wegmanoevreert*
then look *how she leaves*

18

Concluding remarks and future work

We have shown that automatic methods can assist in finding syntactic differences between texts

A paper about this work can be found on <http://ifarm.nl/erikt/papers>

The techniques described in this talk will be build in into the Nederlab framework



19

THE END