

## Appendix

Below is a list of n-grams that are significant at level 0.001 and occur at least 100 times. For highlighted words, we should the original tweets that has the occurrences of code-switching in the following table.

Table I. Significant Positive N-grams. These n-grams are more correlated with code-switching.

Beta	N-Gram	Translation	# times	Beta	N-Gram	Translation	# times
1.655486	ตีแผ่	Expose	198	0.666319	ม้าย	Maybe (particle word)	196
1.587904	เด็ก	kid	226	0.644223	คำ ว่า	the word	1955
1.286285	แอป	App	366	0.642781	เสื้อผ้า	clothes	3780
1.258007	หด	Shrink	524	0.638636	รวม	Total	1726
1.200594	นคร	City	602	0.607498	งาน	Job	17645
1.196338	ไป โหวต	Let's go vote	228	0.592980	สน	Care, interested	5794
1.126960	ปุ่ม	Button	130	0.592954	เกม	Game	1444
1.08817	เพลง	Song	5123	0.581066	โชว์	Show	1035
1.073115	อัป	Upload	603	0.575604	จอ	Screen	769
1.057182	กด	Press	3370	0.573533	ลาย	Pattern	1810
0.980306	อัลบั้ม	Album	187	0.562297	นะ	Please (particle word)	184
0.930855	คาเฟ่	Cafe	144	0.548819	หนุ่มๆ	Boys	115
0.682889	ฟรี	Free	1925	0.543560	คอนเสิร์ต	Concert	318

Table II. Significant positive Thai n-grams and their associated tweets

Thai N-Gram	Translation	Beta	Example Tweets / Translation
งาน	Task/Job/Event	0.607498	การที่ใกล้ปิดงาน แล้วใช้โปรแกรมอะไรในเครื่องเยอะๆ ( ยิ่งพวก adobe เจ้าปัญหา ) จะทำให้คุณไม่กด <b>update</b> อะไรง่าย ๆ // มันล่มมาเยอะแล้ว
			<b>When you use a lot of programs (especially Adobe), it will make it hard to click update.</b>
			รีพีทงานสุด <b>Hot</b> คร่า
			<b>Repeat! Very lucrative job.</b>
			@lftcphyominnn89 ไปงานนี้กัน ก่อนน # นพมาศฟาดกระทง <b>party</b> . จุงมือ 555555555555
			<b>Let's go to this full moon party</b>
ปุ่ม	button	1.126960	แอปเปิลเปลี่ยนคำแสดงผลบนปุ่ม ดาวน์โหลดแอปพลิเคชันฟรีจาก " <b>Free</b> " เป็น " <b>Get</b> " <a href="https://t.co/jyTm4knIWh">https://t.co/jyTm4knIWh</a> - ดิๆ เราไม่เคยชอบปุ่มที่เขียนว่า " ฟรี " เลย
			<b>Apple changed the button 'Free' to 'Get'. I have never liked the button 'Free'.</b>
			อยากสร้างปุ่ม <b>skip</b> ให้ตัวเอง
			<b>I want to create 'skip' button for myself.</b>
			มีปุ่ม <b>Delete</b> แพนเก่าออกไปจากสมองไหม
			<b>Is there a delete button for ex boyfriends/girlfriends for my brain?</b>
เพลง	song	1.088173	#TheVoice ถึงอารมณ์สุดๆ # ปอย <b>vs</b> # จ้อ กับเพลง # ความรัก # ตัวเต็ม #daradaily cr.#thevoicethailand - <a href="https://t.co/OxWy6soqbU">https://t.co/OxWy6soqbU</a>
			<b>Good show! Poy vs. Jor with song # love ...</b>
			สิ่งที่น่ากลัวในการฟังเพลงเยอะๆ คือหลายๆ เพลงแค่ขึ้นดนตรี <b>intro</b> มาแล้วรู้ ว่าคือเพลง อะไร
			<b>What's scary is when you listen to a long of songs but when you hear intro, you know what song it is.</b>
คอนเสิร์ต	Concert	0.543560	สำหรับใครที่มีความสนใจเรื่องการทำ <b>sound</b> งานคอนเสิร์ตเป็นพิเศษ ไม่ควรพลาดกิจกรรมนี้ เรามี <b>workshop</b> พิเศษทำงาน
			<b>For those you are interested in doing sound at concert, please don't miss this special workshop activity.</b>
			จูนซุ : แพนๆคนที่มาดูคอนเสิร์ต พวกเราเมื่อวานนี้ และก็มีวันนี้อีกรอบ มัน <b>incredible</b> มากครับ !!! * น่ารักอีกแล้วววว *
			<b>The fans who came to see our concert yesterday and again today are incredible.</b>

## Interpretation

There are roughly 2000 significant positive Thai n-grams at level 0.001. The words shown in Table I are words that occur frequently enough that the correlation is not a noise and are easier to interpret.

Consider the word (ปุ่ม,button) for instance. This word is significantly positive as there are many instances when users refer to the button name on their electronic device. (which is somewhat non-surprising)

Another interesting word is (งาน,job). This n-gram occurs very frequently (~18K times) but seems to be more positively correlated to code-switching. This word is associated with event, job that the younger or working generation talks about such as ‘going to the full moon event’ or ‘very lucrative job’.

(คอนเสิร์ต, Concert) might also be positively correlated due to the younger user base who more frequently goes to concert or has a job related to concert.

For negative words, based on Table III, it’s hard to interpret as to why these words are negatively correlated with code-switching. However, the highlighted words show types of word that might be associated with lower socio-economic status as the words are highly informal/ slang pronouns that some people would not used comfortably in public.

Table III. Significantly Negative Words

Beta	N-Gram	Translation	# times	Beta	N-Gram	Translation	# times
-0.631654	ส่ง ฟรี	Free Delivery	478	-0.418667	งง	Confused	34463
-0.602323	ขอ เพลง	Request song	342	-0.389046	หรือ	Is it? (particle word)	11579
-0.512966	นอน	Sleep	18079	-0.388566	กู	I (informal)	64995
-0.473843	ฉัน	I	15892	-0.377427	มึง	You (informal)	23057
-0.457594	ทำ งาน	Work	3541	-0.371505	ทุก วัน	Every day	3082