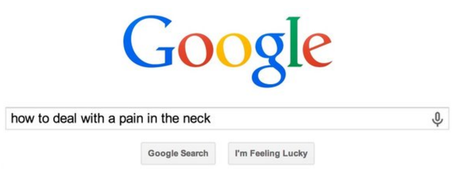
Detecting Multi-word Expressions Through Typing Patterns

Adam Goodkind
Department of Linguistics, CUNY Graduate Center

The Problem



Do they need a chiropractor or a stress ball?

Multi-Word Expressions

The phrase *pain in the neck* can have two different meanings, either the literal meaning, or the figurative meaning.

What is a multi-word expression?

A *multi-word expression* (MWE) is a word collocation that exhibits markedly peculiar linguistic behavior in terms of lexicalization, syntax or semantics. (Kunchukuttan 2007)

In other words:
$\text{def}(word_1) + \text{def}(word_2) \neq \text{def}(word_1 + word_2)$

MWEs in Speech

MWEs are theorized to be stored and retrieved as a single lexical unit (Wray 2005). Often, the pause length is longer surrounding an MWE, and shorter within an MWE. (Dahlmann & Adolphs 2007)

She___looked_up___ the___world_record___

?? Question ??

Can we apply speech prosody to keyboard typed text, to distinguish MWEs from other text?

Keystroke Dynamics to the Rescue!

Keystroke dynamics (KD) measures the timing of every computer keystroke, from the time a key is pressed to the time it is released.

Detecting MWEs in Typing

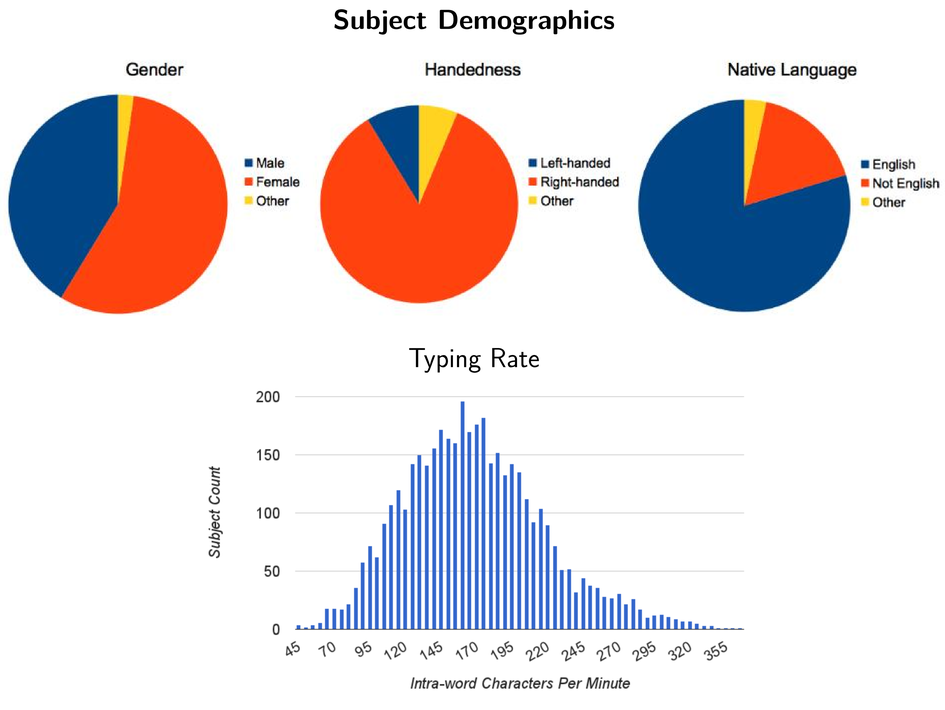
Assumption 1: MWEs have unique prosodic characteristics
Assumption 2: Keystroke Dynamics is the reflection of speech prosody in typing
Conclusion: MWEs should be uniquely characterized in typing

Materials

- 1,013 Louisiana Tech college students
- Recorded subjects clickstream (keystrokes)
- Recorded clickstream timing (key up/down)
- Subjects respond to a randomized question set
- Subjects can skip to next question after typing 300 characters

Methodoology

From the keystroke timing data, we determined the length of pause before each word. Each word was then categorized into one of four positions, relative to an MWE: outside, starting, middle of, and ending an MWE.



!! Results !!

Mean Pauses Before Word Location

		Start	Middle	End	Outside
L1	**Fast**	429	220	31152	8839
	Slow	3787	491	137456	33773
L1 Mean		**1594**	**370**	**62810**	**19375**
L2	**Fast**	799	250	74984	11866
	Slow	2709	187	123149	43385
L2 Mean		**1718**	**519**	**87086**	**27547**
Overall		**1616**	**397**	**67252**	**20888**

All figures are normalized for predictability (Ask me what that is!)

- Typists pause for significantly shorter lengths of time in the middle of MWEs
- While slow typists were 4x slower than fast typists outside of an MWE, the two groups were much closer within MWEs
- Slow L1 English speakers are slower than slow L2 English speakers

Future Steps

- Build an MWE Detector, bottom-up
- Does this improve information Retrieval?
- Can we extend this to other languages?

Contact Information

- Web: htttp://www.adamgoodkind.com
- Email: a.goodkind@gmail.com

Acknowledgements

Thank you to my advisor, Dr. Andrew Rosenberg for his ceaseless advice and guidance, and my research partner David-Guy Brizan. This work was supported in part by DARPA Active Authentication Phase I grant FA8750-12-2-0201 and FA8750-13-2-0274