

Institute of Computational Linguistics

# Programmiertechniken in der Computerlinguistik I

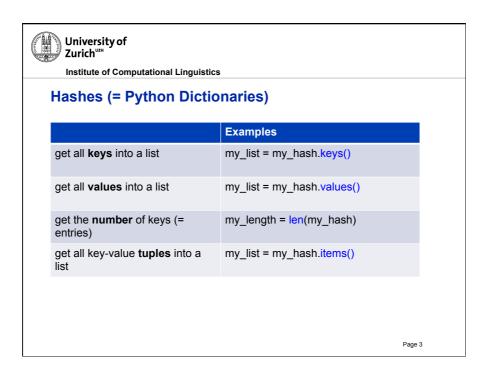
Herbstsemester 2015 5. Sitzung, 22. Oktober 2015 Martin Volk

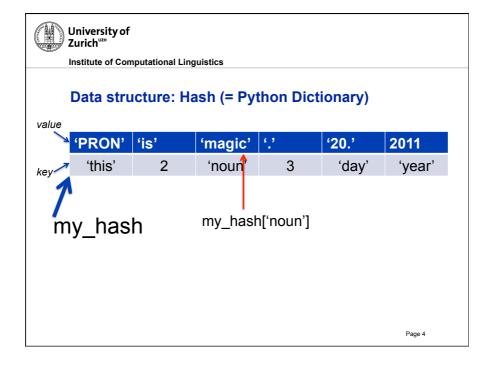


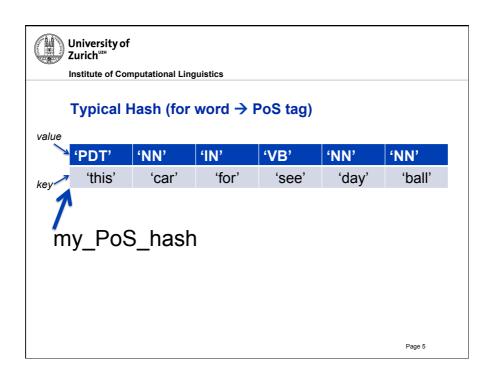
Institute of Computational Linguistics

#### **Hashes (= Python Dictionaries)**

	Examples
Define a hash (= dictionary) - with two entries	my_hash = {} my_hash = {'uni':'noun', 'gute':'adj'}
Enter a key – value pair	my_hash['mit'] = 'praep'
Check if a key is in the hash - and get the value	my_hash.has_key(my_word) my_word in my_hash my_hash.get(my_word)
Loop over a hash	for my_word in my_hash.keys() for my_word in my_hash









## Hashes (= Python Dictionaries): Typical Usages

	Examples
word + PoS tag	<pre>my_hash = {'der':'ART', 'in':'APPR', 'Auto':'NN'}</pre>
item + frequency	my_hash = {'der':4, 'in':12, 'Auto':2} my_hash = {'ART':7, 'NN':12, 'APPR':8}
word + translation	my_hash = {'der':'the', 'in':'in', 'Auto':'car', 'Haus':'house'}
string + dummy value (for quick access)	my_hash = {'der':1, 'in':1, 'Auto':1}



Institute of Computational Linguistics

#### List vs. Hash

#### List

my\_list = []

stores a sequence of values  $x_1, x_2, x_3 \dots$ 

preserves order

quick access via number key to value, but **slow** access to a particular value

#### Hash

my\_hash = {}

stores relations  $x_1 \rightarrow y_1 \ x_2$ 

 $\rightarrow$  y<sub>2</sub>

random order!

quick access via key to value

Page 7



Institute of Computational Linguistics

# Regular Expressions Search and Substitution

	Examples
Regular expression matching	<pre>import re my_match = re.search('abc', my_text) my_match = re.search('^\waa', my_text) print my_match.group()</pre>
Get hit from within a match	<pre>my_match = re.search('ab(\w+)', my_text) print my_match.group(1)</pre>
Get multiple matches in a list	my_matches = re.findall('a\w+', my_text)
Find and substitute	my_new_text = re.sub('t', 'th', my_text)



Institute of Computational Linguistics

#### **Miscellaneous**

	Examples
Defining functions - the main function	<pre>def my_function():   def main():</pre>
Global variables (defined outside a function)	
Remove <b>leading</b> (left) "whitespace" characters	my_line.lstrip()
Remove <b>trailing</b> (right) "whitespace" characters	my_line.rstrip()

Page 9



Institute of Computational Linguistics

## **Programming Techniques in this Lesson**

- Filtering a text and writing output to multiple files
- Person name recognition and gender statistics
- Reading information from a file and applying it to a text
- · Word by word machine translation