Words of a feather flock together

John Goldsmith (Linguistics and Computer Science)
Work done with:
   Jackson Lee (Linguistics)
   Simon Jacobs (RCC)
   Stephen Fitz (Computer Science)
Looking at texts: *documents*

- Texts contain a great deal of *structure*.
- That structure can be divided into:
  - Structure of the *content*;
  - Structure of the *language*.
- Look at a series of books in French: what they share in common is *the structure of French*.
- Look at a series of translations of *Harry Potter and the Sorcerer's Stone*: what they share in common is *the structure of the content*. 
We care about the structure of the language (today)

- Structure of a language includes a *lexicon*: the words of the language.
- The *order* in which words appear is largely determined by the *syntax* of the language.
Word categories

- Noun
- Verb
- Participle
- Article
- Pronoun
- Preposition
- Adverb
- Conjunction
Mad Libs Worksheet
In The Jungle!

I walk through the color jungle. I take out my
____________________(adjective) canteen. There's a
__________(adjective) parrot with a ______________________(adjective)
___________________(noun) in his mouth right in front of me in the
____________________(adjective) trees! I gaze at his _____________(adjective)
________________________(noun). A sudden sound awakes me from my
____________________(adjective) canteen. I turn around and see a ______________________(noun) in front of me...
Mad Libs Worksheet

In The Jungle
Mad Libs Worksheet

In The Jungle

 undue mental
Mad Libs Worksheet

In The Jungle

Name ____________________________  Date __________________

undue mental valid
name __________________________

Date __________________

Mad Libs Worksheet

In the Jungle

undue mental prison valid
Mad Libs Worksheet

In The Jungle:

undue mental prison equitable valid
There are many ________ ways to choose a/an ________ to read. First, you could ask for recommendations from your friends and _________. Just don’t ask Aunt _________—she only reads ________ books with ________-ripping goddesses on the cover. If your friends and family are no help, try checking out the ________ Review in The ________ Times. If the ________ featured there are too ________ for your taste, try something a little more low-______ , like ________: The ________ Magazine, or ________ Magazine. You could also choose a book the
There are many __________ ways to choose a/an __________ to read. First, you could ask for recommendations from your friends and __________. Just don’t ask Aunt ______________—she only reads __________ books with ______________-ripping goddesses. 
There are many ___________ ways to choose a/an ___________ to read. First, you could ask for recommendations from your friends and ___________. Just don’t ask Aunt ________________—she only reads ___________ books with ________________-ripping goddesses.
There are many ___________ ways to choose a/an ___________ to read. First, you could ask for recommendations from your friends and ___________. Just don’t ask Aunt ______________—she only reads ___________ books with ______________-ripping goddesses.
There are many __________ ways to choose a/an __________ to read. First, you could ask for recommendations from your friends and __________. Just don’t ask Aunt ______________—she only reads __________ books with ______________-ripping goddesses.
There are many __________ ways to choose a/an __________ to read. First, you could ask for recommendations from your friends and __________. Just don’t ask Aunt ______________—she only reads __________ books with ______________-ripping goddesses.
There are many ____________ ways to choose a/an ____________ to read. First, you could ask for recommendations from your friends and ____________. Just don’t ask Aunt _________________—she only reads ____________ books with _________________-ripping goddesses.
There are many ____________ ways to choose a/an ____________ to
ADJECTIVE NOUN
read. First, you could ask for recommendations from your friends and
PLURAL NOUN PERSON IN ROOM (FEMALE)
___________. Just don’t ask Aunt __________________—she only
reads ___________ books with _________________-ripping goddesses
ADJECTIVE ARTICLE OF CLOTHING
There are many _______ ways to choose a/an _________ to read. First, you could ask for recommendations from your friends and _________ . Just don’t ask Aunt ________________—she only reads _________ books with ________________ -ripping goddesses.
There are many _______________ ways to choose a/an _______________ to read. First, you could ask for recommendations from your friends and _______________. Just don’t ask Aunt ________________—she only reads _______________ books with _______________-ripping goddesses.
There are many __________ ways to choose a/an __________ to read. First, you could ask for recommendations from your friends and __________. Just don’t ask Aunt ______________—she only reads __________ books with ______________-ripping goddesses.
There are many ___________ ways to choose a/an ___________ to read. First, you could ask for recommendations from your friends and ___________. Just don’t ask Aunt ______________——she only reads ___________ books with ______________-ripping goddesses.

ADJECTIVE

NOUN

PLURAL NOUN

PERSON IN ROOM (FEMALE)

ADJECTIVE

ARTICLE OF CLOTHING
There are many ___________ ways to choose a/an ___________ to read. First, you could ask for recommendations from your friends and ___________. Just don’t ask Aunt ___________—she only reads ___________ books with ___________-ripping goddesses.
There are many ___________ ways to choose a/an ___________ to read. First, you could ask for recommendations from your friends and ___________. Just don’t ask Aunt ____________—she only reads ___________ books with ___________-ripping goddesses.
There are many _________ ways to choose a/an _________ to read. First, you could ask for recommendations from your friends and _________. Just don't ask Aunt ____________—she only reads _________ books with _________-ripping goddesses.
Your startup name: ______________________. Have you ever wanted to ________
(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)
all sorts of ________ with friends, family, colleagues, or even ________, but didn’t
(PLOYAL NOUN)
know how? Maybe you were just strolling the sidewalks of ________ past your
(MINOR CELEBRITY)
favorite ________ bakery when you wished you could pull out your i______
(OBSCURE ETHNICITY)
and use it to quickly ________ all your latest ________ and post it to your
(VERB ENDING IN -ATE)
_____? ________ is the answer. The fact is, we live in a world
(SOCIAL MEDIA PROFILE NO ONE READS) (STARTUP NAME)
where ________ will revolutionize ________, and pretty soon you’ll be able to
(GERUND STARTING WITH “CROWD”) (FAILING INDUSTRY)
________ anything you want through the cloud. All this________
(VERB THAT IMPLIES CREATIVITY) (NOUN FORM OF VERB ENDING IN -ATE)
Your startup name: _________________. Have you ever wanted to ____________ (MISSPelled VERSION OF REAL WORD, ENDING IN -LY OR -IO) __________ (VERB)

all sorts of ________ with friends, family, colleagues, or even ________, but didn’t know how? Maybe you were just strolling the sidewalks of __________ past your favorite ________ bakery when you wished you could pull out your i________ (MINOR CELEBRITY) (PLURAL NOUN) (GENTRIFIED NEIGHBORHOOD) (OBSCURE ETHNICITY) (NOUN)
Your startup name: _______________. Have you ever wanted to __________

(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)

all sorts of _______ with friends, family, colleagues, or even _________, but didn’t know how? Maybe you were just strolling the sidewalks of __________ past your favorite _________ bakery when you wished you could pull out your i_________

(PLURAL NOUN)

(MINOR CELEBRITY)

(GENTRIFIED NEIGHBORHOOD)

(OBSCURE ETHNICITY)

(NOUN)
Your startup name: ___________________. Have you ever wanted to _____ (MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO) _______ (VERB) _____ all sorts of _______ with friends, family, colleagues, or even _______ (MINOR CELEBRITY), but didn't know how? Maybe you were just strolling the sidewalks of _______ past your favorite _______ bakery when you wished you could pull out your i _______ (OBSCURE ETHNICITY).
Your startup name: _________________. Have you ever wanted to ____________
(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)

all sorts of ________ with friends, family, colleagues, or even ____________, but didn’t
(PLURAL NOUN) (MINOR CELEBRITY)

know how? Maybe you were just strolling the sidewalks of ____________ past your
(GENTRIFIED NEIGHBORHOOD)

favorite ____________ bakery when you wished you could pull out your i__________
(OBSCURE ETHNICITY) (NOUN)
Your startup name: _________________. Have you ever wanted to ____________
(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)

Rheummate-ly jump

sneakers

Mitt Romney

Kenwood

jump

all sorts of _______ with friends, family, colleagues, or even _________, but didn’t

(PLURAL NOUN)

know how? Maybe you were just strolling the sidewalks of ______________ past your

(MINOR CELEBRITY)

gentrified neighborhood)

favorite __________ bakery when you wished you could pull out your i_________

(OBSOKE ETHNICITY)
Your startup name: _______________. Have you ever wanted to _______ (MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO) the sneaker Mitt Romney Kenwood Igbo.

All sorts of _______ with friends, family, colleagues, or even _______ (MINOR CELEBRITY), but didn’t know how? Maybe you were just strolling the sidewalks of _______ (GENTRIFIED NEIGHBORHOOD) past your favorite _______ bakery when you wished you could pull out your i_ _______ (OBSCURE ETHNICITY).
Your startup name: _________________. Have you ever wanted to ________

(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)

MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO

jump

Mitt Romney

Kenwood

Soap

all sorts of ______ with friends, family, colleagues, or even ________, but didn’t

(PLURAL NOUN)

Igbo

know how? Maybe you were just strolling the sidewalks of ________ past your

(MINOR CELEBRITY)

past your

(GENTRIFIED NEIGHBORHOOD)

favorite ________ bakery when you wished you could pull out your i________

(OBSCURE ETHNICITY)

(NOUN)
Your startup name: _________________. Have you ever wanted to ________ (MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)

jump

Mitt Romney

Kenwood

Soap

all sorts of _______ with friends, family, colleagues, or even _________, but didn’t know how? Maybe you were just strolling the sidewalks of _________ past your favorite _________ bakery when you wished you could pull out your i_________ (PLURAL NOUN) (MINOR CELEBRITY) (GENTRIFIED NEIGHBORHOOD) (OBSCURE ETHNICITY) (NOUN)
Your startup name: _________________. Have you ever wanted to ___________

(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)

all sorts of _______ with friends, family, colleagues, or even __________, but didn’t

(PLURAL NOUN) (MINOR CELEBRITY)

differentiate pogosticks

Rheummate-ly jump

(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)

sneakers

Mitt Romney

Kenwood

Igbo

Soap

know how? Maybe you were just strolling the sidewalks of __________ past your

(GENTRIFIED NEIGHBORHOOD)

favorite __________ bakery when you wished you could pull out your i__________

(OBSCURE ETHNICITY) (NOUN)
Your startup name: __________________________. Have you ever wanted to ________ (MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)

sneakers

Mitt Romney

Kenwood

Igbo Soap

differentiate pogosticks

GoogleMine

jump

all sorts of ________ with friends, family, colleagues, or even ________, but didn’t know how? Maybe you were just strolling the sidewalks of ________ past your favorite ________ bakery when you wished you could pull out your i ________ (PLURAL NOUN) (MINOR CELEBRITY) (GENTRIFIED NEIGHBORHOOD) (OBSCURE ETHNICITY) (NOUN)
Your startup name: _______________. Have you ever wanted to ____________?

MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO

all sorts of ______ with friends, family, colleagues, or even ________, but didn’t

(PLURAL NOUN) (MINOR CELEBRITY) (GENTRIFIED NEIGHBORHOOD)

know how? Maybe you were just strolling the sidewalks of ____________ past your

differentiate pogosticks

(GENOCIDAL PERIOD) (GENTRIFIED NEIGHBORHOOD)

favorite ___________ bakery when you wished you could pull out your i__________

(OBSCURE ETHNICITY) (NOUN)
Have you ever wanted to

Your startup name: Rheummate-ly

Kenwood

Rheummate-ly

Mitt Romney

differentiate

goosebumps

crowd-licking

(slang)

gender-identity

past your

soap

Pogosticks

Rheummate-ly

��道吗？也许你就是那种边走边沿着人行道

with friends, family, colleagues, or even

all sorts of

(Minor Celebrity)

but didn't

soap

kenwood

Rheummate-ly

sneaker

jump

(Retail Noun)

Misspelled version of Real Word. Ending in -ly or -0

Rheummate-ly
Your startup name: _________________. Have you ever wanted to ____________

(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)

(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)

all sorts of __________ with friends, family, colleagues, or even __________, but didn’t

(PLURAL NOUN)

(MINOR CELEBRITY)

know how? Maybe you were just strolling the sidewalks of __________ past your

(GENTRIFIED NEIGHBORHOOD)

crowd-licking ___________ bakery when you wished you could pull out your i_________

(CROWD-licking)

GoogleMine

(RHEUMMATE-LY)

Rheummate-ly

differentiate

taxi business

pogosticks

Igbo

Soap

Kenwood

Mitt Romney

jump

sneakers

favorite ___________.
Your startup name: _________________. Have you ever wanted to __________ jump
sneakers
(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)
Mitt Romney
Kenwood Soap
all sorts of Igbo ______ with friends, family, colleagues, or even _________, but didn’t
(PLURAL NOUN) thinkify
differentiate pogosticks
GoogleMine Rheummate-ly
crowd-licking taxi business
Rheummate-ly

know how? Maybe you were just strolling the sidewalks of __________ past your
(GENTRIFIED NEIGHBORHOOD)
favorite ________ bakery when you wished you could pull out your i_________
(OSTRANGE ETHNICITY)
Empirically, not so obvious:
'the' versus 'a'
<table>
<thead>
<tr>
<th>Left</th>
<th>Right</th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td></td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>of</td>
<td>9629</td>
<td>of</td>
<td>1467</td>
</tr>
<tr>
<td>in</td>
<td>5567</td>
<td>in</td>
<td>1321</td>
</tr>
<tr>
<td>to</td>
<td>3433</td>
<td>as</td>
<td>914</td>
</tr>
<tr>
<td>on</td>
<td>2315</td>
<td>with</td>
<td>879</td>
</tr>
<tr>
<td>and</td>
<td>2143</td>
<td>is</td>
<td>872</td>
</tr>
<tr>
<td>for</td>
<td>1766</td>
<td>was</td>
<td>783</td>
</tr>
<tr>
<td>at</td>
<td>1511</td>
<td>for</td>
<td>740</td>
</tr>
<tr>
<td>with</td>
<td>1481</td>
<td>to</td>
<td>655</td>
</tr>
<tr>
<td>that</td>
<td>1392</td>
<td>and</td>
<td>615</td>
</tr>
<tr>
<td>from</td>
<td>1352</td>
<td>on</td>
<td>476</td>
</tr>
<tr>
<td>by</td>
<td>1315</td>
<td>by</td>
<td>413</td>
</tr>
<tr>
<td>is</td>
<td>793</td>
<td>be</td>
<td>318</td>
</tr>
<tr>
<td>as</td>
<td>698</td>
<td>at</td>
<td>256</td>
</tr>
<tr>
<td>into</td>
<td>673</td>
<td>have</td>
<td>256</td>
</tr>
<tr>
<td>was</td>
<td>531</td>
<td>had</td>
<td>246</td>
</tr>
<tr>
<td>:</td>
<td>466</td>
<td>like</td>
<td>242</td>
</tr>
<tr>
<td>all</td>
<td>424</td>
<td>from</td>
<td>229</td>
</tr>
<tr>
<td>In</td>
<td>412</td>
<td>such</td>
<td>222</td>
</tr>
</tbody>
</table>
masculine singular nouns  

feminine singular nouns  
exisiele  
simple past verbs  
prenominal modifiers  
plural nouns  
infinitives  

cities  

French  

English  

years  

Encarta (encyclopedia) 4,000 words  

nouns  

plural nouns  
countries  

units  

prepositions  
bare verbs  

past tense verbs  

auxiliary verbs  

john, charles,...
## Dynamic view: English color codes

<table>
<thead>
<tr>
<th>Category</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbs: ‘bare’ verb (jump)</td>
<td>red</td>
</tr>
<tr>
<td>Verbs: past tense (jumped, bought)</td>
<td>blue</td>
</tr>
<tr>
<td>Verbs: auxiliary (should, can)</td>
<td>green</td>
</tr>
<tr>
<td>Prepositions (from, to, up, down)</td>
<td>aqua</td>
</tr>
<tr>
<td>Adjectives</td>
<td>purple</td>
</tr>
<tr>
<td>Cities</td>
<td>gray</td>
</tr>
<tr>
<td>Nouns</td>
<td>pink</td>
</tr>
<tr>
<td>Category</td>
<td>Color</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Infinitives</td>
<td>red</td>
</tr>
<tr>
<td>Prepositions</td>
<td>light blue</td>
</tr>
<tr>
<td>Past participles</td>
<td>blue</td>
</tr>
<tr>
<td>Adjectives</td>
<td>purple</td>
</tr>
<tr>
<td>Cities</td>
<td>gray</td>
</tr>
<tr>
<td>Masculine nouns</td>
<td>pink</td>
</tr>
<tr>
<td>Feminine nouns</td>
<td>light green</td>
</tr>
<tr>
<td>Inflected verbs</td>
<td>light gray</td>
</tr>
</tbody>
</table>
Finding regular and irregular past tense forms