## 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.


## Frank and Ernest



Copyright (c) 1997 by Thaves. Distributed from www.thecomics.com.

## Word categories

## The <br> Grammar of Dionysios <br> Thrax

- Noun
- Verb
- Participle
- Article
- Pronoun
- Preposition
- Adverb
- Conjunction
$\qquad$
$\qquad$


## Mad Libs Worksheet

## In The Jungle!

I walk through the color jungle. I take out my (adjective) canteen. There's a
$\qquad$ (adjective) parrot with a $\qquad$ (adjective)
$\qquad$ (noun) in his mouth right in front of me in the
$\qquad$ (adjective) trees! I gaze at his $\qquad$ (adjective)
$\qquad$
$\qquad$

## Mad Libs Worksheet

## In The Jungle!

I walk through the color jungle. I take out my undue_(adjective) canteen. There's a
$\qquad$ (adjective) parrot with a $\qquad$ (adjective)
$\qquad$ (noun) in his mouth right in front of me in the
$\qquad$ (adjective) trees! I gaze at his $\qquad$ (adjective)

Name $\qquad$ Date $\qquad$

## Mad Libs Worksheet

## In The Jungle!

I walk through the color jungle. I take out my mes (adjective) parrot with a $\qquad$ (adjective)
$\qquad$ (noun) in his mouth right in front of me in the
$\qquad$ (adjective) trees! I gaze at his $\qquad$ (adjective)

Name $\qquad$ . Date $\qquad$

## Mad Libs Worksheet

## In The Jungle!

I walk through the color jungle. I take out my
$\square$ (adjective) canteen. There's. a Mental(adjective) parrot with a $\qquad$ (adjective)
$\qquad$ (noun) in his mouth right in front of me in the
$\qquad$ (adjective) trees! I gaze at his $\qquad$ (adjective)

Name $\qquad$ Date $\qquad$

## Mad Libs Worksheet

## In The Jungle!

I walk through the color jungle. I take out my
$\square$ (adjective) canteen. There's. a mental(adjective) parrot with a $\qquad$ (adjective)
PrISOn_(noun) in his mouth right in front of me in the (adjective) trees! I gaze at his $\qquad$ (adjective)

Name $\qquad$ . Date $\qquad$

## Mad Libs Worksheet

## In The Jungle!

I walk through the color jungle. I take out my
$\square$ (adjective) canteen. There's. a
ment al lajective) parrot with a $\qquad$ (adjective)
PrISOn_(noun) in his mouth right in front of me in the equitable (adjective) trees! I gaze at his $\qquad$ (adjective)

# MAD $(9) L I B S^{\circ}$ <br> <br> 至BOOKish Mad Libs 

 <br> <br> 至BOOKish Mad Libs}

There are many $\qquad$ ways to choose a /an $\qquad$
ADJECTIVE
NOLIN
read. First, you could ask for recommendations from your friends and
. Just don't ask Aunt $\qquad$ —she only books with $\qquad$ -ripping goddesses reads $\qquad$ ARTICLE OF CLOTHING
on the cover. If your friends and family are no help, try checking out the
$\qquad$ Review in The $\qquad$ Times. If the $\qquad$ NOLN PLLLRAL NOLN
featured there are too $\qquad$ for your taste, try something a little ADJECTIVE
more low- $\qquad$ , like $\qquad$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

## MAD $(9) L I B S$

## 至BOOKish Mad Libs

There are many $\underbrace{S}_{\text {ADJECTIVE }}$ ways to choose a/an $\square$ read. First, you could ask for recommendations from your friends and PLLRAL NOLN $\cdot$ Just don't ask Aunt $\begin{aligned} & \text { PERSON IN ROOM (FEMALE) }\end{aligned}$-she only
reads $\begin{aligned} & \text { ADJECTIVE }\end{aligned}$ books with $\begin{aligned} & \text { ARTICLE OF CLOTHING }\end{aligned}$ ripping goddesses on the cover. If your friends and family are no help, try checking out the
$\qquad$ Review in The $\qquad$ Times. If the $\qquad$ NOLN A CITY PLLIRAL NOLIN
featured there are too $\qquad$ for your taste, try something a little ADJECTIVE more low- $\qquad$ , like $\qquad$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

## MAD $(9) L I B S^{\circ}$

## 至BOOKish Mad Libs

There are many $\underset{\text { ADJECTIVE }}{S i \|}$ ways to choose a/an bagel $_{10}$ read. First, you could ask for recommendations from your friends and
 on the cover. If your friends and family are no help, try checking out the
$\qquad$ Review in The $\qquad$ Times. If the $\qquad$ featured there are too $\qquad$ for your taste, try something a little ADJECTIVE more low- $\qquad$ , like $\qquad$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

# MAD $(9) L I B S$ 

## 至BOOKish Mad Libs

There are many $\operatorname{Sid}_{\text {ADJECTIVE }}$ bagel $_{10}$ read. First, you could ask for recommendations from your friends and OS. Just don't ask Aunt $\qquad$ —she only PLLRAL NOLIN PERSON IN ROOM (FEMALE) reads $\qquad$ books with $\qquad$ -ripping goddesses on the cover. If your friends and family are no help, try checking out the
$\qquad$ Review in The $\qquad$ Times. If the $\qquad$ NOLN A CITY PLLIRAL NOLIN
featured there are too $\qquad$ for your taste, try something a little ADJECTIVE more low- $\qquad$ , like $\qquad$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

# MAD $(9) L I B S$ 

## 至BOOKish Mad Libs

There are many $\operatorname{Sin}_{\text {ADJECTIVE }}$ bagel $_{10}$ read. First, you could ask for recommendations from your friends and
 on the cover. If your friends and family are no help, try checking out the
$\qquad$ Review in The $\qquad$ Times. If the $\qquad$ A CITY PLLRAL NOLIN
featured there are too $\qquad$ for your taste, try something a little ADJECTIVE more low- $\qquad$ , like $\qquad$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

## MAD $(9) L I B S$

## 至BOOKish Mad Libs

There are many $\int_{\text {ADJECTIVE }}^{\text {Pu }}$ ways to choose a/an bagel $_{10}$ read. First, you could ask for recommendations from your friends and $\underset{\text { PLLIRAL NOLIN }}{903}$. Just don't ask Aunt $\frac{\text { Mary }}{\text { PERSON IN ROOM (FEMALE) }}$ —she only reads $\underbrace{\text { G }}_{\text {ADJECTIVE }}$ -ripping goddesses on the cover. If your friends and family are no help, try checking out the
$\qquad$ Review in The $\qquad$ Times. If the $\qquad$ NOLN A CITY PLLLRAL NOLIN
featured there are too $\qquad$ for your taste, try something a little ADJECTIVE more low- $\qquad$ , like $\qquad$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

## MAD $(9) L I B S$

## 至BOOKish Mad Libs

There are many $\operatorname{Sin}_{\text {ADJECTIVE }}$ bagel $_{10}$ read. First, you could ask for recommendations from your friends and e9. $\underbrace{}_{\text {PLLLRAL NOLN }}$. Just don't ask Aunt $\frac{\text { Mary }}{\text { PERSON IN R POM (FEMALE) }}$ —she only $\operatorname{reads} \underbrace{C \gamma}_{\text {ADJECTIVE }} y_{\text {fooks with }}^{S c a r t}$ ARTICLE OF CLOTHING ripping goddesses on the cover. If your friends and family are no help, try checking out the
$\qquad$ Review in The $\qquad$ Times. If the $\qquad$ PLLLRAL NOLIN
featured there are too $\qquad$ for your taste, try something a little ADJECTIVE more low- $\qquad$ , like $\qquad$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

## MAD $(9) L I B S$

## 

There are many $\frac{5 i\left\|\|_{\text {ADJECTIVE }}\right.}{\text { ways to choose a/an } \frac{\text { bagel }}{\text { NOLN }} \text { to }}$ read. First, you could ask for recommendations from your friends and e9 $\underbrace{3}_{\text {PLLIRAL NOLN }}$. Just don't ask Aunt Mary —she only
 on the cover. If your friends and family are no help, try checking out the
slide-rule ${ }_{\text {Review in }}$ The $\qquad$ Times. If the $\qquad$ noun A CITY PLLLRAL NOLIN
featured there are too $\qquad$ for your taste, try something a little ADJECTIVE
more low- $\qquad$ , like $\qquad$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

## 

## 三BOOKish Mad Libs

There are many Sidways to choose alan $_{\text {ADJECTIVE }}^{\text {NOLN }}$
read. First, you could ask for recommendations from your friends and
egg ${ }_{\text {Just don't ask Aunt }}$ Mary $^{9}$
PLURAL NOLN PERSON IN RCOM (FEMALE)

on the cover. If your friends and family are no help, try checking out the
 PLURAL NOLIN featured there are too $\qquad$ for your taste, try something a little ADJECTIVE
more low- $\qquad$ , like $\qquad$ : The $\qquad$ Magazine. You could also choose a book the

Magazine, or $\qquad$ PLURAL NOLIN

## MAD © $(\mathbb{O})\left[\mathrm{IBS}^{\circ}\right.$

## 三BOOKish Mad Libs

There are many $\left.\frac{\text { sill }}{\text { wojerrive }} y_{\text {ways to chose alan }}^{\text {bage }}\right|_{\text {to }}$
read. First, you could ask for recommendations from your friends and
eg9S Just don't ask Aunt Mary
PERSON IN RCOM (FEMALE)

on the cover. If your friends and family are no help, try checking out the $\underbrace{\text { side-ruleview in the }}_{\text {NoLn }} \underbrace{\text { Hyde park }}_{\text {ACITY }}$ fimes. If the $\underbrace{\text { bookcases }}_{\text {PLLIRAL Noun }}$
featured there are too $\qquad$ for your taste, try something a little ADJECTIVE —she only
LLIRAL NOLIN


## MAD © $(\mathbb{O})\left[\mathrm{IBS}^{\circ}\right.$

## 至BOOKish Mad Libs

There are many silly bagel. read. First, you could ask for recommendations from your friends and egg ${ }_{\text {Just don't ask Aunt }}$ Mary $^{9}$ PLURAL NOLN PERSON IN RCOM (FEMALE)
 on the cover. If your friends and family are no help, try checking out the
 more low- $\qquad$ , like $\qquad$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

## MAD (®) LIBS

## 至BOOKish Mad Libs

There are many silly bagel. read. First, you could ask for recommendations from your friends and

# eggs matem Mary 

PERSON IN RCOM (FEMALE)

on the cover. If your friends and family are no help, try checking out the
 more low- $\frac{\text { stomach }}{\text { PART OF THE BODY }}$, like $\qquad$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

## MAD (®) LIBS

## 至BOOKish Mad Libs

There are many silly bagel. read. First, you could ask for recommendations from your friends and

# eggs matem Mary 

PERSON IN RCOM (FEMALE)

on the cover. If your friends and family are no help, try checking out the
 more low- $\frac{s^{\dagger} O M 2 C A^{\text {ADJECTIVE }}}{\text { PART OF THE BODY }}$, like $\frac{M}{\text { LETTER OF THE ALPHABET }}$ : The $\qquad$
Magazine, or $\qquad$ Magazine. You could also choose a book the

## MAD (e) LIBS

## 至BOOKish Mad Libs

There are many silly bagel. read. First, you could ask for recommendations from your friends and

# eggs $S_{\text {Ier torit ate ant. }}$ Mary 

PERSON IN RCOM (FEMALE)

on the cover. If your friends and family are no help, try checking out the
 more low- $\frac{s^{\top} O M 2 C h}{\text { PART OF THE BODY }}$, like $\frac{M}{\text { LETTER OF THE ALPHABET }}$
:The $\frac{\text { Barack }}{\text { CELEBRITY }}$

Magazine, or $\qquad$ PLLLRAL NOLN Magazine. You could also choose a book the

#  

## 至BOOKish Mad Lis

There are many silly bagel． read．First，you could ask for recommendations from your friends and

# eggs $S_{\text {Ier tort ate ant．}}$ Mary 

PERSON IN ROM（FEMALE）

on the cover．If your friends and family are no help，try checking out the
 more low－$\frac{s^{\dagger} 0 M a C h}{\text { PART OF ANE }}$ ADJECTIVE,$~ l i k e \frac{M}{\text { LETTER OF THE ALPHABET }}$
：The $\frac{\text { Barack }}{\text { CELEBRITY }}$
Magazine，or Magazine．You could also choose a book the

## Your startup name:

all sorts of ___ with friends, family, colleagues, or even (PLURAL NOUN)
$\qquad$ but didn't
$\qquad$
 favorite
$\qquad$ bakery when you wished you could pull out your i $\qquad$
and use it to quickly $\qquad$ all your latest $\qquad$ and post it to your (TRIVIAL STUFF YOU LIKE)
(SOCIAL MEDIA PROFILE NO ONE READS)
where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to (GERUND STARTING WITH "CROWD")

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

(MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)
all sorts of ___ with friends, family, colleagues, or even (PLURAL NOUN)
$\qquad$ but didn't
$\qquad$
 favorite
$\qquad$ bakery when you wished you could pull out your i $\qquad$
and use it to quickly $\qquad$ all your latest $\qquad$ and post it to your (TRIVIAL STUFF YOU LIKE)
(SOCIAL MEDIA PROFILE NO ONE READS)
where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to (GERUND STARTING WITH "CROWD")

> (MISSPELLED VERSION OF REAL WORD, ENDING IN -LY OR -IO)
> all sorts of ___ with friends, family, colleagues, or even (PLURAL NOUN) , but didn’t (MINOR CELEBRITY)
> know how? Maybe you were just strolling the sidewalks of past your favorite $\qquad$ bakery when you wished you could pull out your i $\qquad$
and use it to quickly $\qquad$ all your latest $\qquad$ and post it to your (TRIVIAL STUFF YOU LIKE) is the answer. The fact is, we live in a world
(SOCIAL MEDIA PROFILE NO ONE READS)
where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to (GERUND STARTING WITH "CROWD")

Your startup name: Rheummate-ly. Have you eve
all sorts of Sneakers with friends, family, colleagues, or even (PLURAL NOUN) , but didn't (MINOR CELEBRITY)
know how? Maybe you were just strolling the sidewalks of $\qquad$ past your favorite bakery when you wished you could pull out your i
(OBSCURE ETHNICITY) and use it to quickly $\qquad$ all your latest $\qquad$ and post it to your (TRIVIAL STUFF YOU LIKE)
$\qquad$
? $\qquad$ is the answer. The fact is, we live in a world (SOCIAL MEDIA PROFILE NO ONE READS)
where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to
(GERUND STARTING WITH "CROWD")
(STARTUP NAME) anything you want through the cloud. All this

# Your startup name: Rheummate-1y. Have you ever wanted tojuminin 

know how? Maybe you were just strolling the sidewalks of $\qquad$ past your (GENTRIFIED NEIGHBORHOOD)
bakery when you wished you could pull out your i $\qquad$
$\qquad$ and post it to your
(VERB ENDING IN -ATE) (TRIVIAL STUFF YOU LIKE) is the answer. The fact is, we live in a world
(SOCIAL MEDIA PROFILE NO ONE READS)
where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to (GERUND STARTING WITH "CROWD")

#  

 know how? Maybe you were just strolling the sidewalks of Kenwood past your (GENTRIFIED NEIGHBORHOOD) favorite bakery when you wished you could pull out your $i$(OBSCURE ETHNICITY) and use it to quickly $\qquad$ and post it to your
(VERB ENDING IN -ATE) (TRIVIAL STUFF YOU LIKE) is the answer. The fact is, we live in a world
(SOCIAL MEDIA PROFILE NO ONE READS)
where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to
(GERUND STARTING WITH "CROWD")
 all sorts of $\frac{\text { SneakerS }}{\text { (puradal noun) }}$ with friends, family, colleagues, or even $\frac{\text { Romney }}{\text { (mmor celegerin }}$, but didn't know how? Maybe you were just strolling the sidewalks of Kenwood past your favorite $\frac{\text { IgDO }}{\text { (Ooscune ETHNOCTM) }}$ bakery when you wished you could pull out your i $\qquad$ and use it to quickly $\qquad$ all your latest $\qquad$ and post it to your $\xrightarrow{\text { socall mena }}$ ? $\qquad$ is the answer. The fact is, we live in a world where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to (GERUNo Startina with "Crowo) anything you want through the cloud. All this $\qquad$

Your startup name: Rheummate-ly. Have you ever wanted to 1 Mimp all sorts of $\frac{\text { SneakerS }}{\text { (puvant noum) }}$ with friends, family, colleagues, or even $\frac{\text { Romney }}{\text { (mmor celegerin }}$, but didn't know how? Maybe you were just strolling the sidewalks of Kenwood past your
 and use it to quickly $\qquad$ all your latest $\qquad$ and post it to your
 $\qquad$ is the answer. The fact is, we live in a world where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to (GERUND STARTING WITH "CROWD") anything you want through the cloud. All this $\qquad$

Your startup name: Rheummate-ly. Have you ever wanted to 1 Mimp
 know how? Maybe you were just strolling the sidewalks of Kenwood past your
 and use it to quickly $\frac{\text { differentiate }}{\frac{\text { Nererenonanan-are }}{}}$ all your latest $\qquad$ and post it to your ? $\qquad$ is the answer. The fact is, we live in a world (SOCIAL MEDIA PROFILE NO ONE READS) (STARTUP NAME) where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to (GERUND STARTING WITH "CROWD") anything you want through the cloud. All this $\qquad$

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

 ? $\qquad$ is the answer. The fact is, we live in a world (soctal meial proofle No one neans (STARTUP NAME) where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to (GERUND STARTING WITH "CROWD") anything you want through the cloud. All this $\qquad$

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

 GoogleMine ? $\qquad$ is the answer. The fact is, we live in a world $\overline{(S O C I A L ~ M E D I A ~ P R O F I L E ~ N O ~ O N E ~ R E A D S) ~}$ where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to
$\qquad$ anything you want through the cloud. All this $\qquad$

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


 where $\qquad$ will revolutionize $\qquad$ , and pretty soon you'll be able to
$\qquad$ anything you want through the cloud. All this $\qquad$
 all sorts of Sneakers with friends, family, colleagues, or even $\frac{\text { Rompant noum) }}{\text { (mmoocecelearin) }}$, but didn't know how? Maybe you were just strolling the sidewalks of Kenwood past your

 $\frac{\text { GoogleMine }}{\text { (socomumela Profluno one beans) }}$ ? $\frac{\text { Rheummate-ly }}{\text { (startup name) }}$ is the answer. The fact is, we live in a world where crowd-licking will revolutionize $\qquad$ , and pretty soon you'll be able to (GERUND STARTING WITH "CROWD")
$\qquad$ anything you want through the cloud. All this $\qquad$

Your startup name: Rheummate-ly. Have you ever wanted to jump all sorts of $\frac{\text { Sneakers }}{\text { puvenchown }}$ with friends, family, colleagues, or even $\frac{\text { Romney }}{\text { smmonectearann }}$, but didn’ know how? Maybe you were just strolling the sidewalks of Kenwood past your favorite Igbo bakery when you wished you could pull out your i Soap
 GoogleMine ? Rheummate-ly is the answer. The fact is, we live in a world

$\qquad$ anything you want through the cloud. All this $\qquad$

Your startup name: Rheummate-ly. Have you ever wanted to jump all sorts of $\frac{\text { Sneakers }}{\text { puvenchown }}$ with friends, family, colleagues, or even $\frac{\text { Romney }}{\text { smmorectearan }}$, but didn’ know how? Maybe you were just strolling the sidewalks of Kenwood past your favorite Igbo bakery when you wished you could pull out your i Soap
 GoogleMine ? Rheummate-ly is the answer. The fact is, we live in a world
 thinkify anything you want through the cloud. All this $\qquad$

## Empirically, not so obvious: 'the' versus 'a'

the
Right

| of | 9629 | same | 601 | of | 1467 | few | 344 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| in | 5567 | frst | 547 | in | 1321 | little | 292 |
| to | 3433 | other | 388 | as | 914 | man | 253 |
| on | 2315 | most | 380 | with | 879 | new | 248 |
| and | 2143 | United | 367 | is | 872 | good | 221 |
| for | 1766 | world | 338 | was | 783 | small | 161 |
| at | 1511 | *j | 264 | for | 740 | great | 153 |
| with | 1481 | time | 243 | to | 655 | very | 138 |
| that | 1392 | two | 232 | and | 615 | long | 135 |
| from | 1352 | new | 231 | on | 476 | number | 117 |
| by | 1315 | way | 226 | by | 413 | large | 110 |
| is | 793 | only | 207 | be | 318 | year | 99 |
| as | 698 | end | 204 | at | 256 | single | 99 |
| into | 673 | last | 202 | have | 256 | moment | 86 |
| was | 531 | next | 185 | had | 246 | lot | 86 |
| $:$ | 466 | state | 182 | like | 242 | more | 84 |
| all | 424 | whole | 178 | from | 229 | time | 79 |
| In | 412 | old | 171 | such | 222 | matter | 79 |



Encarta (encyclopedia) 4,000 words


## English

# Dynamic view: English color codes <br> Verbs: 'bare' verb (jump) red <br> Verbs: past tense(jumped, bought) blue <br> Verbs: auxiliary (should, can) green <br> Prepositions (from, to, up, down aqua Adjectives <br> purple <br> Cities <br> gray <br> Nouns <br> pink 

Dynamic view: French color codes
Infinitives red
Prepositions light blue
Past participles blue
Adjectives purple
Cities
gray
Masculine nouns pink
Feminine nouns light green
Inflected verbs light gray

## Finding regular and irregular past tense forms





