

**facebook**

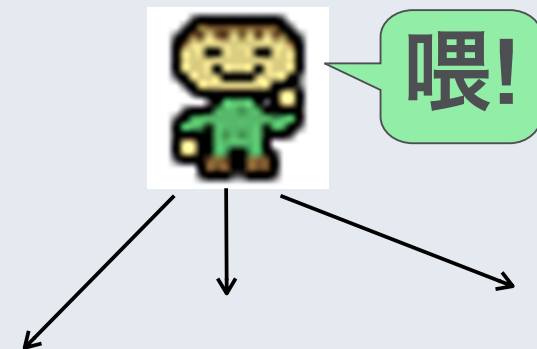
**facebook**

# Social (distributed) language modeling, clustering and dialectometry

David Ellis  
August 7, 2009  
TextGraphs (@ACL)

# Why?

- Users
  - connect w/ wider audience
  - access to more info (search)
- Facebook
  - growth, open communication
- Personal
  - Google (mach.) => FB (crowd)
- Research
  - corpora (public, anonymized)
  - tools (API, open source)

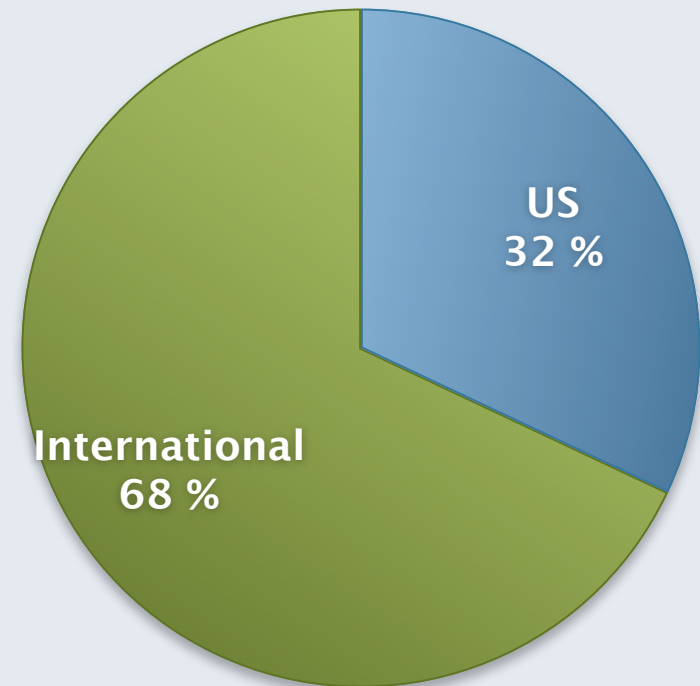


# Sharing with the world

- Facebook is live in 64 languages
- 55 more are in translation (مال شما؟)
- Over 2/3 of users are outside the US
- >50% of Internet users in Chile are active Facebook members
- Help 350,000 apps go global



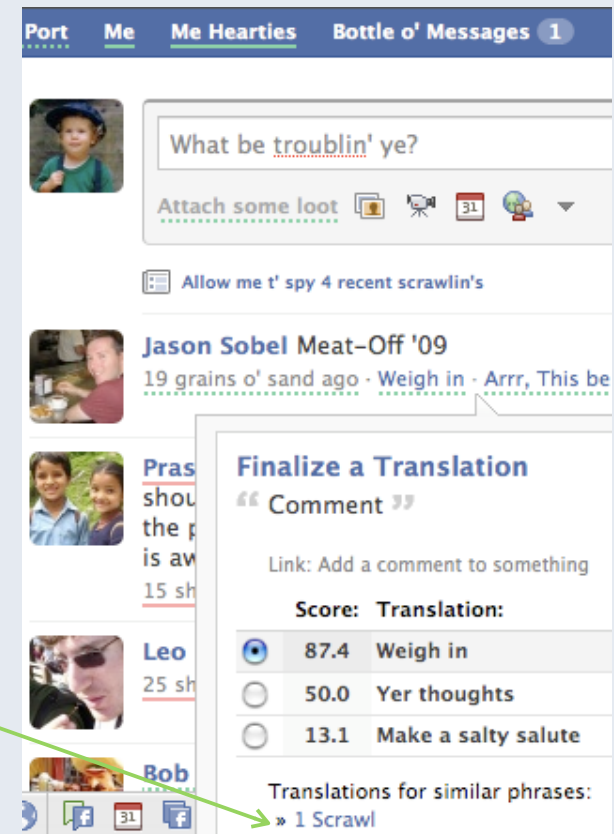
- Share docs (social productivity)
- Local insight => travel info
- Explore common interests
- Play games (MasMultilingualOn)
- Taste virtual cuisine (and learn to cook the real kind)



# Translations on Facebook

- Community-driven, open
  - glossary => inline (voting) => local UI
- French <24h
- English (Pirate)
  - ~30,000 users
  - >50,000 (phrase) translations submitted
- Similar suggestions
  - fuzzy clustering, regional sharing, MT\*
- आप दुनिया चैट मदद कर सकता है!

\*experiments with open source toolkits in progress



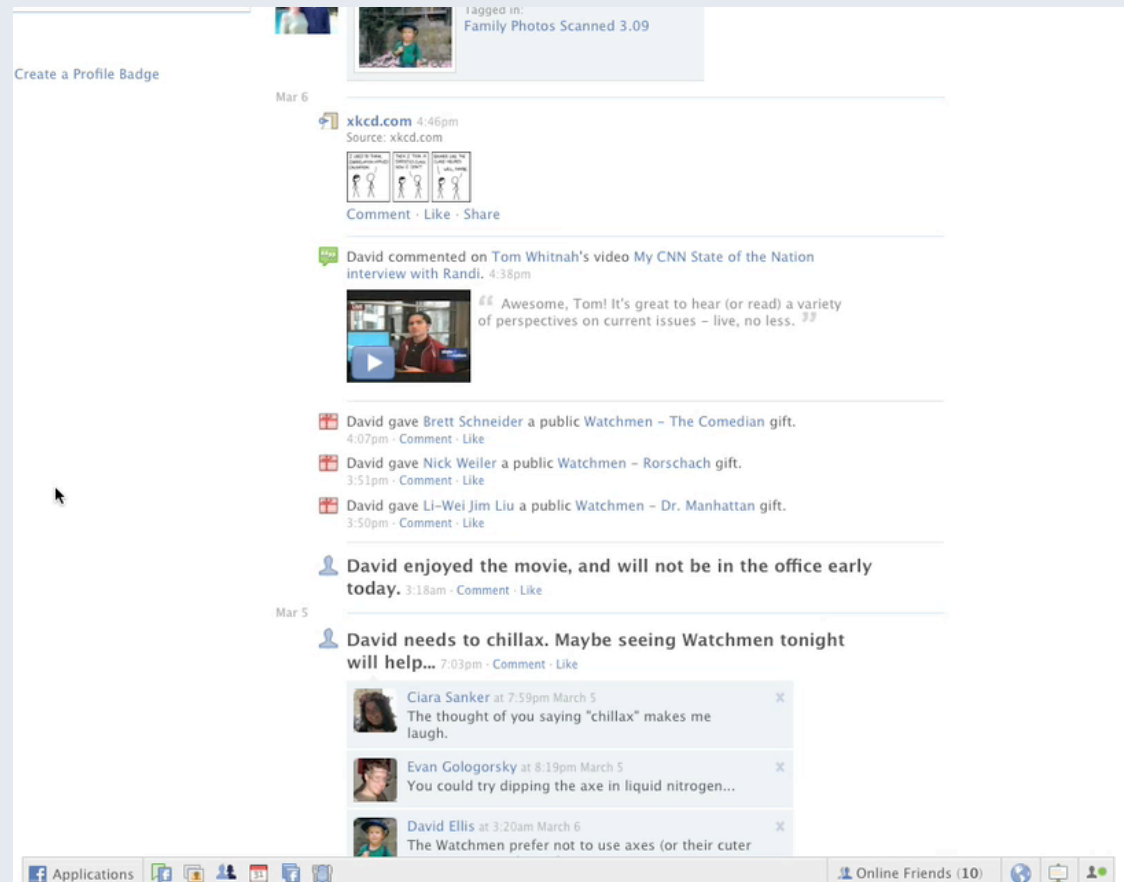
# Dynamic text as {token}s

- {name} depends on
  - person
  - context
- food pyramid app
  - “Eat a(n) {fruit}!”
  - kiwi? apple?
- translations
  - “Syö(kää) {fruit}!”
  - sing/plural?
    - addressee (of imperative)



# Linguistic tools

- Rules
  - orthographic
  - phonological
- Features
  - gender
    - Català
  - number (si, du, pl)
    - Russian, Arabic
- Models
  - language
  - pronunciation



# Linguistic tools

- Variation
  - native explosion
    - 1-to-many
  - cluster folding
    - many-to-1
- detection
  - spam (text)
  - botnet (usage)
  - authentic?
- “Güd [pet]tRn’d!” –EG

facebook Home Profile Friends Inbox 4 David Ellis Settings Logout Search

Home | Using Translations | Go Translate | Leaderboard | Admin Panel Help | About

All strings (translated, approved) Application: Facebook

Dashboard Browse Phrases Top Pages Preferences

Add Glossary Term Search for: number in: Original English Search

Advanced Search

Displaying 25 phrases out of 74 phrases. 1 2 3 Next

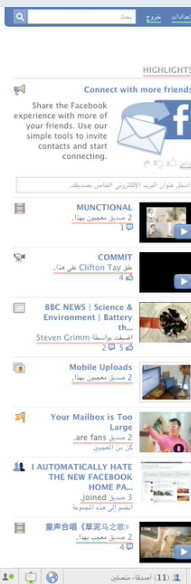
Created: Anytime ▾	Type: All Strings ▾	Language: Arabic ▾	Score	Status ▾	More
{name1} changed his phone number to {new-phone-number}. It has been added to your {phonebook-link}.	{name1} غير رقم هاتفه إلى {new-phone-number}. و تمت إضافته إلى {phonebook-link}	0	Approved	1 more	
Online Friends {number-available}	{number-available} أصدقاء متصلين	-8	Approved	2 more	
This error occurred because you have reached the maximum number of albums.	حدث هذا الخطأ لأنك بلغت العدد الأقصى للصور.	0	Approved		
{number} photos	Click here to see/modify the token variation rules			<a href="#">See All Variants</a>	
{name} added a new mobile number: {phone-number}.	{name} أدخل رقم هاتف محمول جديد: {phone-number}	0	Approved	1 more	
Could not validate the your phone number, please check the number and try again.	تعذر التحقق من صحة رقم الهاتف، يرجى التحقق من الرقم والمحاولة من جديد.	0	Approved		
Existing user account matches the phone number supplied. If this is an error, please contact mobile-info@facebook.com	يتطابق حساب المستخدم الموجود مع رقم الهاتف المتوفر. إذا كان هذا خطأ، يرجى الاتصال بـ mobile-info@facebook.com	0	Approved		
Could not send confirmation sms, please check phone number and try again.	تعذر إرسال رسالة تأكيد، يرجى التحقق من رقم الهاتف والمحاولة من جديد.	0	Approved		
{name1} and {number others} like this.	{name1} و {number others} معجبين.	0	Approved		
{number} of your friends	{number} من أصدقائك	4	Approved	3 more	



# Deep localization

- Transform layout (not just text)

مُد



أرحب في أعماق روحك...  
قبل أن يغادر ، انقر هنا  
لالأقران في منجم  
هل لي أن تنضم في  
المستقبل (في سنغافورة ،  
كاليفورنيا وراء ذلك الوقت)؟

# Fu(ture|n)

- Open innovation
  - experimental API
    - access [via Hive] to publicly shared, anonymized data
      - social graph
      - language model
  - parallel (dynamic) corpus
    - named entity annotated
    - growing, with snapshots
- Instant, free multimodal interpretation
  - <http://www.colips.org/blog/acl-ijcnlp-2009/index.php/lieutenant-commander-data-star-trek/comment-page-1/#comment-292>

**Loc**

||  
v

**Fuzz**

||  
v

**SMT**

||  
v

**?**

**Questions?**  
**Kiitti kaikille!**

# facebook

(c) 2007 Facebook, Inc. or its licensors. "Facebook" is a registered trademark of Facebook, Inc.. All rights reserved. 1.0