



Multilingualism at Google

Better serving users in more than one language

Luke Swartz
Product Manager, Google Internationalization Engineering
lswartz@google.com

Some myths about Multilingualism...

Myth: Most people speak only 1 language

Estimated >50% of world interacts with >1 language regularly

More people speak English as a 2nd language than natively (nearly 2X)

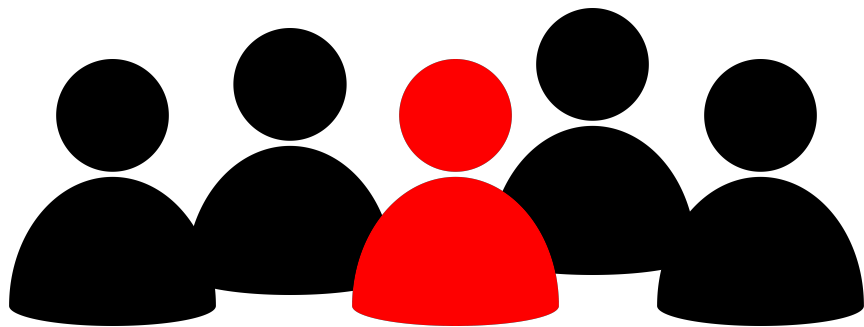
54% of EU residents are able to hold conversation in at least 2 languages

India has >20 “official” languages

sources: [François Grosjean](#), [Wikipedia/Ethnologue](#), [European Commission](#),
Eighth Schedule to [Indian Constitution](#)

Myth: OK, but most *Americans* speak only 1 language

1 in 5 US residents speak a language other than English *at home*



source: [2011 US Census Community Survey](#)

Myth: OK, but most *internet users* only use 1 language

>30%

of Google users have activity
(searches, etc.) in >1 language



Multilingual Challenges & Opportunities

Reasons People are Multilingual

National Diversity

Migration

Travel

People moving among language communities



Business

Content

Education

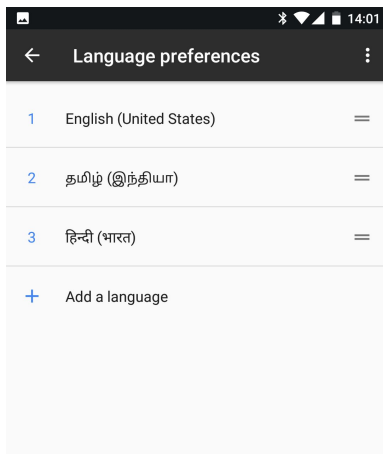
Ideas/things moving among language communities



Only a few users explicitly tell us they're multilingual



myaccount.google.com
Language Settings



Android 7.0+ (Nougat)
Language Settings

Fluency versus Preference

Different problems:

- **Fluency:** (Offer to) translate this content?
- **Preference:** Predict what language the user will use / want

Sample reasons fluency $\neq$ preference:

- Mutually ~intelligible languages (Czech versus Slovak, zh-Hans versus zh-Hant)
- “Immersion” language learners / “fully assimilated” immigrants
- ...also preference is more context-dependent
 - *both* vary reading / writing / listening / speaking

Multilingualism in Context

Example contexts:

- **Person:** I chat in Japanese to Bob and in French to Alice
- **Place:** I speak Tamil at home and Hindi at work
- **Medium:** I watch TV in German and read in Turkish
- **Topic:** I search for recipes in Italian and religious texts in Arabic

...and code switching in single utterance (e.g. [cuando es el black friday 2017])

Deep Dive: India

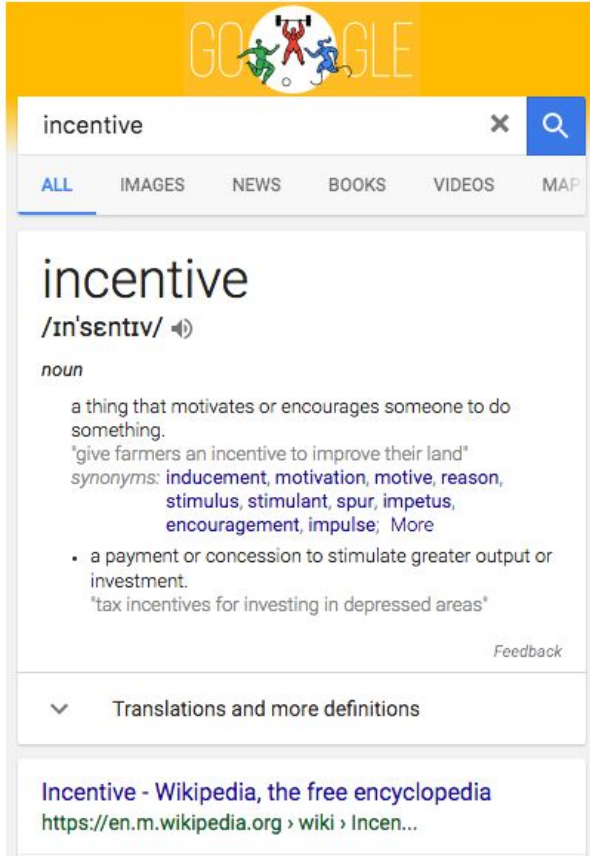
Story: Google Search in India

User at a parent-teacher conference...

Teacher: “Your child needs an incentive to learn”

User: Googles [incentive]

Story: Google Search in India



User at a parent-teacher conference...

Teacher: "Your child needs an incentive to learn"

User: Googles [incentive]

"Great, before I didn't know 1 word...now I don't know 20 words!"

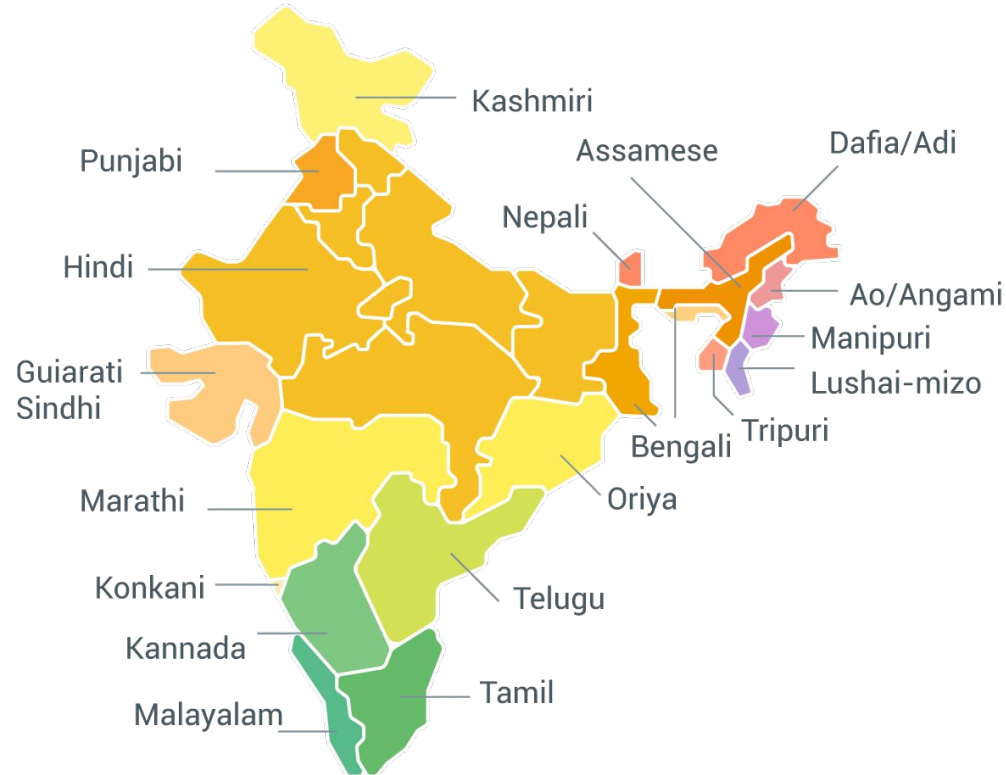
Story: Google Search in India



With the right data, we could recognize that this user knows Hindi better than English

...and show Translate instead of/in addition to Dictionary

India: High Linguistic Diversity



Myths about Indian language use

The device/UI language is the user's primary/native tongue

OK, well the user at least *understands* the device/UI language

Users know how to input text in the languages they speak

A smartphone is used by one person

Users use 1 language at a time, and want content in 1 language at a time

The language someone uses is the language they want (query lang == desired content lang)

“Hinglish”

- Hindi in Latin Script >> Hindi in Devanagari for Google queries
- Mixing Hindi & English → more and more common in Indian culture
 - [Study](#) compared 1982, 1992, 2004 films → Hindi & English mixed dialog went from 28% to 47%!
 - Another study looked at Big Boss (~Indian Big Brother): “...zero contestants who were able to produce Monolingual Hindi”
- Variations: English in Hindi, Hindi in English, and a pure mix
 - thus: just because someone uses English words doesn't mean they're fluent
 - also: English is frequently mis-spelled (in different ways from how people in US misspell!)
 - important to include (some) English words in, for example, keyboard dictionaries

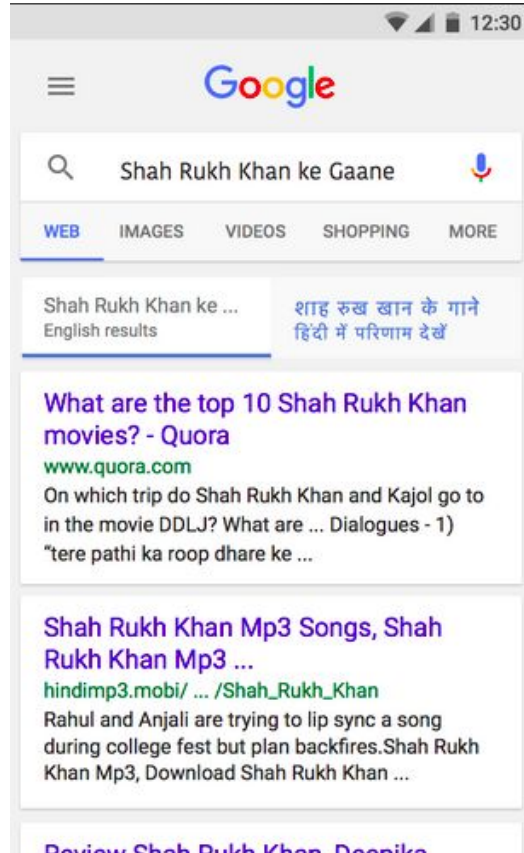
● ...and similar phenomena for other Indic languages (+ other countries)

India: Most phones are set to English...why?

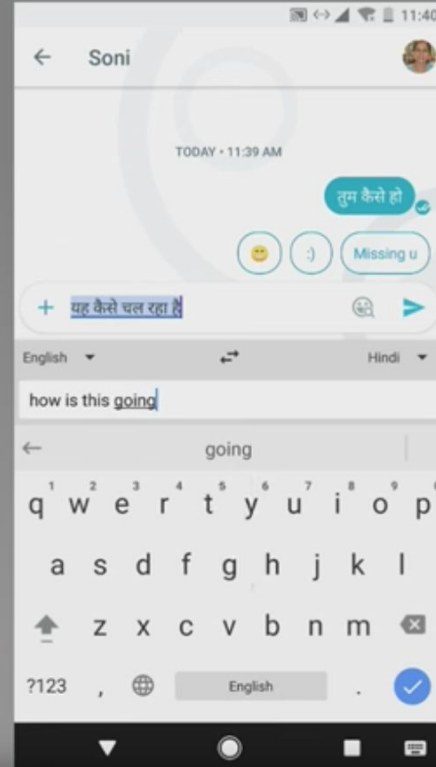
- “Neutral”
- “Aspirational”
- “Status”
- Setup / Settings
- Bad Localization

Multilingualism and Google products

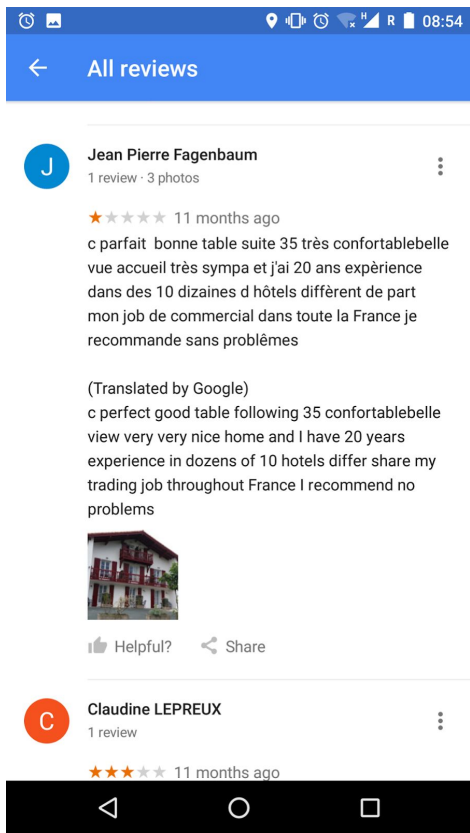
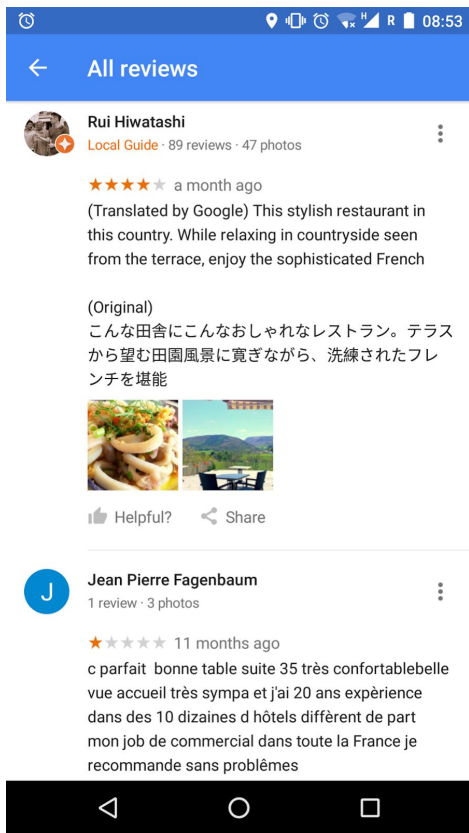
“Tabbed Browsing” Search for Hindi-speaking India



Android Go Multilingual gBoard Keyboard



Local Review Translation on Google Maps



Luke's language
profile = {en, it, fr}

Google Noto Font

No Tofu (~all scripts in Unicode supported)

Similar style: mix and match scripts (no “ransom note” effect)



More Multilingual Challenges

Language Identification

- cv bien hmd w enti
 - French/Arabic: ça va bien alhamdulillah wa enti (الحمد لله و إنتي) $\approx$ ok how r u?
- safety tutor
 - Italian (autostrade speed control system)
- handy
 - German (“mobile phone”)
- iphone 8
 - almost any language!

Spoken versus written language

Rising importance of audio-primary/only interfaces (Google Assistant, etc.)

What is “Arabic”? “Chinese”?

Language versus dialect / accent

Lingua francas versus native (spoken) tongues

Very long tail—how to scale? How to protect minority languages?

有疑问？

¿Tiene alguna pregunta?

Bạn có câu hỏi?

هل لديك أسئلة؟

Frågor?

Haben sie noch Fragen?

질문이 있으세요?

Onko sinulla kysyttävää?

Máte otázky?

Des questions ?

Ada pertanyaan?

Questions?

Spørsmål?

Perguntas?

หากมีข้อสงสัย

Har du spørsmål?

Vragen?

Domande?

שאלות?

質問があればどうぞ

Pytania?

Есть вопросы?

Sorularınız mı var?

Маєте запитання?

Έχετε ερωτήσεις;

પ્રશ્ન?