Developing Multilingual OCR and Handwriting Recognition at Google

Observations and Reflections

Ashok Popat

Research Scientist, Google Inc.

IAPR Summer School, Jaipur: Jan 23 2017

Google

Joint work with

Jon Baccash

Marcos Calvo

Victor Cărbune

Thomas Deselaers

Karel Driesen

Sandro Feuz

Yasuhisa Fujii

Philippe Gervais

Pedro Gonnet

Patrick Hurst

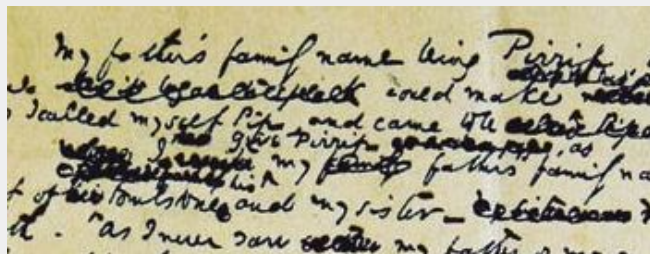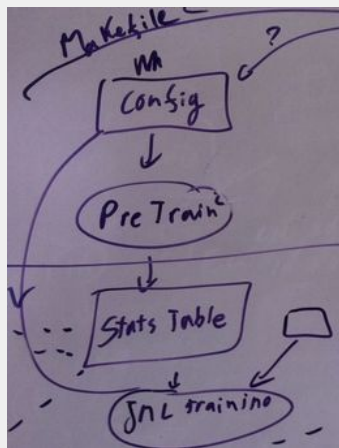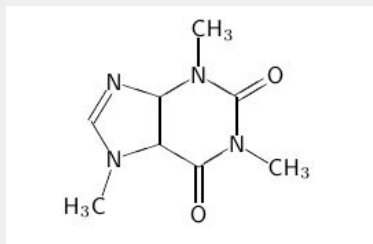Henry Rowley

Li-Lun Wang

Optical Character Recognition

Ashok Popat, Jan 23, 2016

OCR in Google Products

Google Translate

Google books

Google Keep - notes and lists

Use your camera for instant text translation

SORTIE

EXIT

Google Drive
Keep everything. Share anything.
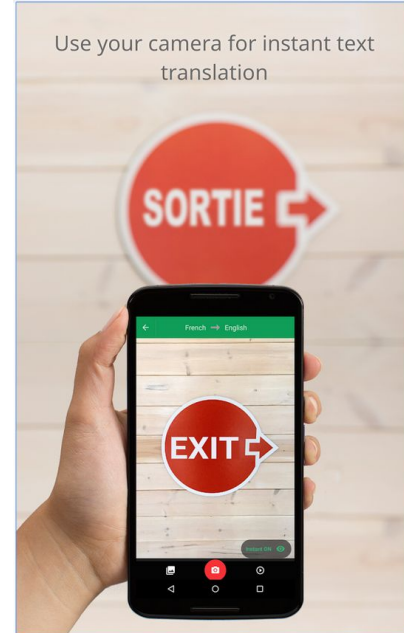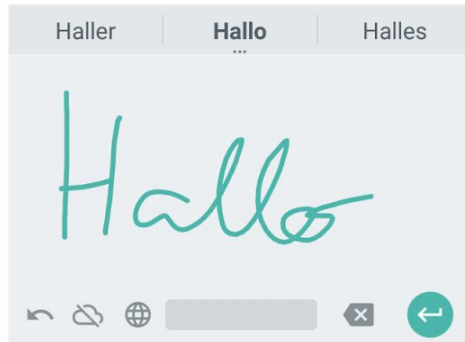
Google Handwriting Input

on-device recognition
> 80 languages + emoji
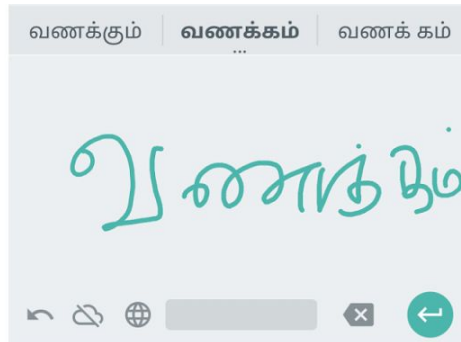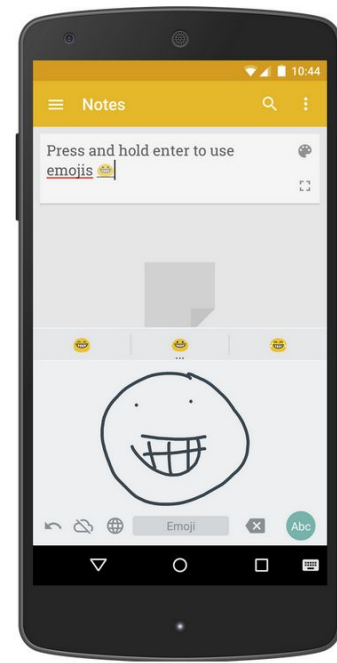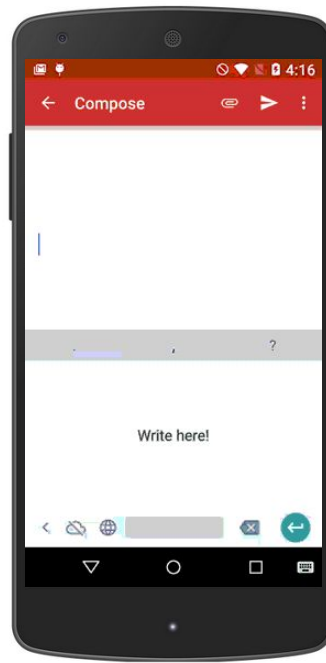
Google Translate for Android

Translate 2.3

enabled by default only for CJ
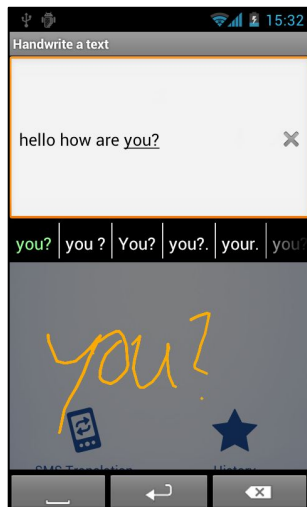
Translate 2.4+

enabled for all supported lang.
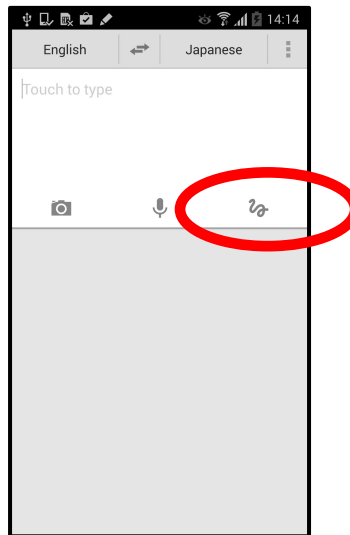
Handwrite for Mobile Search

write your search right on the
Google homepage

available on Google.com from
smartphone or tablet

can be activated or disabled in
mobile search settings

Google

Other Applications

Other input methods for Android

Input tools

… and more
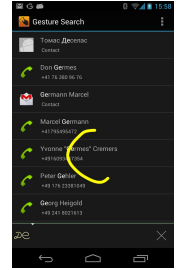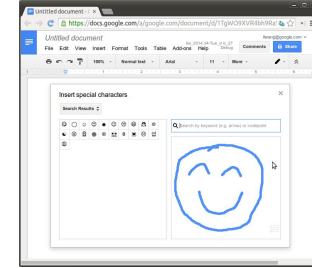
Google

Outline

- Multilingual OCR and On-line handwriting systems
- Research at Google
- Personal observations, reflections

Google

Part 1a: A multilingual OCR system

Examples from Google Books

Multiple scripts / languages on a page:

Examples from Google Books (cont.)

Per-word script and language variation:

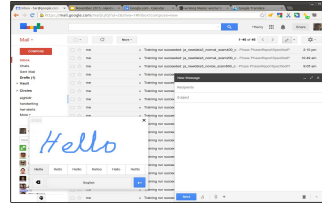Dieser Familie nähern sich die Slawonischen Worte лапа, лопасть und vorzüglich лопата (lapa, lopast und lopata, Pfote, Flügel einer Mütze, Schaufel), mit welcher ein Blatt durch seine Fläche eine grosse Aehnlichkeit hat. Daher nennen auch wir die grossen Blätter einiger Gewächse лапушникъ (lapuschnik, Klette). Wie sehr sich aber auch die letztern Worte

Examples from Google Books (cont.)

§. 6. Lesezeichen:

1. virâma, „Ruhezeichen" ˛, ein unten rechts an den Konsonanten angesetzter Strich, nimmt dem Konsonanten den a-Laut, mit welchem er nach § 1 und § 4 an und für sich zu sprechen ist. Der Virâma findet sich namentlich am Ende eines Wortes und manchmal anstatt einer Ligatur; z. B. यत् yat, आसीत् âsît, आगमद् देवी âgamad dêvî; क्र ktra kann auch geschrieben werden क्र.

2. avagraha, „Apostroph" ऽ, steht am Anfange eines Wortes zur Bezeichnung eines vorn abgefallenen a: z. B. दुःखितो ऽभवत् duḥkhitô 'bhavat.

Some of the 26 scripts of interest

Starting point: Markov-model-based approaches

- Document image decoding [Kopec and Chou, 1994]
  - Explicit model of typesetting process: seek to invert
  - Influenced by speech recognition methods
  - Extremely high accuracy when models match the data

- BBN Byblos system [Schwartz et al., 1996]
  - Treat text line like a speech waveform
  - Built on existing speech recognition system
  - First successful Arabic OCR

Ashok Popat, Jan 23, 2016

Generalization of the noisy channel model

- Speech approach

$$m' = \arg\max_m P(m)^{\alpha} P(x|m)$$
$$= \arg\max_m \alpha \log P(m) + \log P(x|m)$$

- Generalize to multiple feature functions

$$m' = \arg\max_m \sum_i \lambda_i h_i(x, m)$$

- Learn $\{\lambda\}$ via minimum error-rate training [Macherey et al. '08, Och '03]

Minimum Error Rate Training
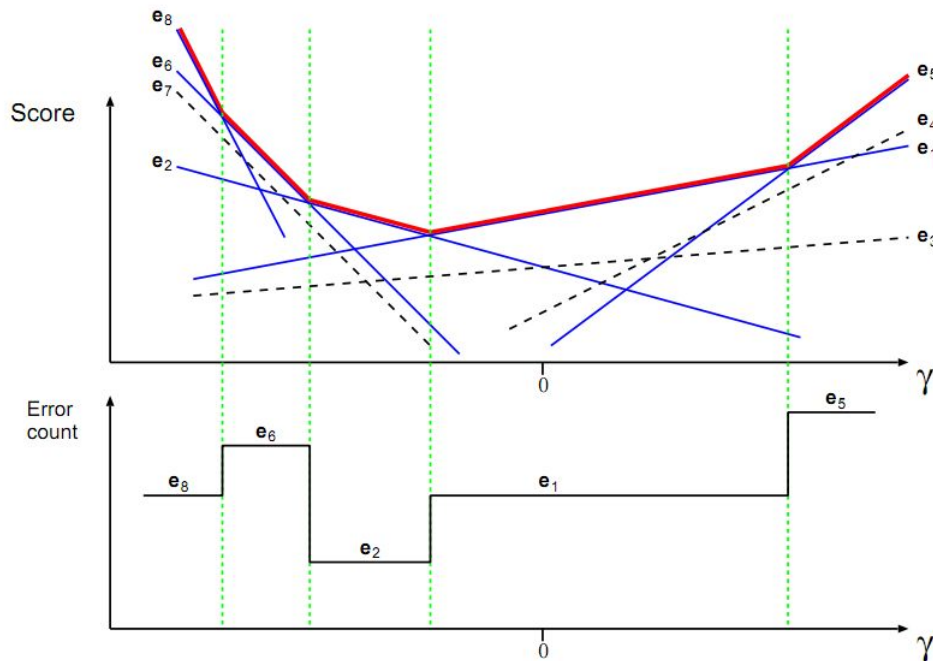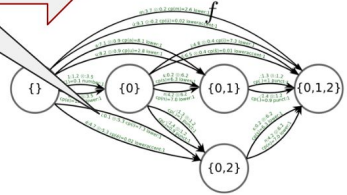
Macherey, Och, Thayer, Uszkoreit: Lattice-based Minimum Error Rate Training for Statistical Machine Translation. EMNLP 2008.

Training flow

Google

Technical evolution

- Optical model
  - GMM -> DNN
  - DNN -> LSTM
  - Sequential discriminative training of DNN/LSTM
- Language model
  - N-gram -> RNN-LM
- Decoding
  - Pruning algorithms designed for OCR
  - Automatic decoding parameter optimization
  - Fujii et al., ICDAR'15

Ashok Popat, Jan 23, 2016

Script ID (Li et al., 2015)

Ashok Popat, Jan 23, 2016
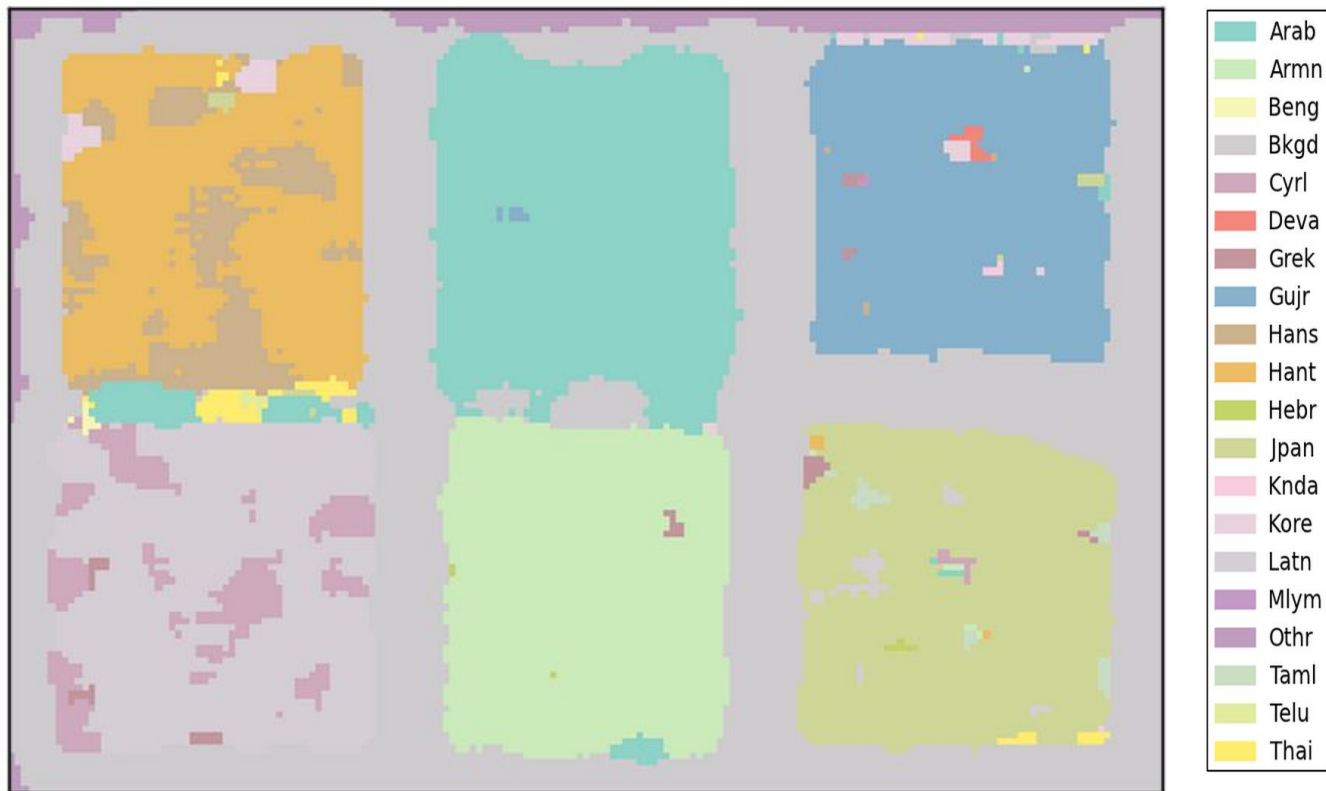
Regions not covered

Ashok Popat, Jan 23, 2016

Part 1b: A multilingual handwriting recognition system

Segment and Decode

Hidden Markov Models
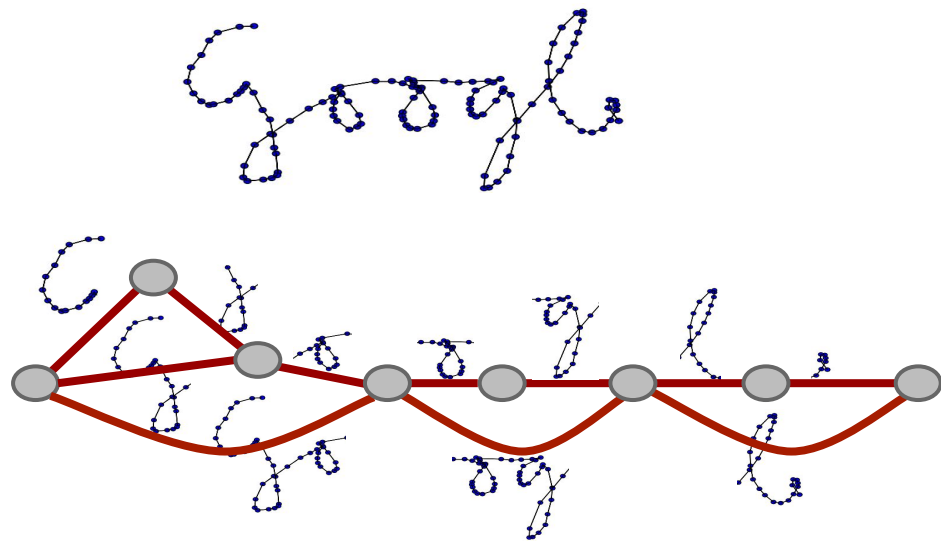
Apple Newton [Yaeger 1996]
Microsoft Tablet PC / Vista [Pittman 2007]
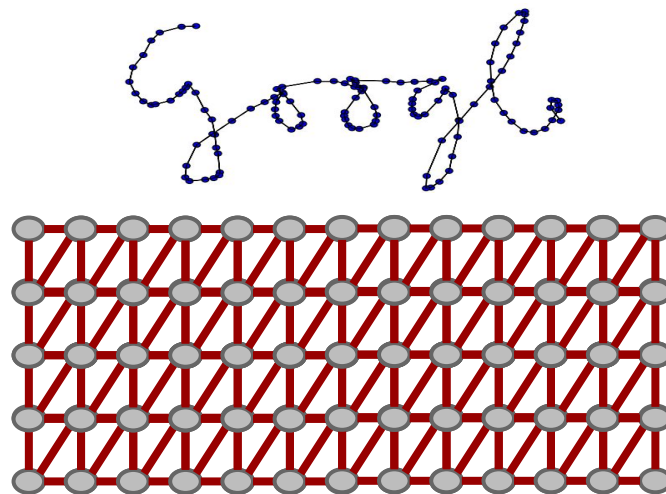
neural network variants:
Recurrent, Time-Delay,
Long Short-term Memory
[Jaeger 2001], [Graves 2009], ...

Google

Segment and Decode 1: Creating a segmentation lattice
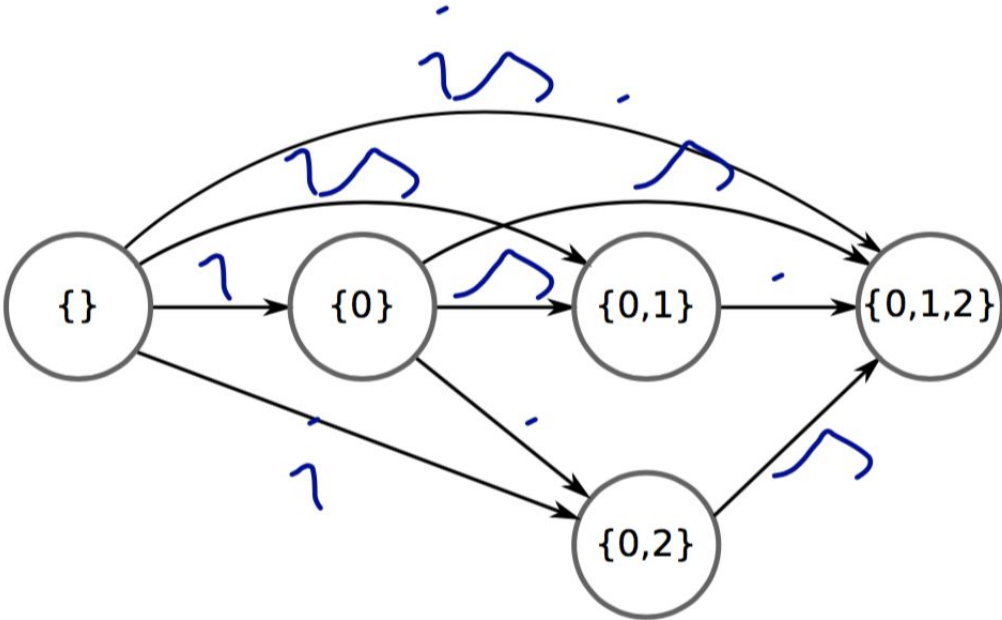
Segment and Decode 2: recognizing character hypotheses
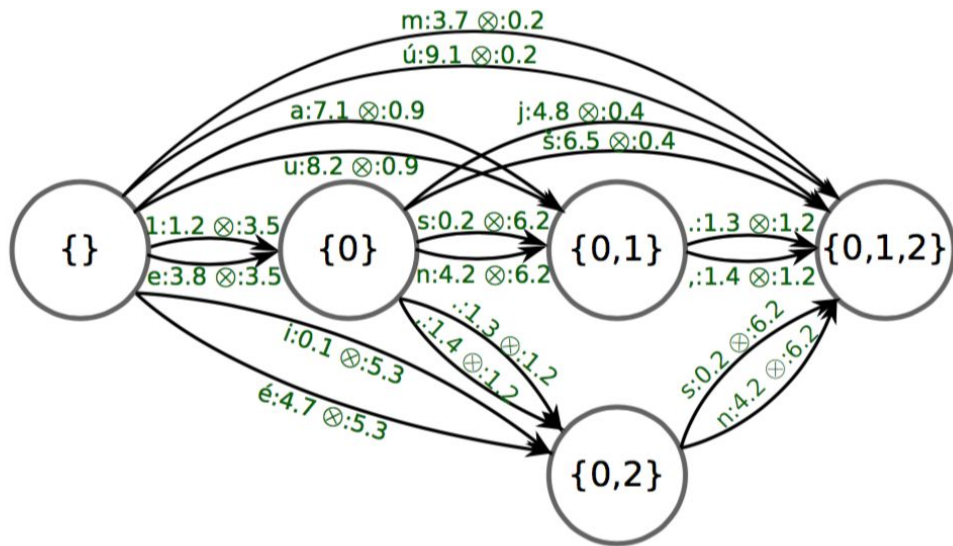
Segment and Decode 3: Decoding

Graph nodes: {}, {0}, {0,1}, {0,2}, {0,1,2}

Edges:
- m:3.7 ⊗:0.2 cp(m)=2.6 lower:1
- ú:9.1 ⊗:0.2 cp(ú)=0.02 loweraccent:1
- a:7.1 ⊗:0.9 cp(a)=8.1 lower:1
- j:4.8 ⊗:0.4 cp(i)=7.3 lower:1
- u:8.2 ⊗:0.9 cp(u)=2.8 lower:1
- š:6.5 ⊗:0.4 cp(š)=0.01 loweraccent:1
- 1:1.2 ⊗:3.5 cp(1)=0.1 number:1
- s:0.2 ⊗:6.2 cp(s)=6.3 lower:1
- .:1.3 ⊗:1.2 cp(.)=1 punct:1
- e:3.8 ⊗:3.5 cp(e)=12.2 lower:1
- n:4.2 ⊗:6.2 cp(n)=7.0 lower:1
- ,:1.4 ⊗:1.2 cp(,)=0.9 punct:1
- i:0.1 ⊗:5.3 cp(i)=7.3 lower:1
- .:1.3 ⊗:1.2 cp(.)=1 punct:1
- é:4.7 ⊗:5.3 cp(é)=0.02 loweraccent:1
- ,:1.4 ⊗:1.2 cp(,)=0.9 punct:1
- s:0.2 ⊗:6.2 cp(s)=6.3 lower:1
- n:4.2 ⊗:6.2 cp(n)=7.0 lower:1
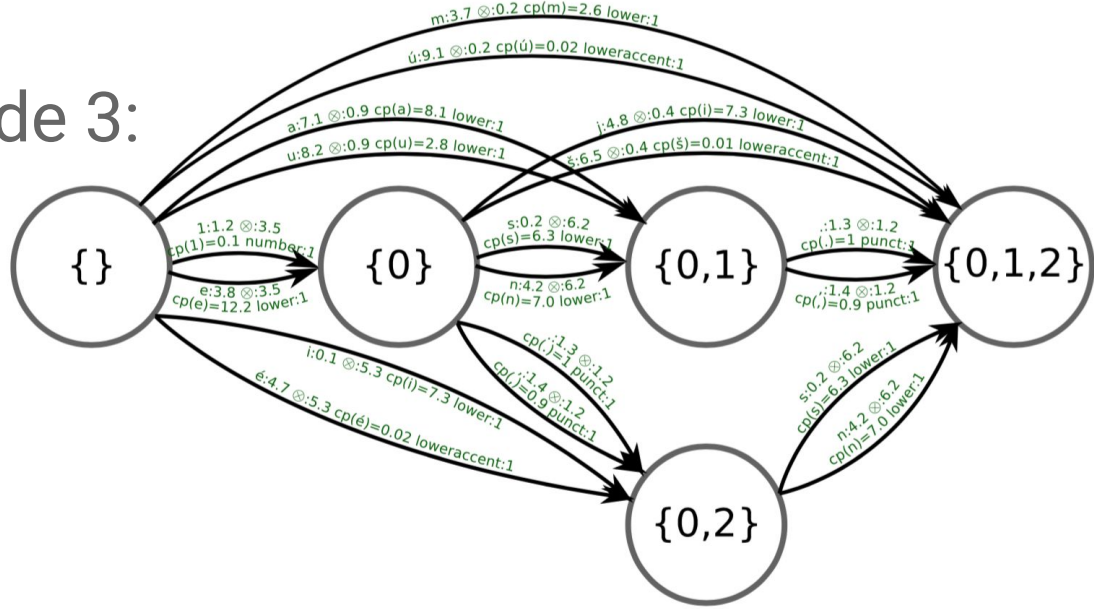
Beam 0 steps:

label	edge cost	lm cost	total cost	path
	0.0	0.0	0.0	{}

Beam 1 steps:

label	edge cost	lm cost	total cost	path
1	-1.2	9.7	8.5	{}-0
i	3.1	7.3	10.4	{}-02
é	0.4	13.5	13.9	{}-02
m	7.1	13.4	20.5	{}-012
u	11.1	9.9	21.0	{}-01
e	13.5	7.9	21.4	{}-0
ú	9.9	12.0	21.9	{}-012
a	15.3	7.0	22.3	{}-01

Beam 2 steps:

label	edge cost	lm cost	total cost	path
is	1.3	8.3	19.9	{}-02-012
1.	2.1	9.4	20.0	{}-0-02
m			20.5	{}-012
1,	2.1	10.5	21.1	{}-0-012
1s	1.3	13.5	23.3	{}-0-01
in	6.0	7.9	24.2	{}-02-012
1n	6.0	14.0	28.5	{}-0-01
és	1.3	17.7	32.9	{}-02-012
1j	12.7	14.0	35.2	{}-0-02
1š	7.1	24.0	39.6	{}-0-02
én	6.0	19.9	39.8	{}-02-012

Beam 3 steps:

| label	edge cost	lm cost	total cost	path
is			19.9	{}-02-012
m			20.5	{}-012
1.s	1.3	36.3	37.6	{}-0-02-012
1,s	1.3	41.2	42.5	{}-0-02-012
1.n	6.0	39.2	45.2	{}-0-02-012
1,n	6.0	40.1	46.1	{}-0-02-012

Google

Feature Function Weights

Label **"i"**
Feature functions values:
0.1 – character score
0.9 – language model score
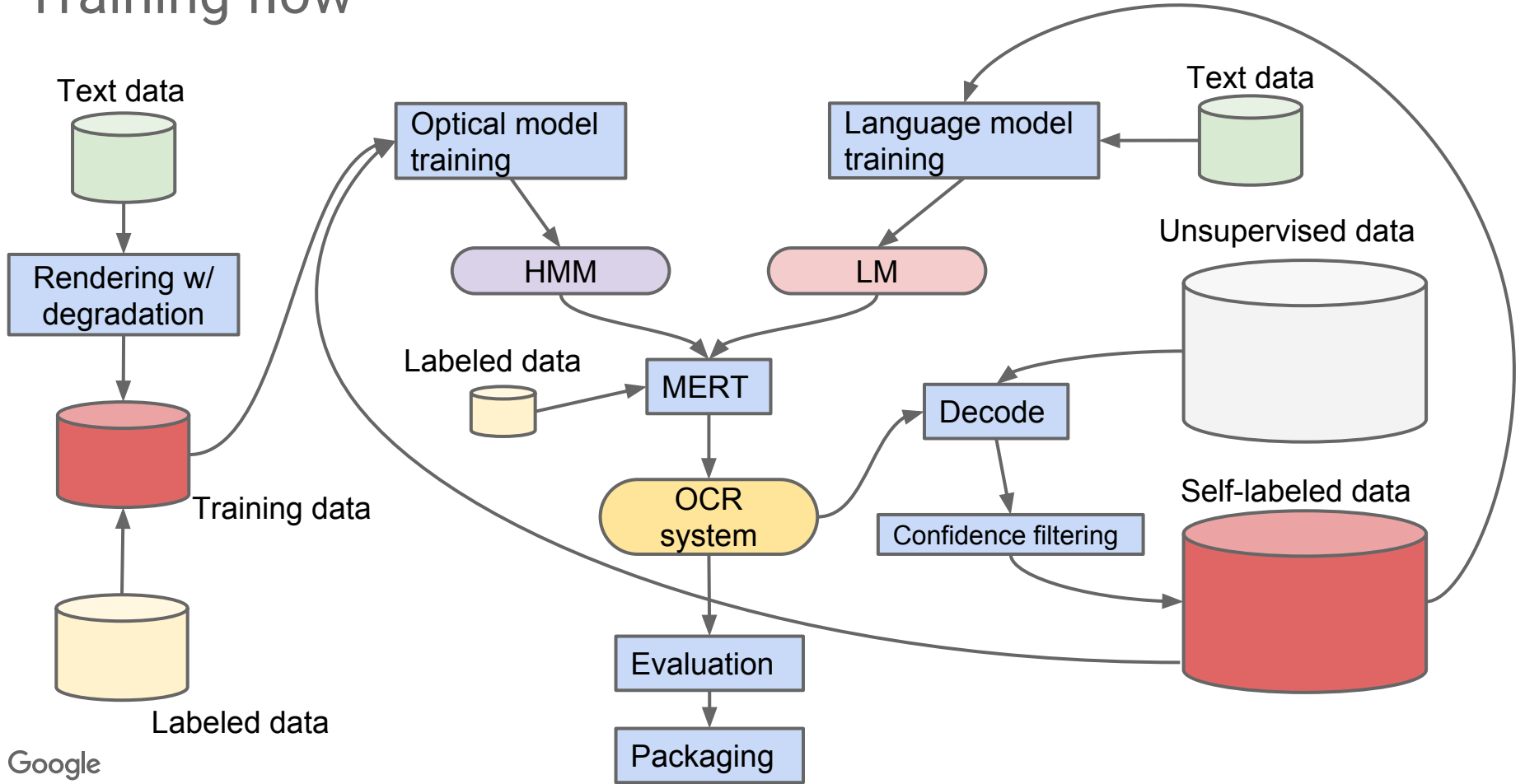2.3 – relative size to neighbors
0.2 – cut score

Label **"é"**
[...]

determine edge score as weighted sum

$$c = \sum_f \gamma_f c_f$$

{}   {0}   {0,1}   {0,1,2}   {0,2}
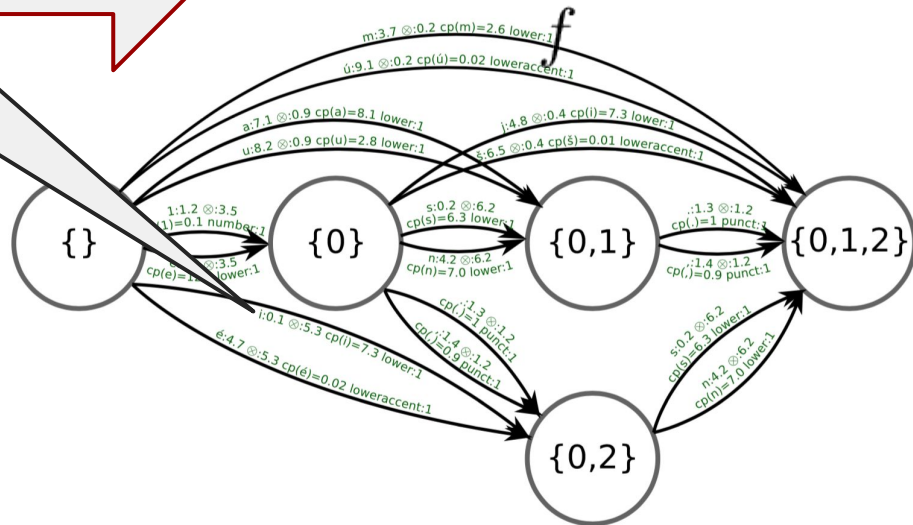
Features: Per character hypothesis

- Histograms of point features (3210 dimensional)

- Bitmap features: 3x8x8 pixels (192 dimensional)

- Simple statistics (384 dimensional)

- Water reservoir features (64 dimensional)

- Stroke direction (180 dimensional)

- Quantized stroke direction maps (512 dimensional)

More feature functions

- string length
- character prior
- segmenter cut features
- relative size

Part 2: Research at Google

Google's Hybrid Approach to Research

Spector, Norvig, Petrov '12 *Comm. of the ACM*

- Pattern 2: Small research team builds a system that gets deployed.

  "This pattern applies best when continuing research can further improve and extend the resulting products."

Enablers

- Single code base, wide range of library functions
- Infrastructure
- Expertise and skills of other teams
- Data

Ashok Popat, Jan 23, 2016

Enablers (cultural)

- Transparency and cooperation
- Peer review
- Respect and psychological safety
- Team- and personal-level pace and execution
- Data-centrism

Software engineering

- Respected and valued
- If it's not checked in, it doesn't exist
- Toy prototypes versus production-quality code
- A day in the life: 80/20

Ashok Popat, Jan 23, 2016

Google

Part 3: Observations and Reflections

Improvement over time (May - July): Arabic

Ashok Popat, Jan 23, 2016
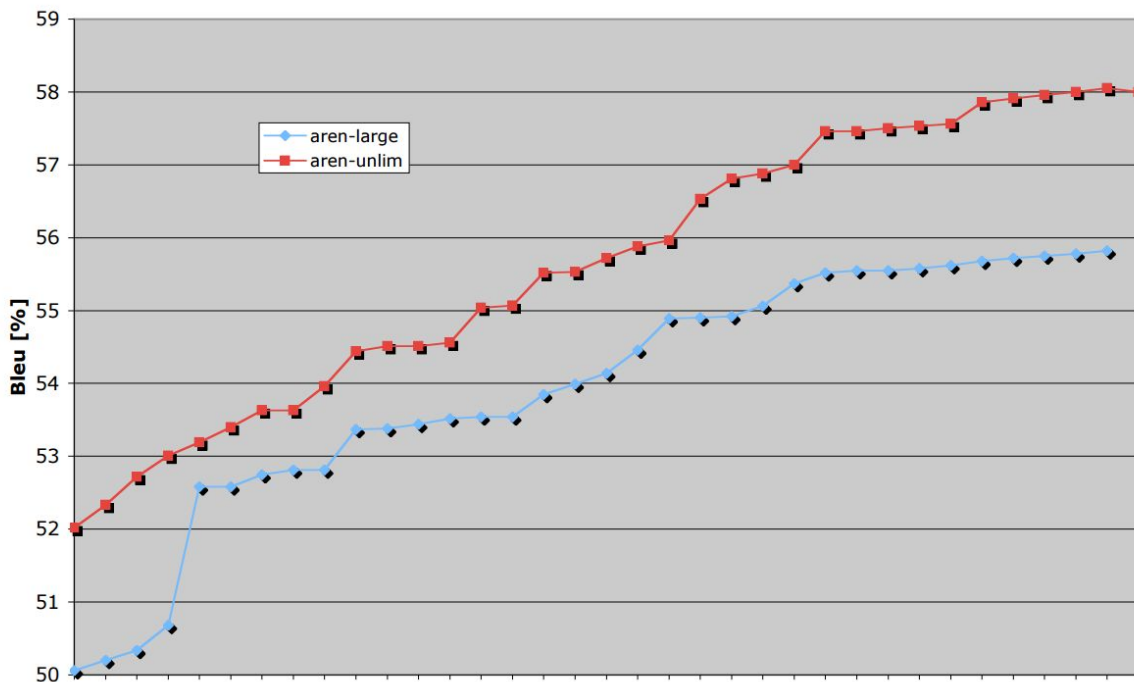
Rapid real progress

- Multiple contributors, one system
- Industry folks at NIST'06 meeting were startled
- Incentive: get a real gain, check it in quickly
- From each according to ability
- Data is important; eval data is paramount

Ashok Popat, Jan 23, 2016

Keeping it real

- Working, deployed system that solves a whole problem
- Tight feedback loop
- Everything that matters gets measured

Ashok Popat, Jan 23, 2016

Pedestrian approaches versus cutting edge

- Translate: world-beating and obsolete
- Data versus Syntax
- Language modeling: "Stupid Backoff" (Brants et al., 2007)
- When and how to invest in promising researchy approaches?

Ashok Popat, Jan 23, 2016

Reward and recognition

- Cleverness, independence, origination of new ideas?
- Cooperation, generosity, communication, productivity, risk taking?
- Imposter syndrome
- Happiness

Google

Ashok Popat, Jan 23, 2016

Summary: what's worked for me?

- Work on real systems
- Measure what matters
- Incent the right things
- Keep aware of new research while investing conservatively

Ashok Popat, Jan 23, 2016

Then and now

Ashok Popat, Jan 23, 2016

Thank you!