You bungling nincompoops!
What's happened to our snibbers?
You idiots couldn't organise your way out of a paper bag!
etc. etc. . . .

Messieurs,
Nous avons l'hommeur de vous informer que les snibloeurs ne sont pas arrivés...

Bridging the Language Barrier with Intelligent Systems
A paraphrase

The canonical paraphrase

- *The glass is half-empty.*
- *The glass is half-full.*
Inverses of relationships

• The company *acquired* four daily newspapers from Worrell Enterprises.

• Worrell Enterprises *sold* four daily newspapers to the company.

• *Only 30%* of the students *passed* the exam.

• *Most* of the students *failed* the exam.
Syntactic variations

Variation in focused component

- At least 13 people were killed by a suicide bomber on a bus in downtown Jerusalem this morning.

- A suicide bomber blew himself up on a bus in downtown Jerusalem this morning, killing at least 13 people.
Lexical variation

Translation as paraphrase (Barzilay and McKeown 2001)

- Emma pleurait, et il s’efforçait de la consoler, enjolivant de calembours ses protestations. (Gustave Flaubert, Madame Bovary, 1857)
- Emma cried, and he tried to console her, adorning his words with puns.
- Emma burst into tears and he tried to comfort her, saying things to make her smile.
More paraphrases

The canonical paraphrase revisited

- The glass is half-empty.
- The glass is half-full.
- The glass is bigger than it needs to be.
Interpretation and viewpoint

- The U.S.-led *invasion* of Iraq …
- The U.S.-led *liberation* of Iraq …
- The U.S.-led *occupation* of Iraq …
- “You’re getting in the way.”
- “I’m only trying to help.”
Definition of paraphrase

- To paraphrase a text is to “talk about the same situation in a different way”.
- Different words or different syntax.
What’s “the same situation”?

- Same truth conditions? **Sometimes**
- Same details? **Sometimes**
- Same message? **Rarely**
Paraphrases aren’t synonymous

- Pragmatic differences in paraphrases:
  - Viewpoint, evaluation.
  - Fine-grained denotation.
  - Style, connotation, implication.
Viewpoint and evaluation

Large-scale paraphrasing
Viewpoint

- Viewpoint includes:
  - The **mental model** used to interpret or frame the situation.
  - The relative **emphasis** placed on the components of the model — especially, the component taken as **central**.
What we don’t have

What we have
• At least 13 people were killed by a suicide bomber on a bus in downtown Jerusalem this morning.

• A suicide bomber blew himself up on a bus in downtown Jerusalem this morning, killing at least 13 people.
Evaluation

(Hunston and Thompson 2000)

• Evaluation is speaker’s **attitude, stance, opinion, appraisal.**

• An implicit (or explicit) **judgement** of the goodness or badness of some aspect of the situation.
Bombardier Inc said that it can make a thinner-walled subway car, capable of cramming in an extra 35 people. (Macleans, 19 May 2003)
Evaluation and ideology

(Malrieu 1999)

- Evaluation must be based on system of values.
- Part of the writer’s mental model or its enclosing ideologies.
In a dramatic escalation of its assault on music piracy, the recording industry plans to start suing individuals who illegally swap copyrighted songs over the Internet. Recording Industry Association of America officials said that beginning today, they will start preparing for the lawsuits by gathering information on people who make songs available for sharing over the Internet.

(Austin American-Statesman, 26 June 2003)
Example (2b)

The RIAA (Recording Industry Association of America) has **issued the biggest threat** to date against online **file-traders**, saying it will **sue thousands of individuals into submission**.

Starting Thursday, **pigopolist grunts** will begin combing P2P networks in search of **industrious file traders**.  *(The Register, 25 June 2003)*
Frameworks of interpretation (1)

(Lakoff 1987)

- Categorization, interpretation within model or framework is basis of cognition.
- Depends on culture and individual.
- Important topic (under various different names) in social psychology, sociology, cognitive science, political science, law, ....
Frameworks of interpretation (2)

- Paraphrase by change of framework.
- Basis of much literature, poetry, and humor.
- Cause of many political, religious, and inter-personal disputes.
- Mediator: Person who interprets and paraphrases across frameworks.
• **Pauline** (Hovy 1988):
  - NLG system (not analysis).
  - Could describe a situation in many different ways.
  - Choice of content and words depended on viewpoint and conversational goals.
Viewpoint and paraphrase in CL

(Paris 1993; Hirst, DiMarco, Hovy, Parsons 1997; Reiter et al 1999)

- Tailoring text to the individual reader (with known user model).
  - Emphasize details of interest to reader.
  - Maximize persuasiveness by matching reader’s viewpoint and value-system.
- Tutoring, health communication.
Viewpoint and paraphrase in CL

(Malrieu 1999)

- Evaluative semantics:
  - Check a text for “ideological consistency”.
  - Compare the ideological or evaluative aspects of different paraphrases.
Viewpoint and paraphrase in CL

- Related work in knowledge representation:
  - Faceted representations.
  - Resolution of conflicting beliefs.
  - Reasoning in and about context.
  - Structure mapping.
Summary

• Conceptual restatement is a ubiquitous and pervasive kind of paraphrase.

• Viewpoint, emphasis, and evaluation may change in this kind of paraphrase.

• Truth conditions may be different but should be compatible.
Lexical nuances of meaning

Small-scale paraphrasing
Lexical choice

- Paraphrase by replacing words with synonyms?

The President addressed the nation.
Language dislikes synonymy

- Absolute synonymy is rare, useless.
- “Every two forms contrast in meaning.”
  (Clark 1992; cf Saussure 1916)
- Language evolves to avoid synonymy, differentiate linguistic forms by nuances.
- **Nuance**: Small difference.
Lexical nuances

- Most “synonyms” are near-synonyms.
- Differences:
  - Denotation: blunder, slip.
  - Emphasis: enemy, foe.
  - Connotation: fib, lie.
  - Register: intoxicated, drunk, plastered
  - Evaluation: skinny, slender.
How do we know?

- Hard for language users to articulate distinctions between near-synonyms.
- Can’t learn distinctions from corpora.
- Distinctions are explained in dictionaries of “synonyms”.
- Lexicographers as skilled users.
Webster's New Dictionary of Synonyms

A Merriam-Webster

Thousands of synonyms defined, discriminated and illustrated with quotations. Plus antonyms, analogous words and contrasting words. To help you use the right word in the right place.
erratic eccentric, odd, queer, *strange, singular, peculiar, unique, quaint, outlandish, curious
Auran aberrant, *abnormal, atypical, irregular, unnatural, anomalous capricious, bizarre, mercurial, eccentric
Con normal, *regular, typical, natural, *usual, customary, wonted, habitual, common, ordinary, familiar; conventional, formal (see CEREMONIAL) decorrect, decent, seemingly proper
error mistake, blunder, slip, lapse, faux pas, bull, holler, boner are comparable when they denote something (as an act, statement, or belief) that involves a departure from what is, or what is generally held to be, true, right, or proper. Error implies a straying from a proper course and suggests such guilt as may lie in failure to take proper advantage of a guide (as a record or manuscript, a rule or set of rules, or a principle, law, or code); thus, a typographical error results when a compositor misreads a manuscript; an error in addition involves some failure to follow the rules for addition; an error in conduct is an infraction of an accepted code of manners or morals (those who, with sincerity and generosity, fight, and in an evil cause, posterity can only compassionately as victims of a generous but fatal error—Scott) (without understanding grievous and irreparable errors can be made—Donald Harrington) MISTAKE implies misconception, misunderstanding, a wrong but not always blameworthy judgment, or inadvertence; it expresses less severe criticism than error (he made a serious mistake when he chose the law as his profession) (a child makes many mistakes in spelling) (there is a medium between truth and falsehood, and I believe the word mistake expresses it exactly. I will therefore say that you were mistaken—Cowper) Blunder is harsher than mistake or error; it commonly implies ignorance or stupidity, sometimes blameworthiness (we usually call out our blunders mistakes, and our friends style our mistakes blunders—Wheatley) One’s translation is sure to be full of gross blunders, but the supreme blunder is that of translating at all when one is trying to catch not a fact but a feeling—Henry Adams) Slip carries a stronger implication of inadvertence or accident than mistake and often, in addition, connotes triviality (the wrong date on the check was a slip of the pen) (a slip which makes us feel hot all over—L. P. Smith) Often, especially when it implies a transgression against morality, the word is used euphemistically or ironically (let Christian’s slips before he came hither... be a warning to those that come after—Bunyan) (the minister... comes when people are in extremis, but they don’t send for him every time they make a slight moral slip—tell a lie, for instance, or smuggle a silk dress through the customs—Holmes) Lapse, though sometimes used interchangeably with slip, stresses forgetfulness, weakness, or inattention more than accident; thus, one says a lapse of memory or a slip of the pen, but not vice versa (writes well, despite occasional lapses into polysyllabic humor—Geographical Jour.) When used in reference to a moral transgression, an asterisk (*) indicates place of treatment of each guilt it carries a weaker implication of triviality than slip at a stronger one of a fall from grace or from one’s standards (for all his... lapses, there was in him still nobility, an even ascetic firmness and purity of character—Ellis) Faux pas is most frequently applied to a mistake or an etiquette (she was carefully instructed so that there was no danger of her making a faux pas when she was presented at the Court of St. James’s) (Goha and I, horrified, hustled him out before he could commit further faux pas—S. H. Adams) Bull, holler, and boner are all three rather informal terms applicable to blunders, especially to blunders in speech or writing. The language is typically an amusing aspect. A bull may be a grotesque blunder in language typically characterized by some risible incongruity (the well-known bull nag that “one man is just as good as another—and something more so”) or it may be a mere stupid or gauche blunder (he really committed a bull when he solemnly introduced his new friend to the latter’s ex-wife) A holler is any oracular or ludicrous error based on ignorance or confusion of ideas; the term is used especially of laughable errors in scholastic recitations or examinations (a collection of schoolboy hollers) A holler that turns the tide “Imitations of Immortality” into “Imitations of Immortality”. A boner may be a grammatical, logical, or factual blunder in a piece of writing that is usually so excessive to be funny (a few historical boners... such as dunam surviving until medieval times—Coulton Wood). It may be a ridiculous or embarrassing slip of the kind the results from a sudden lapse (as of attention or facility or decorum) (the proprietor of a large and vast selection of diplomatic boners—Rosenthal)
errorless flawless, faultless, *impeccable
Ana correct, accurate, exact, precise, right, nice
erussat nit “artificial, synthetic, factitious,
erudite learned, scholarly
erudition learning, scholarship, *knowledge, *science, information, lore
escape vb 1. Escape, flee, fly, decamp, abscond means run away especially from something which limits or freedom or threatens one’s well-being. To escape means the idea of flight from confinement or restraint that very often conveys no suggestion of wrongdoing or danger (one of the most powerful motives that make people to science and art is the longing to escape everyday life—Ellis) (eager to escape from the air and go back to his hometown—Wester) The trip haste and often abruptness in departure (there was evidence that the burglars had been frightened and fled. If it often connotes disappearance, especially extended to things (the mist fled before the rising sun) Fly is interchangeable with flee but its use is restricted in idiomatic English to the present tense (go, father, for all your friends are fled—Shak.) Decamp usually suggests a sudden departure to elude discovery or arrest; it commonly carries a disparaging or belittling connotation charging impertinent to my companion. I found it high time for us to decamp—Smollett) (to town, took orders, received advances of gold and money, and then decamped—Jones) Abscond with decamp the distinctive implications of clandestine subterfuge and concealment usually to avoid the consequences of fraudulent action (he had the appearance of a bankrupt tradesman absconding—Meredith, determined to be a poet at any price, he absconded to college with his clothes and took refuge in a lonely fast house—Brooke)
Can *follow, chase, pursue, trail, tag
2. Escape, avoid, evade, elude, shun, eschew are comparable
error, mistake, blunder, slip, lapse, faux pas, bull, howler, boner are comparable when they denote something (as an act, statement, or belief) that involves a departure from what is, or what is generally held to be, true, right, or proper. Error implies a straying from a proper course and suggests such guilt as may lie in failure to take proper advantage of a guide (as a record or manuscript, a rule or set of rules, or a principle, law, or code); thus, a typographical error results when a compositor misreads a manuscript; an error in addition involves some failure to follow the rules for addition; an error in conduct is an infraction of an accepted code of manners or morals.
Mistake expresses it exactly. I will therefore say that you were mistaken—Cowper—Blunder is harsher than mistake or error; it commonly implies ignorance or stupidity, sometimes blameworthiness (we usually call our blunders mistakes, and our friends style our mistakes blunders—Wheatley) (one's translation is sure to be full of gross blunders, but the supreme blunder is that of translating at all when one is trying to catch not a fact but a feeling—Henry Adams) Slip carries a stronger implication of inadvertence or accident than mistake and often, in addition, connotes triviality (the wrong date on the
Erreur : Fausse opinion. *Erreur* se dit dans tous les cas où l'on prend le faux pour le vrai (et aussi le mal pour le bien : → Écart). Également, erreur considérable, due à une sorte d'extravagance, de démence, par rapport au vrai et au bien : *Les égarements des sophistes (Acad.). Flot-tant entre la sagesse et l'égarment (J.-J. R.).* Illusion (→ ce mot), erreur des sens ou de

Henri Bénac, *Dictionnaire des synonymes*, 1956
Notes

• Typical entry describes a cluster of near-synonyms.
• Describes common meaning, and then differences — often in relative terms.
• Cluster does not form a hierarchy.
• Differences may relate to arbitrary facets of meaning.
Representing and using nuances

- For computational use, need:
  - **Representation formalism** for knowledge of lexical differences.
  - **Acquisition** of this knowledge.
  - Methods to **use** the knowledge in analysis and generation.
Representing near-synonymy

(Philip Edmonds 1999, Edmonds and Hirst 2002)

- Format of lexical knowledge base follows format of the dictionaries.
- A coarse-grained hierarchy of core meanings of clusters.
- Members of cluster are explicitly differentiated.
(defcluster Error–Cluster
  :syns (error mistake blunder slip lapse)
  :core (ROOT Generic–Error)
  :periph ((P1 Stupidity) (P2 Blameworthiness)
    (P3 Criticism (ATTRIBUTE (P3–1 Severity)))
    (P4 Misconception) (P5 Accident) (P6 Inattention))
  :distinctions
  ((blunder usually medium implication P1)
    (mistake sometimes medium implication (P2 (DEGREE 'medium)))
    (blunder sometimes medium implication (P2 (DEGREE 'high)))
    (mistake always medium implication (P3–1 (DEGREE 'low)))
    (error always medium implication (P3–1 (DEGREE 'medium)))
    (blunder always medium implication (P3–1 (DEGREE 'high)))
    (mistake always medium implication P4)
    (slip always medium implication P5)
    (mistake always low implication P5)
    (lapse always low implication P5)
    (lapse always medium implication P6)
    (blunder always medium pejorative)
    (blunder high concreteness)
    (error low concreteness)
    (mistake low concreteness)))
Acquiring the lexical knowledge


- Knowledge acquisition from a dictionary of near-synonyms as an information extraction task.
Choose the Right Word

A CONTEMPORARY GUIDE
TO SELECTING
THE PRECISE WORD
FOR EVERY SITUATION

S.I. HAYAKAWA
EUGENE EHRLICH, REVISING EDITOR

"To the careful writer and speaker, this book can be invaluable." — Saturday Review

**abjure.** Do not confuse the verb *abjure* (renounce under oath) with the verb *adjure* (urge solemnly).

**abrogate.** Do not confuse the verb *abrogate* (cancel or repeal) with the verb *arrogate* (claim a power, privilege, etc., unduly).

**absorb**
- assimilate
- digest
- imbibie
- incorporate
- ingest

These verbs, all relatively formal, indicate the taking in of one thing by another. *Absorb* is slightly more informal than the others and has, perhaps, the widest range of uses. In its most restricted sense it suggests the taking in or soaking up specifically of liquids: the liquid *absorbed* by the sponge. In more general uses *absorb* may imply the thoroughness of the action: not merely to read the chapter, but to *absorb* its meaning. Or it may stress the complete disappearance of the thing taken in within the encompassing medium: once-lovely countryside soon *absorbed* by urban sprawl. *Ingest* refers literally to the action of taking into the mouth, as food or drugs, for later absorption by the body. Figuratively, it designates any taking in and suggests the receptivity necessary for such a process: too tired to *ingest* even one more idea from the complicated philosophical essay she was reading. *To digest* is to alter food chemically in the digestive tract so that it can be *absorbed* into the bloodstream. In other uses, *digest* is like *absorb* in stressing thoroughness, but is even more emphatic. [You may completely *absorb* a stirring play in one evening, but you will be months *digesting* it.]

**Assimilate** is even more emphatic about the thoroughness of the taking in than either *absorb* or *digest*—in both its specific physiological and general uses. Physiologically, food is first *digested*, then *absorbed* by the bloodstream, and then *assimilated* bit by bit in each cell the blood passes. In more general uses, *assimilate*, unlike the previous verbs, often implies a third agent beside the absorber and the absorbed—an agent that directs this process: architects who *assimilate* their buildings to the environment. The process, furthermore, often implies the complete transformation of the absorbed into the absorbing medium. *Assimilate* also suggests a much slower process than *digest* and certainly than *absorb*, which can be nearly instantaneous: It would take the city generations to *assimilate* the newcomers into the patterns of a strange life.

**Incorporate** is the only verb here that does not have a specific use pertaining to the taking in of liquids or of food, meaning literally embody. It resembles the aspect of *assimilate* that stresses the loss of separate identity for the absorbed quantity: incorporating your proposals into a new system that will satisfy everyone. It is unlike *assimilate* in lacking that verb's suggestion of necessarily careful, time-consuming thoroughness.

**Imbibie**, while capable of uses comparable to those for *assimilate*, is mainly rooted still to its specific use for the taking in of liquids. Even this use, and certainly any others, sound slightly archaic and excessively formal: Do you *imbibe* alcoholic beverages? See *EAT*.

**Antonyms:** disgorge, disperse, dissipate, eject, emit, exude.

**abstain**
- forbear
- refrain

The verb *abstain* means withhold oneself from an action or self-indulgence. [There were six votes in favor, two against, and two *abstaining*. She *abstained* from drinking.] *Refrain* has to do with withholding an action temporarily, or checking a momentary desire. He *refrained* from scolding his child until the company left. To *forbear*, in its intransitive sense, is to exercise self-control, often out of motives of patience or charity. [Though impatient, the customer *forbore* to upbraid the harried sales clerk; The teacher *forbore* to report Johnnie's misbehavior to his parents.] See *FORGO*, *FORSWEAR*.

**Antonyms:** begin, permit.
absorb
assimilate
digest
imbibe
incorporate
ingest

These verbs, all relatively formal, indicate the taking in of one thing by another. **Absorb** is slightly more informal than the others and has, perhaps, the widest range of uses. In its most restricted sense it suggests the taking in or soaking up specifically of liquids: the liquid *absorbed* by the sponge. In more general uses *absorb* may imply the thoroughness of the action: not merely to read the chapter, but to *absorb* its meaning. Or it may stress the complete disappearance of the substance absorbed within the
Acquiring the lexical knowledge


- Knowledge acquisition from a dictionary of near-synonyms as an information extraction task.
- Disambiguate (using WordNet senses).
  - Many clusters refer to just one sense of an ambiguous word.
Refining the lexical knowledge


- Add more denotational distinctions. (from the Macquarie Dictionary and WordNet)
- Add knowledge of collocations and anti-collocations for each word. (from Web-based counts).
Using near-synonyms

(Diana Inkpen 2003, Inkpen and Hirst 2003)

- Xenon: NLG system built on HALogen
  \((n\text{-gram–based generation, Langkilde 2000})\).

- Uses knowledge of near-synonyms to choose the best word to express the desired nuances.

- Overrides \(n\)-gram probabilities where necessary.
Syntactic nuances of meaning

(Chrysanne DiMarco 1990, DiMarco and Hirst 1993)

- Syntactic structure affects style of text.
  - Same structure might give different styles in different languages.

- Rules to relate syntactic structure to:
  - Staticness vs dynamism.
  - Abstraction vs concreteness.
  - Clarity vs obscurity.

- Paraphrase to maintain style in translation.
Translation as lexical paraphrase

(Vinay and Darbelnet 1995; Manfred Stede 1996)

- Paraphrase to get most synonymous possible translation in context.

Emma *pleurait* → Emma *burst into tears*

Tom *disconnected* the wire from the sparkplug.
Tom *zog* das Zündkabel von der Zündkerze *ab.*

- Situation → Result of action → Explication
- Action (‘pull off’)
Conclusion

“That’s not what I said”
Paraphrasing Paraphrased

- Viewpoints on paraphrase:
  - As change of viewpoint or evaluation.
  - As change of truth conditions.
  - As change of words and syntax.
  - As change of connotation, style, other pragmatic aspects of the message.
Some things I didn’t get to say

- Analysis of paraphrase:
  - … as a linguistic phenomenon.
  - … as a literary phenomenon.
  - … as a social and psychological phenomenon.
- “Semantic prosody”
Some applications I didn’t mention

- Higher-quality machine translation.
- Finer-grained analysis of content and opinion in text.
- Finer-grained NLG.
- Intelligent thesauri and writers’ aids.
- Summaries, event tracking, QA, …
Research challenges (1)

For each aspect of paraphrase:
- Representation of knowledge.
- Acquisition of knowledge.
Recognition challenges (2)

- Recognizing paraphrases.
- *Imposing* recognition of paraphrase.
- Measuring paraphrase distance.
- Explicating paraphrase differences.
- Distinguishing paraphrase from misunderstanding.
Research challenges (3)

- Generating paraphrases on demand.
  - Linguistic restatement.
  - Conceptual restatement.
- Bringing it all together:
  The automatic dispute mediator.
- Peace in the Middle East.
Acknowledgements

• Content was improved by Diana Inkpen, Nadia Talent, Kazuko Nakajima, and Yuki Johnson.

• Research and preparation of the talk was funded by the Natural Sciences and Engineering Research Council of Canada.

• Travel was funded by the Nara Institute of Science and Technology. Many thanks to Kentaro Inui and Ulf Hermjakob.
Contact information

- gh@cs.toronto.edu
- www.cs.toronto.edu/compling
Technical references (1)

Technical references (2)

Technical references (3)


Technical references (4)

Technical references (5)

- Saussure, Ferdinand de. Cours de linguistique générale. 1916.
Dictionaries cited