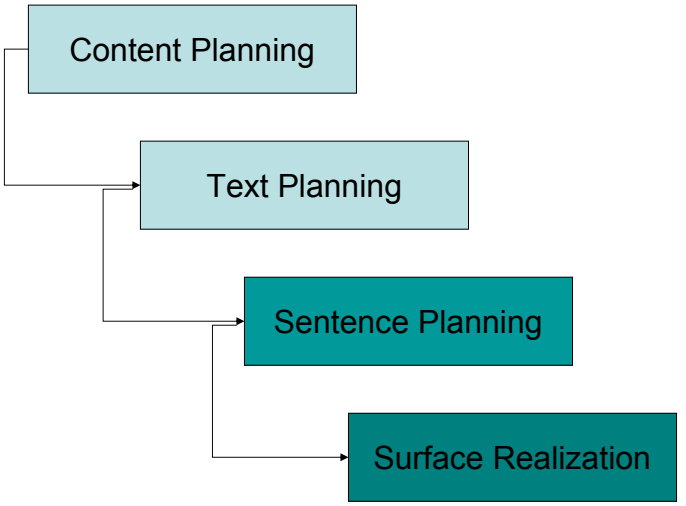
Sentence Planning e Generazione Superficiale

Natural Language Generation



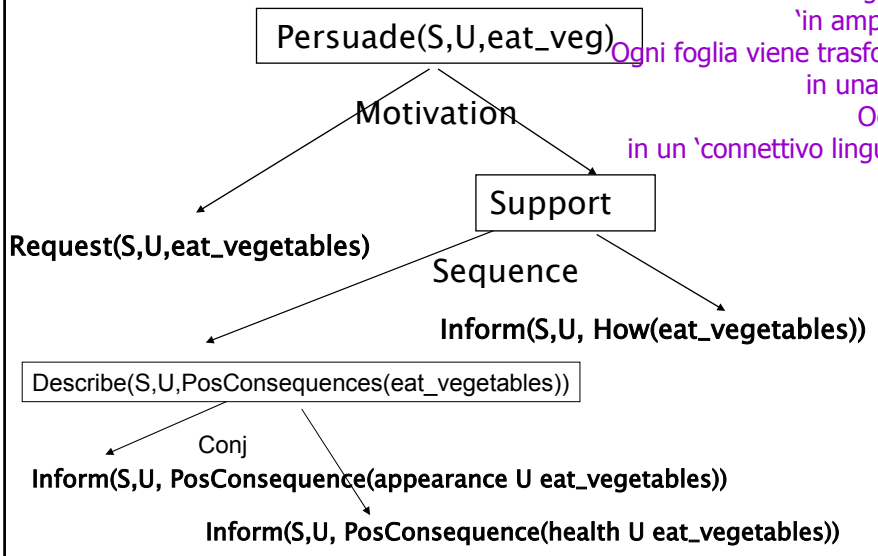
Testo da Generare (dopo l'analisi del corpus)

You should eat 5 portions a day of vegetables and fruit.

This is good for your appearance and health.

Il Piano

Il piano viene esplorato dal Generatore Superficiale con un algoritmo 'in ampiezza'. Ogni foglia viene trasformata in una frase. Ogni RR in un 'connettivo linguistico'



Piano da “Realizzare”

Request(S,U,eat_vegetables)
-> You should eat vegetables and fruit.

Inform(S,U, PosConsequence(appearance U eat_vegetables))
->Eating vegetables and fruit is good for your appearance.

Inform(S,U, PosConsequence(health U eat_vegetables))
->Eating vegetables and fruit is good for your health.

Inform(S,U, How(eat_vegetables))
->You should eat 5 portion a day of vegetables and fruit.

Sentence Planning
Migliorare la struttura del testo.

- Se facciamo il mapping uno-a-uno dalle foglie del piano in frasi -> testo ridondante e poco “fluente”.
- E’ necessario combinare le foglie in maniera appropriata per proddure testi migliori.
- Il risultato di questa fase è il SENTENCE PLAN in cui vengono specificate le frasi.

Aggregazione: un esempio di sentence planning

Se usiamo una frase per ogni fatto da dire, otteniamo testi innaturali:

`likes(john, mary)`	*John likes Mary.*
`likes(jim, mary)`	*Jim likes Mary.*

Combinare due fatti in una singola frase migliora la situazione, ma non sempre il testo diventa più scorrevole:
John likes Mary and Jim likes Mary.

Aggregazione è il processo di combinare fatti e contemporaneamente eliminare informazione ridondante:
John and Jim like Mary.

Semplici regole di aggregazione

Aggreghiamo ogni volta che due parti della frase sono uguali:

* **Soggetto:**
  John likes Mary +
  John hates Jim. =
  John likes Mary and hates Jim.
* **Verbo+Oggetto:**
  John likes Mary +
  Jim likes Mary =
  John and Jim like Mary
* **Soggetto+Verbo:**
  John likes Mary +
  John likes Sue. =
  John likes Mary and Sue.
* **Verbo+Oggetto Diretto+Oggetto Indiretto**
  John gives Mary a book +
  Jim gives Mary a book. =
  Jim and John give Mary books

Usiamo insomma regole del tipo:
ab + ac + .. an = a(b+c+..n)
an + bn + ...mn = (a+b+..m)n

Torniamo al nostro piano

- Senza Aggregazione:

  You should eat vegetables and fruit. **Eating vegetables and fruit is good for your appearance. Eating vegetables and fruit is good for your health.** You should eat 5 portion a day.

- Con l'aggregazione:

  You should eat vegetables and fruit. **Eating vegetables and fruit is good for your appearance and health.** You should eat 5 portion a day.

Torniamo al nostro piano

- Posso ancora aggregare?:

  You should eat vegetables and fruit. Eating vegetables and fruit is good for your appearance and health. **You should eat 5 portion a day.**

Con una aggregazione non proprio immediata:

You should eat 5 portion a day of vegetables and fruit. Eating vegetables and fruit is good for your appearance and health.

Lessicalizzazione

- Una volta determinato il contenuto testuale e la strutturazione delle informazioni in paragrafi e frasi

dobbiamo

- "lessicalizzare": la lessicalizzzione determina le parole da usare per esprimere i contenuti e le relazioni

? Good O Positive

Generazione Superficiale: Realizzazione Sintattica e Morfologica.

- Ogni linguaggio naturale è governato da regole grammaticali che formalizzano il modo in cui vengono costruite le frasi.
- **Morfologia**: regole per la formazione delle parole
- **Sintassi**: regole per la formazione delle frasi

Regole Moforlogiche

Esempi in inglese:

- Per formare il passato di un verbo aggiungi ed
  - walk + ed = walked
- Per formare il plurale di un nome aggiungi s
  - `train + s = trains`
- Se la radice di una parola finisce in e ed il suffisso inzia con una vocale, cancella la e finale nella radice
  - `like + ed = liked`

Realizzazione Sintattica

Esempi di Regole in Inglese:

- Il soggetto va prima del verbo
- Il soggetto e il verbo devono concordare in numero (sing. plur.)

Realizzazione Ortografica

- Inseriemento di punteggiatura
- Veste tipografica: font size, column width …

Esempi di regole:

- Le frasi devono iniziare con le lettere maiuscole e finire con il “.”.
- Ecc.

Generazione Superficiale: Come Realizzare Ogni Atto Comunicativo

a = (‘eat vegetables and fruit’)

Request(S, U, a):

Utilizziamo uno stile di ‘suggerimento’ (la mamma)

"You should eat vegetables and fruit"

Inform(S, U, How(a))

Immaginiamo che, nella ‘base di dati sul dominio’, l’attributo ‘How’ dell’azione ‘eat_vegetables’ sia rappresentata indicando il numero di porzioni giornaliere(5)

"5 portions a day"

Inform(S, U, PosConsequence(c a)):

Immaginiamo che, nella ‘base di dati sul dominio’, uno degli attributi ‘PosConsequence’ dell’azione a sia rappresentata da c

"good for your appearance"

Come Rendere le Singole Frasi
(Generazione Superficiale)

- con un 'testo precompilato':
  a *Request (S, U, eat_vegetables)* viene associata la frase:
  "You should eat vegetables and fruit"
- *con una 'cornice linguistica'*
  a *Request (S, U, (eat <object>))* viene associata la cornice:
  "You should eat <object>"
- *con un 'generatore di linguaggio naturale'*
  che conosca le regole della grammatica e della sintassi

Esempio

Inform(S,U, PosConsequence(c U action))

<action> has a positive consequence on your <c>
In questo caso:
Eating vegetables has a positive consequence on your appearance!

Oppure:

<action> is good for <c>
In questo caso:
Eating vegetables is good for your appearance!

Le Parti del Discorso Non Sono Separate Fra Loro!

Quali relazioni?

*You should eat fruit and vegetables **because** it is good for you health and appearance.*

Motivation

*it is good for you health **and** appearance.*

conjunction

Relazioni fra le parti del discorso

Secondo la *'RST'*,
le Parti del Discorso Sono Legate fra Loro da 'Relazioni Retoriche', che vengono espresse da 'Connettivi Linguistici'

Esempi:

Joint: ',' 'e', ….
Purpose: 'allo scopo di', 'per',…
Motivation: 'perché',…
Elaboration Object-Attribute: 'il quale',…
Solutionhood: 'Per risolvere questo problema',…
Ordinal Sequence: 'Per prima cosa', 'per seconda cosa',...

Esempio di Sistema: STOP

- Scopo:
  - Produrre foglietti informativi personalizzati riguardanti il fumo (stop-smoking leaflets).
- Input:
  - Questionario sulle abitudini, credenze, ecc del fumatore;
- Target user:
  - NHS (Ministero della Sanita' britannico)
- Sviluppo:
  - Aberdeen University (CS, Medicine, GP Depts)
  - Sito http://www.csd.abdn.ac.uk/research/stop/onlineQ.htm
- Stato:
  - Sviluppato ma non viene usato nella pratica clinica.

STOP: Input

SMOKING QUESTIONNAIRE

Please answer by marking the most appropriate box for each question like this: ☒

Q1 Have you smoked a cigarette in the last week, even a puff?

YES ☒ NO ☐

Please complete the following questions

Please return the questionnaire unanswered in the envelope provided. Thank you.

Please read the questions carefully. If you are not sure how to answer, just give the best answer you can.

Q2 **Home situation**:

Live alone ☐ Live with husband/wife/partner ☒ Live with other adults ☐ Live with children ☒

Q3 **Number of children** under 16 living at home …… boys ……1……. girls

Q4 **Does anyone else in your household smoke?** *(If so, please mark all boxes which apply)*

husband/wife/partner ☒ other family member ☒ others ☐

Q5 **How long have you smoked for?** …10… years

Tick here if you have smoked for less than a year ☐

STOP: Output

Dear Ms Cameron

Thank you for taking the trouble to return the smoking questionnaire that we sent you. It appears from your answers that although you're not planning to stop smoking in the near future, you would like to stop if it was easy. You think it would be difficult to stop because *smoking helps you cope with stress, it is something to do when you are bored, and smoking stops you putting on weight*. However, you have reasons to be confident of success if you did try to stop, and there are ways of coping with the difficulties.

Output: Testo senza Formattazione

When time is limited, travel by limousine, unless cost is also limited, in which case go by train. When only cost is limited a bicycle should be used for journeys of less than 10 kilometers, and a bus for longer journeys. Taxis are recommended when there are no constraints on time or cost, unless the distance to be travelled exceeds 10 kilometers. For journeys longer than 10 kilometers, when time and cost are not important, journeys should be made by hire car.

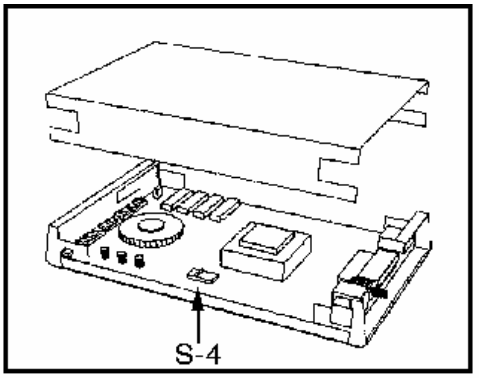
Output:Testo con formattazione

When only time is limited:
travel by Limousine
When only cost is limited:
travel by Bus if journey more than 10 kilometers
travel by Bicycle if journey less than 10 kilometers
When both time and cost are limited:
travel by Train
When time and cost are not limited:
travel by Hire Car if journey more than 10 kilometers
travel by Taxi if journey less than 10 kilometers

Output: Testo

* Push the code switch S-4 to the right. The code switch is located in front of the transformer.

Testo e Immagini



Push the code switch S-4 to the right.

Output: Ipertesti (PagineWeb)

- La Generazione dinamica di ipertesti puo' essere vista come una forma limitata di dialogo fra un sistema e l'utente.
- Il click dell'utente su un link rappresenta una forma di *follow-up question*.
- Decidere quale informazione inserire nella pagina e quale rendere accessibile attraverso un link
- Decidere come "etichettare" un link
- Esempi: ILEX , PEBA-II
  - Costruiti prima di xml, dei cookies e dei linguaggi di scripting per il web.

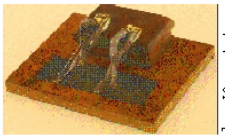
Esempio: ILEX

- Scopo:
  - Descrizioni *context-sensitive* di oggetti presenti in un museo virtuale
- Input:
  - DB del museo + storia dell'interazione
- Target user:
  - National Museums of Scotland
- Sviluppo:
  - Edinburgh University
  - Sito http://www.hcrc.ed.ac.uk/ilex/systemintro.html
- Stato:
  - Funziona, possibile sviluppo commerciale

Esempio: ILEX

ILEX=Intelligent Labelling EXplorer

- Usa metodi e tecniche di NLG per generare descrizioni di oggetti esposti in una galleria di un museo (National Museums of Scotland's 20th Century Jewellery Gallery).
- **Interazione**:
  - i visitatori inizialmente hanno a disposizione un indice costituito dalle immagini di tutti gli oggetti presenti nel museo.
  - Selezionando una immagine si ottiene la descrizione dell'oggetto. Questa viene generata considerando quali pezzi della galleria il visitatore ha gia' visto.
  - Le descrizioni hanno lo scopo di "educare" il visitatore. Questo si realizza cercando di rispettare i GOAL del sistema che il generatore cerca di soddisfare ogni volta che si presenta l'opportunita'.



Pair of brooches on mount

Silver, gold, mahogany, walnut and perspex

This item was made in 1979 and is made of silver, gold, mahogany, walnut and perspex. It was designed by Martin Page who was English. Like the necklace designed by Flockinger, this item is in the Organic style. Organic jewels tend to be coarsely textured. However, this item has smooth surfaces.

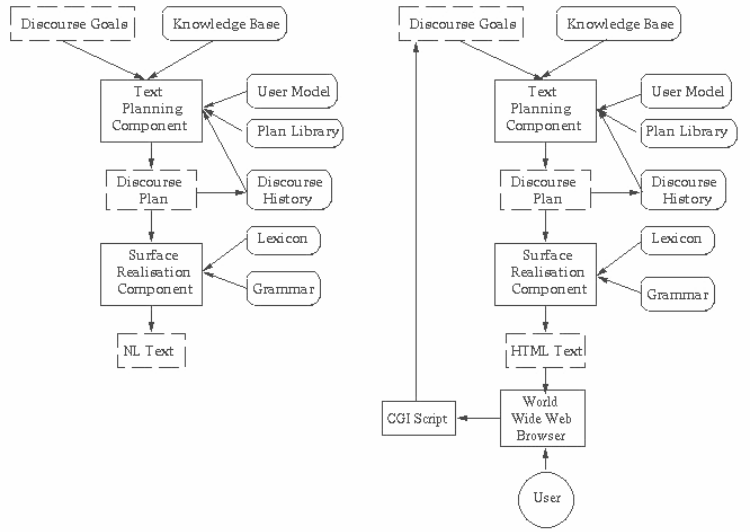
With a piece like this, the boundary between `jewellery' and `sculpture' or `art' starts to become quite indistinct. One important theme across 20th Century jewellery has been what to do with a piece of jewellery when it is not being worn. From the 1970s onwards, jewellers have started exploring the idea of turning jewellery into sculpture--so that you can hang it on the wall, or prop it on the mantelpiece when you are not wearing it. This piece works equally well whether it is being worn or being displayed (as at present).

Other jewels in the organic style include:

- a pendant necklace designed by Bjorn Weckstrom
- the necklace designed by Flockinger
- a bracelet designed by Flockinger
- a finger ring designed by Frances Beck
- a finger ring designed by Jacqueline Mina
- the previous item
- a finger ring designed by Ernest Blyth

Figure 3: A description produced by ILEX-1.2.

Come cambia l'architettura



Peba-II

- Peba-II è una enciclopedia on-line che produce descrizioni e confronti di animali.
- Genera descrizioni relative a centinaia di animali ma si focalizza soprattutto su:
  - Australian Monotremes (Echidna e Platypus) e Marsupiali (Kangaroo, Koala e Wombat).
  - Ci sono anche alcuni felini e altri mammiferi (Alligatori, Porcospini, Elefanti, Conigli,ecc).
- Le descrizioni sono prodotte tenendo in considerazione il livello di esperienza dell'utente e quello che ha gia' visto nel sistema.

The Echidna

The Echidna, also known as the spiny Anteater, is a type of Monotreme that is covered in stiff, sharp spines mixed with long, coarse hairs.

The Echidna has the following subtypes:

- the short-beaked Echidna and
- the long-beaked Echidna.

The Echidna is about the same length as a domestic cat. It ranges from 2 kg to 7 kg in weight. It has a browny black coat and paler-coloured spines. It has a small head. It has a prolonged, slender snout. It has no teeth. It uses its extensible, sticky tongue catching ants, termites and other small insects. It is a carnivore and eats ants, termites and earthworms. It has powerful claws allowing for rapid digging of hard ground. It is found in Australia. It is active at dawn and dusk. It lives by itself. It has an avera lifespan in captivity of 50 years.

This text is generated for the novice user level. If you would like the text for the expert user level click here.

Figure 2: A description of the echidna produced by PEBA-II.

Generazione di Descrizioni Comparative

Importante non solo per l'apprendimento di concetti ma anche per il commercio elettronico.

Tipi di Confronti: Milosavljevic e Dale descrivono 3 tipi di confronti trovati nel loro corpora (Enciclopedia dei Mammiferi):

• *Direct comparison* (*bi-focal*e): le entita' che vengono descritte sono alla pari.

  • Ad esempio: descrizione di due sottotipi di una classe: *There are two kinds of camels: the dromedary, or Arabian camel, which has one hump, and the Bactrian camel, which has two humps*.

  • Oppure quando le due entita' sono molto simili (i.e. coniglio e lepre).

PEBA-II genera confronti diretti se l'utente lo richiede e confronta "punto-punto" le proprieta' delle due entita' fornendo prima le analogie e poi le differenze.

Generazione di Descrizioni Comparative

The Rabbit and the Hare

The Rabbit is a member of the Leporidae Family that has very helpless young which are born naked and with closed eyes. The Hare is a member of the Leporidae Family that has young which are born furred and with open eyes.

Some comparisons:

Like the Hare, the Rabbit has **long ears**.
Like the Hare, the Rabbit **feeds on herbs, tree bark and vegetables**.
Like the Hare, the Rabbit has a short, upturned tail.
The Hare **is longer than** the Rabbit.
The Hare **weighs twice as much as** the Rabbit.
The Rabbit lives in underground burrows whereas the Hare lives in a simple nest.
The Rabbit lives in colonies whereas the Hare rarely lives socially.

Esempio

The Echidna

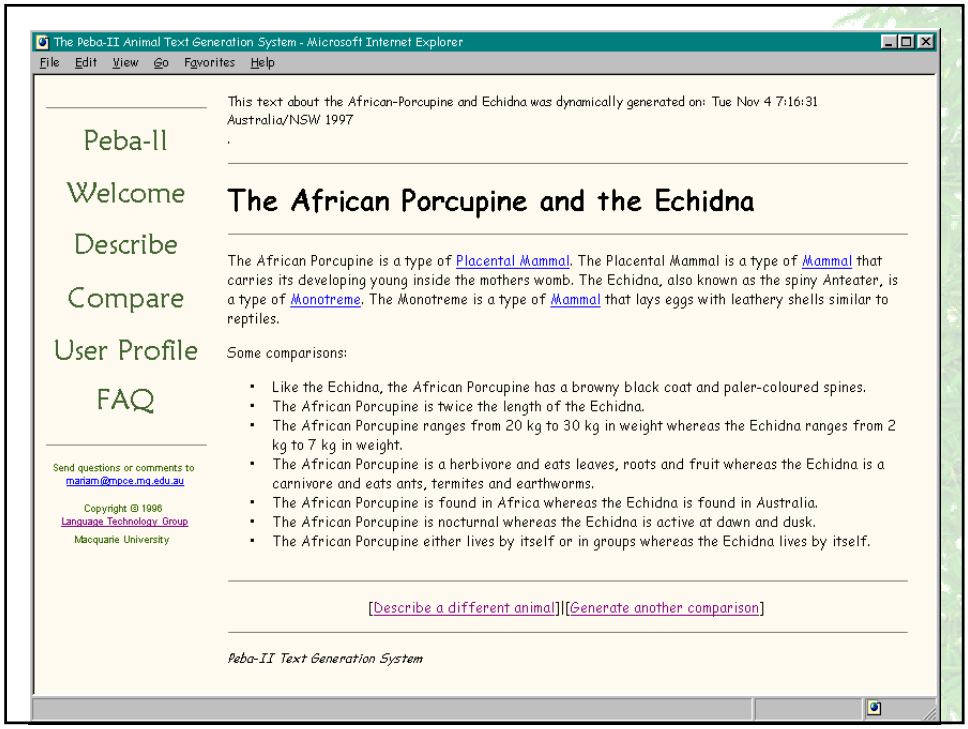
The Echidna, also known as the spiny Anteater, is a type of Monotreme that is covered in stiff, sharp spines mixed with long, coarse hairs. Although it is similar in appearance to the African Porcupine it is not closely related. The African Porcupine is a type of Rodent that has long sharp spines, up to 50cm long, which cover its whole back and can be raised by muscles under the skin. Like the African Porcupine, the Echidna has a browny black coat and paler-coloured spines. The African Porcupine is twice the length of the Echidna (80.0 cm vs 47.5 cm). The Echidna has an average weight of 4.5 kg whereas the African Porcupine has an average weight of 25.0 kg. The Echidna is a carnivore and eats ants, termites and earthworms whereas the African Porcupine is a herbivore and eats leaves, roots and fruit.

…

Generazione di Descrizioni Comparative

- *Illustrative comparison*: utile se esiste una entita' comunemente nota (o nota in maniera specifica dall'utente) che condivide particolari proprieta' con l'entita' descritta.

Esempio:
The Echidna is about the same length as a domestic cat.



Conclusioni

- NLG è un area di ricerca interessante e piena di sviluppi commerciali e di interesse per le applicazioni "reali".
- NLG ormai comprende non solo il linguaggio naturale ma anche altre forme di espressione
- Ora esistono tecnologie che rendono piu' semplice la generazione.