

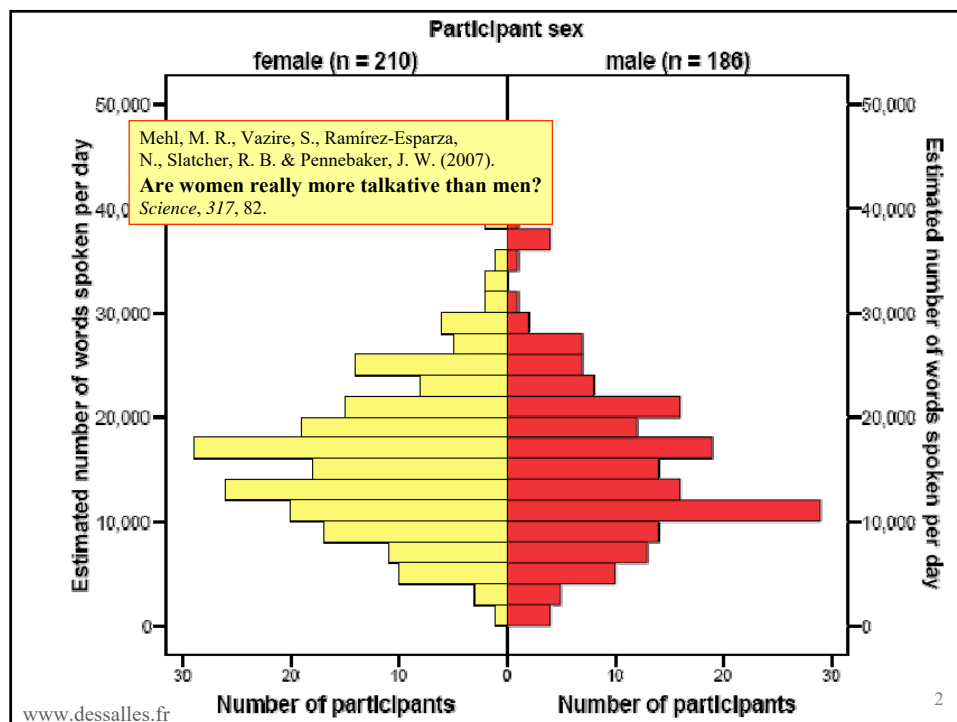
Symbolic Artificial intelligence

SD213

## Language competence

Jean-Louis Dessalles

[www.dessalles.fr](http://www.dessalles.fr)



**Table 1.** Estimated number of words spoken per day for female and male study participants across six samples.  $N = 396$ . Year refers to the year when the data collection started; duration refers to the approximate number of days participants wore the EAR; the weighted average weighs the respective sample group mean by the sample size of the group.

Sample	Year	Location	Duration	Age range (years)	Sample size ( $N$ )		Estimated average number (SD) of words spoken per day	
					Women	Men	Women	Men
1	2004	USA	7 days	18–29	56	56	18,443 (7460)	16,576 (7871)
2	2003	USA	4 days	17–23	42	37	14,297 (6441)	14,060 (9065)
3	2003	Mexico	4 days	17–25	31	20	14,704 (6215)	15,022 (7864)
4	2001	USA	2 days	17–22	47	49	16,177 (7520)	16,569 (9108)
5	2001	USA	10 days	18–26	7	4	15,761 (8985)	24,051 (10,211)
6	1998	USA	4 days	17–23	27	20	16,496 (7914)	12,867 (8343)
Weighted average							16,215 (7301)	15,669 (8633)

Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B. & Pennebaker, J. W. (2007). Are women really more talkative than men?. *Science*, 317, 82.

[Experience sampling] studies have pinpointed students' amount of daily conversations at around 6 hr—about **38%** of the day (Reis & Wheeler, 1991)—whereas the EAR data suggest **27.9%**.

(Mehl & Pennebaker (2003)

(Dunbar 1998)

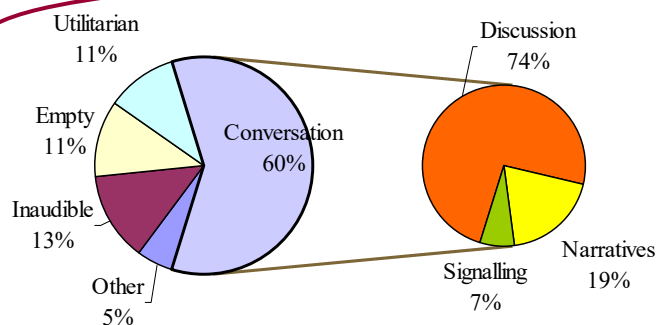
**TABLE 1.1** *Time spent in language use in different cultures*

Society	Economy	Activity	% of waking hours
Dundee (UK)	industrial	conversation	20.6
Kapanara (PNG)	horticultural	social interaction	19.4
Maasai (Tanzania)	pastoralist	leisure	17.5
Central African Republic	agricultural	non-work (leisure, dances, visits)	16.8
Nepal	agricultural	leisure/social	32.3
Ivory Coast	agricultural	social	7.2
Upper Volta	agricultural	free time (social, religion, errands)	23.6

Conversation type	%
Storytelling	43.4
Observation/Comment	19.75
Opinion	16.8
Gossip	13.8
Joke-telling	6.3

*Distribution of conversational topic types in Eggins and Slade's corpus (1997 p. 265)*

Distribution of utterance types in our own corpus, assessed through a sampling method



www.dessalles.fr

5

Fabian Suchanek

## Natural Language ... appears everywhere on the Internet

WIKIPEDIA

English: The Free Encyclopedia, 3 468 000 articles

Deutsch: Die freie Enzyklopädie, 1 100 000 Artikel

France: L'encyclopédie libre, 880 000 articles

Italiano: L'enciclopedia libera, 700 000 voci

Polaki: Wolna encyklopedia, 100 000 artykułów

Nederlands: De vrije encyclopedie, 600 000 artikelen

Español: La enciclopedia libre, 940 000 artículos

Pyccов: Свободная энциклопедия, 800 000 статей

Português: A enciclopédia livre, 610 000 artigos

The Universal Declaration of Human Rights

On December 10, 1948 the General Assembly of the United Nations adopted and proclaimed the Universal Declaration of Human Rights the full text of which appears in the following pages. Following this historic act the Assembly called upon all Member countries to publicize the text of the Declaration and "to cause it to be disseminated, displayed, read and expounded principally in schools and other educational institutions, without distinction based on the political status of countries or territories."

Mise à jour 16:33

**LE FIGARO • fr** ACTUALITÉ ÉCONOMIE

INFO

- Politique
- International
- Environnement
- Santé
- Auto
- Société
- Médias
- Science et Tech
- Web
- Météo

DÉ

(1 trillion Web sites;  
1 trillion =  $10^{12}$   
≈ number of cells in the human body)

Relational Transducers for Electronic Commerce

Serge Abiteboul\*  
I.N.R.I.A.-Rocquencourt  
Serge.Abiteboul@inria.fr

Victor Vianu\*  
U.C. San Diego  
vianu@cs.ucsd.edu

Brad Fordham  
Oracle Corporation  
bfordham@us.oracle.com

www.dessalles.fr

6

## Natural Language

**Inbox:** 3726 unread messages

(250 billion mails/day;  
 $\approx$  number of stars in our galaxy;  
 80% spam)



**me:** Would you like to have dinner with me tonight?  
**Cindy:** no.

(1 billion chat msg/day on Facebook;  
 1 billion =  $10^9$  = distance Chicago-Tokio in cm)

## Natural Language: Tasks

- Automatic text summarization

Let me first say how proud I am to be the president  
 of this country. Yet, in the past years, our country...  
 [1 hour speech follows]

→ Summary: Taxes will increase by 20%.

- Machine translation  
 librairie → book store

- Information Extraction  
 Elvis Presley lives on the moon.  
 → lives(ElvisPresley, moon)

## Natural Language: Tasks

- Natural language understanding  
Close the file! Clean up the kitchen!
- Natural language generation  
Dear user, I have cleaned up the kitchen for you.
- Text Correction  
My hardly loved mother-in law  
→ My heartily loved mother-in-law
- Question answering  
Where is Elvis?  
→ On the moon

## Views on Natural Language

Elvis will be on concert tomorrow in Paris!

For humans:



For a machine: 45 6C 76 69 73 20 77 69 6C...

# Linguistics

**Linguistics** is the study of language.

Linguistics studies language just like biology studies life



www.dessalles.fr

11 11

## Languages



Mandarin (850m)  
Spanish (330m)  
English (330m)  
Hindi (240m)  
Arabic (200m)  
Bengali (180m)  
Portuguese (180m)  
Russian (140m)  
French (120m)  
Japanese (120m)  
Punjabi (100m)  
German (100m)  
Javanese (90m)  
Shanghainese (90m)

- around 6000 languages
- around 20 language families
- European languages are mostly Indo-European

www.dessalles.fr    Counts depend a lot on the definition and may vary.

12 12

**Kivunjo (langue bantou) : Năikîm' lyîîă (Pinker, 1994)**  
**“ He is eating it for her ”**

- N-: A marker indicating that the word is the "focus" of that point in the conversation.
- -ă-: A subject agreement marker. It identifies the eater as falling into Class 1 of the sixteen classes, "human singular." Other classes: several humans, thin or extended objects, objects that come in pairs or dusters, the pairs or clusters themselves, instruments, animals, body parts, diminutives (small or cute versions of things), abstract qualities, precise locations, and general locales.
- -î-: Present tense. Other tenses in Bantu can refer to today, earlier today, yesterday, no earlier than yesterday, yesterday or earlier, in the remote past, habitually, ongoing, consecutively, hypothetically, in the future, at an indeterminate time, not yet, and sometimes.
- -kî-: Object agreement marker, in this case indicating that the thing eaten falls into gender Class 7.
- -m-: A benefactive marker, indicating for whose benefit the action is taking place, in this case a member of gender Class 1.
- -lyî-: The verb, "to eat."
- -î-: An "applicative" marker, indicating that the verb's cast of players has been augmented by one additional role, in this case the benefactive. (As an analogy, imagine that in English we had to add a suffix to the verb *bake* when it is used in *I baked her a cake* as opposed to the usual *I baked a cake*.)
- -ă : A final vowel, which can indicate indicative versus subjunctive mood.

www.dessalles.fr

13

## La complexité du langage

**Kivunjo (langue bantou) : Năikîm' lyîîă**  
**“ He eats it for her ”**

- N- *marqueur de focus*
- ă- *accord de classe avec le sujet. Ici: humain singulier*
- î- *temps présent*
- kî- *accord avec l'objet, classe animaux (classe 7)*
- m' - *marqueur indiquant que le bénéficiaire de l'action est de la classe 1*
- lyî- *verbe manger*
- î- *modification du verbe pour indiquer la présence d'un rôle de bénéficiaire*
- ă *mode indicatif (vs subjonctif)*

(16 classes : humain singulier, humain pluriel, objets fins ou étendus, objets en paquets ou par paires, paires ou paquets, instruments, animaux, parties du corps, diminutifs, qualités abstraites, localisations précises, localisations vagues).

(temps : aujourd'hui, plus tôt aujourd'hui, hier, pas plus tôt qu'hier, hier ou avant, dans le passé lointain, habituellement, en train, consécutivement, hypothétique, futur, temps indéterminé, pas encore, parfois).

www.dessalles.fr

14

chinois : yì-zhì-qiānbǐ  
*un-objet(en forme de bâton)-crayon*

Chinese has class markers

## Language complexity

Donaudampfschiffahrtskapitaen  
= Donau + Dampf + Schiff + Fahrt + Kapitaen  
[captain of a steam ship ride on the Danube]

Rechtschutzversicherungsgesellschaften

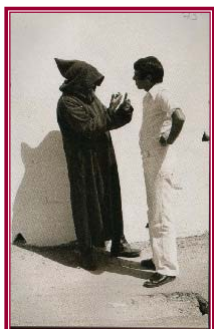
Donaudampfschiffahrtselectrizitätenhauptbetriebswerkbauunterbeamtengesellschaft

Der Vater des Schülers schämt sich.  
*Le père de l'écolier a honte.*

Der Vater schämt sich des Schülers.  
*Le père a honte de l'écolier.*



## Language processing levels



phonology  
 morphology (derivation)  
 morphology (inflection)  
 syntax  
 semantics  
 relevance  
 sociology  
 psychology

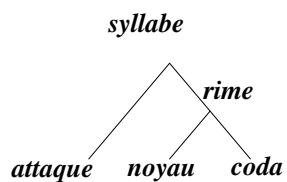
These levels correspond to cognitive modules

## Language processing levels

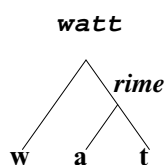
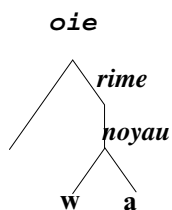
<i>Level</i>	<i>counter-example</i>	<i>example</i>
Phonology	Bning	Ulbrage
Morphology	Coolth	Warmth
Syntax	House money pay	The door garden eats the sky
Semantics	The apple reads a boy	My sister bought a bike
Relevance	Two years ago, I lost 20 cts in the street.	I've lost my wallet in the metro.

These levels correspond to cognitive modules

## Phonological competence



Syllable structure

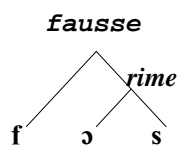
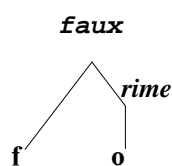


www.dessalles.fr

19

## Phonological competence

*Syllabe ouverte → voyelle fermée*  
*Syllabe fermée → voyelle ouverte*



*aveugle*  
*aveuglement*  
*aveuglement*

*entier*  
*entière*

*sot*  
*sotte*

*prestidigitateur*  
*préstimulation*

*rastole*  
*rastolin*

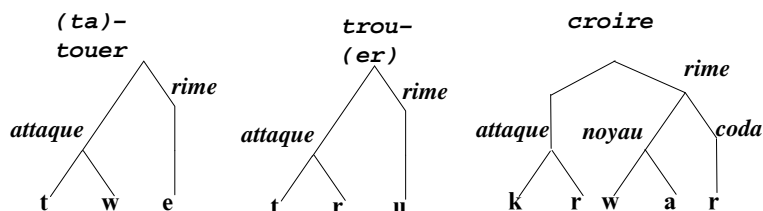
French Southern accent is characterized by a (simple) rule!

(Durand 1990, p. 223)

20

www.dessalles.fr

## Phonological competence



louer	pierre	trouer	prouesse	croire
rieur	lueur	influer	prière	trois
ruelle	mouette	plier	brouette	pluie
	fier		bleuet	groin

La présence d'une liquide dans l'attaque bloque la  
synérèse (contraction de deux syllabes au prix d'une  
fusion de deux voyelles en une semi-voyelle) (Kaye & Lowenstamm 1984, p. 146)

www.dessalles.fr

## Competence in morphology

<i>mot</i>	<i>Signification</i>	<i>âge</i>
enrailler	contraire de dérailler	3
senter	sentir	3
treindre	tenir	3
déprocher	éloigner	4
serrer	servir	4
écriture	texte écrit	4
explicages	explications	5
habillation	acte de s'habiller	5
se dépyjamer	enlever son pyjama	5
levure	niveau de gonflage d'un matelas	5
poissonneur	poissonnier	5
senture	odeur	5
affaciler	faciliter	6
découdre	découdre	6
détransformer	redonner la forme initiale	6
rallonguer	rallonger	6

"Errors" by young children result  
from the wrong application of rules.

www.dessalles.fr

22

## Competence in morphology

<i>mot</i>	<i>Signification</i>	<i>âge</i>
recouer	recoudre	6
recouser	recoudre	6
se rejoindre	se rejoindre	6
sécherie	sécheresse	6
abritement	abri	7
chantement	acte de chanter	7
décorages	décorations	7
joliesses	choses jolies	7
jupitersien	sur le modèle: martien	7
courantation	courant (rivière)	8
eruptionnaient	avaient une éruption	8
inapercevable	ne peut être aperçu	8
inobligé	non obligatoire	8
proposément	proposition	8
fortifiage	fortification	9
interrompation	interruption	9
menteries	mensonges	9
virager	virer	

“Errors” by young children result from the wrong application of rules.

23

www.dessalles.fr

## Competence in morphology

<i>mot</i>	<i>Flexion</i>	<i>âge</i>
naquis	naissé	4.5
failli	faillu	5-6/6-8.5
buvait	boivait	6.5
vainqueur	vainqueuse	5
senti	sentu	5
découvert	découvrie	5
vaudrait	valerait	7/10
à l'endroit	à l'endroites	5
devait	doivait	5.5
prennent	prendent	5.5/6
buvant	boivant	7.5
offert	offri	5.5
servirait	serrait	5.5
plu	pleuvu	5.5/7.5
noué	noeué	6
ouverte	ouvrie	6-8/3
molles	mous	6
perdra	perdera	

“Errors” by young children result from the wrong application of rules.

24

www.dessalles.fr

## Competence in morphology

<i>mot</i>	<i>Flexion</i>	<i>âge</i>
mettez	mêtes	8/6-7
dorment	dorent	6
veuille	veule	8
s'asseoir	s'assir	8
playing	jouing	6
buvez	boyez	6/3.5
écris	écrié	3.5
lise	lie	6.5
faites	faisez	8.5
fui	fuyé	8.5
près	prête	7
soit	soive	7
surprise	surprenue	7
détruisaient	détruiseraient	9.5
vieillira	vieillissera	7.5
bu	boyé	3.5-4.5
vécu	vis	9.5
tenu	tiendu	

“Errors” by young children result from the wrong application of rules.

25

www.dessalles.fr

## Competence in morphology

<i>mot</i>	<i>Flexion</i>	<i>âge</i>
rejoint	rejointu	4.5
sortira	sortra...sorra	5.5
recousant	recoudant	6
étudiant	étudent	11.5
partira	partera	6.5
séché	séké	6.5
venait	viendait	7
revécu	revis	7
sortirai	sortrai...sorterai	7.5
boirai	buverai	12

“Errors” by young children result from the wrong application of rules.

26

www.dessalles.fr

## Syntactic competence

<i>phrase</i>	<i>âge</i>
Où d'y venait, ce ballon?	6
Il faudra que tu enlèveras l'échelle.	5
Je m'ai réveillé.	5.5
Je suis jamais y allé.	5.5/6-13
Ça fait rien pas.	5.5-6
Mange-les pas!	6
Si tu voudras,	6
J'essaie que ça part	8
C'est pas la peine de te déplacer. Tu peux que te lever.	7
C'est celui que tu m'as fait pris!	8
Arrête de toucher à le mien	6.5
C'est presque du fin bois	6.5
Nous sommes y allés hier	6.5
Bon, je l'ai cherché. [Je suis allé le chercher]	8
C'est pas nous qui le savent!	6.5
Hé, ils s'en aillent!	4
C'est vachement différent que la peinture	8.5
Ben moi, j'aurais préféré d'y aller cette sem	

“Errors” by young children result from the wrong application of rules.

27

www.dessalles.fr

## Syntactic competence

<i>phrase</i>	<i>âge</i>
Pourtant t'étais en retard d'elle	8
T'as vu à quel kilomètre ils sont	8
J'en n'ai pas regardé un seul [j'en ai raté un seul]	8
Non, je préfère que tu es là	8
J'ai cassé tout!	3.5
Je vais lui la rendre.	8.5
Où elle est, la montre à toi?	3.5
Il veut pas embêter moi.	3.5
Je veux me sabiller	3.5
Ça te préfère, comme ça, Quentin	9
C'est pour pas la Corse	3.5
Non. Va pas t'en avec ça !	3.5
Non, on va faire pas un câlin.	3.5
Tu dis à maman de lui la montrer	4
Hé, t'es pas s'en allée, toi	4
Je veux tu le prennes	4.5
Quelqu'un mange pas mes miel pops !	4.5
Je suis pas faillu me faire écraser	

“Errors” by young children result from the wrong application of rules.

28

www.dessalles.fr

## Syntactic competence

### phrase

### âge

Nous on a plein brosses à dents	4.5
Est-ce que je peux te taider ?	4.5
Je veux je suis sur tes épaules.	4.5
[C'est pas encore les vacances] Si, c'est encore les vacances !	4.5
C'est un copain de moi	5
Je voudrais que c'était le bas	5
Pourquoi tu l'as mis pas là ?	5
En va toi	5
Je vais sur le rouge toboggan	5.5
Oui, c'est qu'est-ce que je pense.	5.5
[C'est pas sûr qu'il y aura ce que tu veux]	
Ben si y aura qu'est-ce que je veux !	6.5
Non, y a pas tout le monde qui sont restés	6.5
Qu'est-ce que ça représente qu'est-ce que je tiens là ?	7
J'ai pas trouvé quelque chose d'autre	7.5
T'aurais pas une fine trousse ?	7.5
Pas même les classes sont allumées	7.5
Tu sais c'est quand que...	
Elle est voisine au salon.	

"Errors" by young children result from the wrong application of rules.

29

www.dessalles.fr

## Constraints on structure

- *The fact that he talked about the guy to whom John's brother sold his car, saying he was a crook, did not alter Jack's reputation.*

•He ?



- John
- John's brother
- buyer
- Jack

Coreference is blocked between "he" and "John", "the guy", but not with "Jack"

30

www.dessalles.fr





## Autonomy of syntax

\* Il prétend à ses amis d'être un peintre.  
Le jardin de la porte défait la lampe.

\* She is approaching of the door.  
The garden of the door undoes the lamp.

Non-syntactic sentences can be meaningful,  
and conversely syntactic sentences can be meaningless.

## Compétence syntaxique

Vous imaginez qu'il a parlé à Pierre (1)  
Vous imaginez qu'il a parlé à qui ?  
Imaginez-vous à qui il a parlé ?  
A qui imaginez-vous qu'il a parlé ? (1')  
Vous imaginez ce qu'il a dit à Pierre (2)  
Vous imaginez ce qu'il a dit à qui ?  
A qui a-t-il dit ce que vous imaginez ?  
\* A qui imaginez-vous ce qu'il a dit ? (2')

Movement is blocked in (2')

## Acquisition

### ☀ **Pauvreté du stimulus**

- *corrections: énoncés erronés > impolis > incompréhensibles  
> erreurs de prononciation > erreurs de vocabulaire*

### ☀ **10 mots par jour jusqu'à l'adolescence**

### ☀ **Mécanismes spécifiques**

- Principes et paramètres

Structures and Rules explain how fast learning is possible.

## Acquisition

(Piatelli-Palmarini, M. 1988).

- John ate an apple -- John ate
- John received a gift from his mother - John received a gift
- John wears a hat -- \*John wears
- idem say, tell, mais pas speak, write
- John catches rabbits - John catches
- John washes his hands- John washes
  
- John is too clever for us to catch Fido -- John is too clever for us to catch
- The pool is too cold for us to wash the car -- The pool is too cold for us to wash
  
- Lisa considers John too clever to expect us to catch Bill
- Lisa considers John too clever to expect us to catch

Learning does not proceed by induction from example.  
Learning is guided by structure and rules.