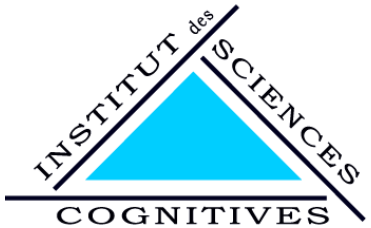


Morphological Decomposition Model (MDM): First Insights and Orthographic Simulation

Ph.D Candidate: Gustavo Lopez ESTIVALET
Thesis Advisor: Fanny MEUNIER

ALP Metting 13th January 2014.

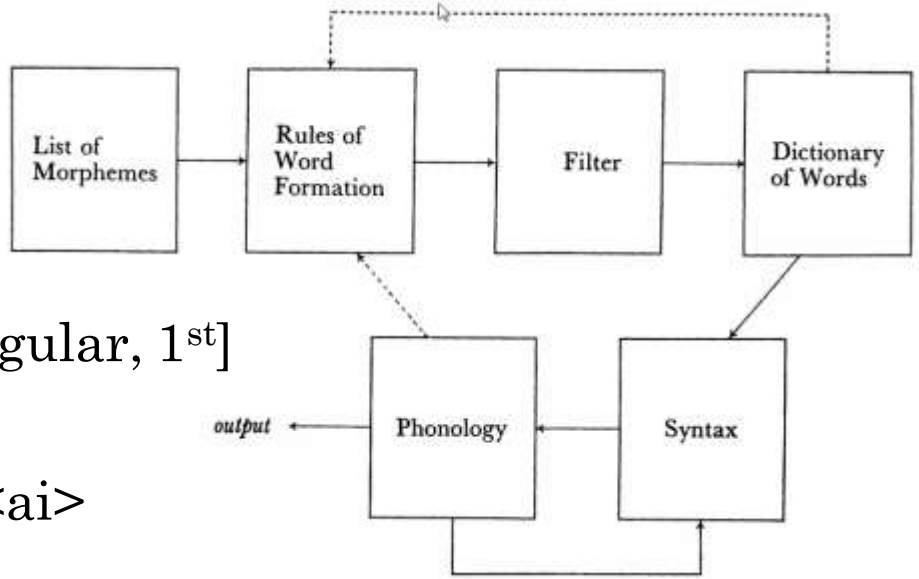
Contextualizing Morphology

parl-er-i-ons, *ons-i-er-parl, *parl-er-ons-i, *i-parl-er-ons

- How is represented the word structure?

  "A grammar must include a list of morphemes as well as rules of word formation or morphology"[1]

- *potential* → filter → *actual*
- phonological rules (SPE)



- **Syncretism:**<s> → [1st], [2nd]

  <ai> → [past], [singular, 1st]

  [past] → <ai>, <i>

  [1st] → <ø>, <s>, <ai>

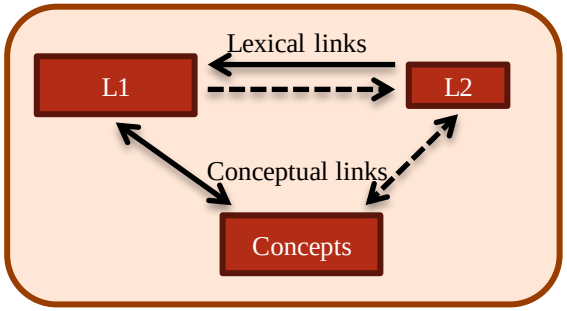
- **Separate phonological form and logical form**

[1]Halle. (1973). Prolegomena to a theory of word formation.

Motivations

"A standard empirical hypothesis is that one component of the mind/brain is a *parser*, which assigns a percept to a signal (abstracting from other circumstances relevant to interpretation). The parser presumably incorporates the language and much else, and the hypothesis is that interpretation involves such a system, embedded in others"[2].

- **BSBM:** How is a bilingual lexicon? Does it have a "language feature"?



Adapted from 3

"UG determines possible symbolic representations and derivations. A language consists of a lexicon and a computational system. The computational system draws from the lexicon to form derivations, presenting items from the lexicon in the format X-bar theory. Each derivation determines a linguistic expression, an Structure Description, which contains a pair (π, λ) meeting the interface conditions"[2].

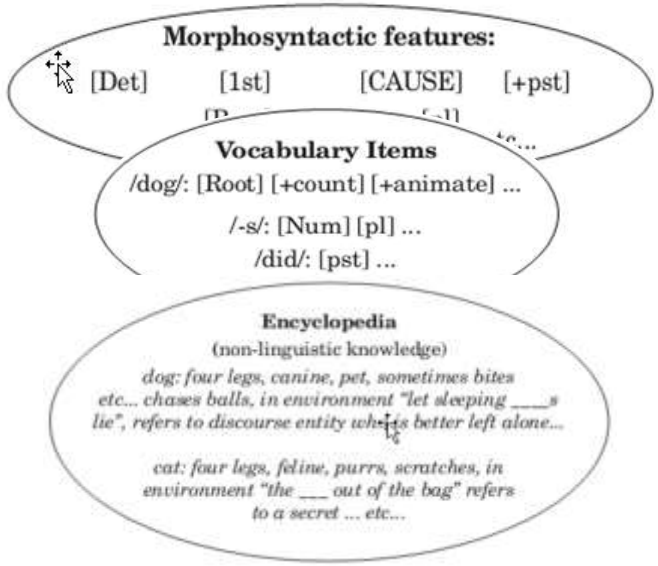
[2]Chomsky. (1995). *The Minimalist Program*. p. 19; p. 171

[3]Kroll. (1993). Accessing conceptual representations for words in a second language.

Distributed Morphology Theory

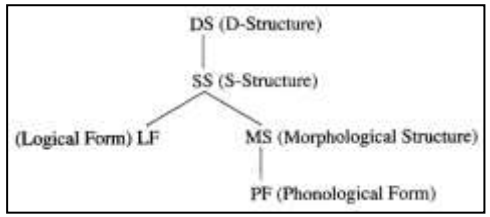
Lexicon[4]

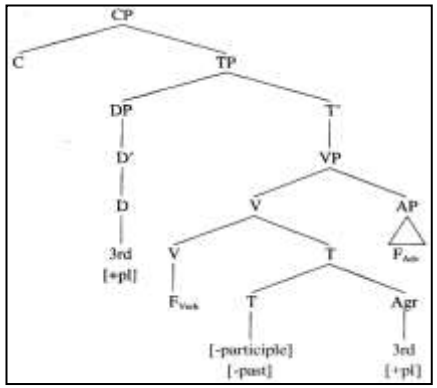
- **List A – Morphosyntactic Features:** inventory of abstract features
- **List B – Vocabulary Items:** relation between abstract features and phonological string
- **List C – Encyclopedia:** relation between Vocabulary Items and general knowledge



Core Properties[5]

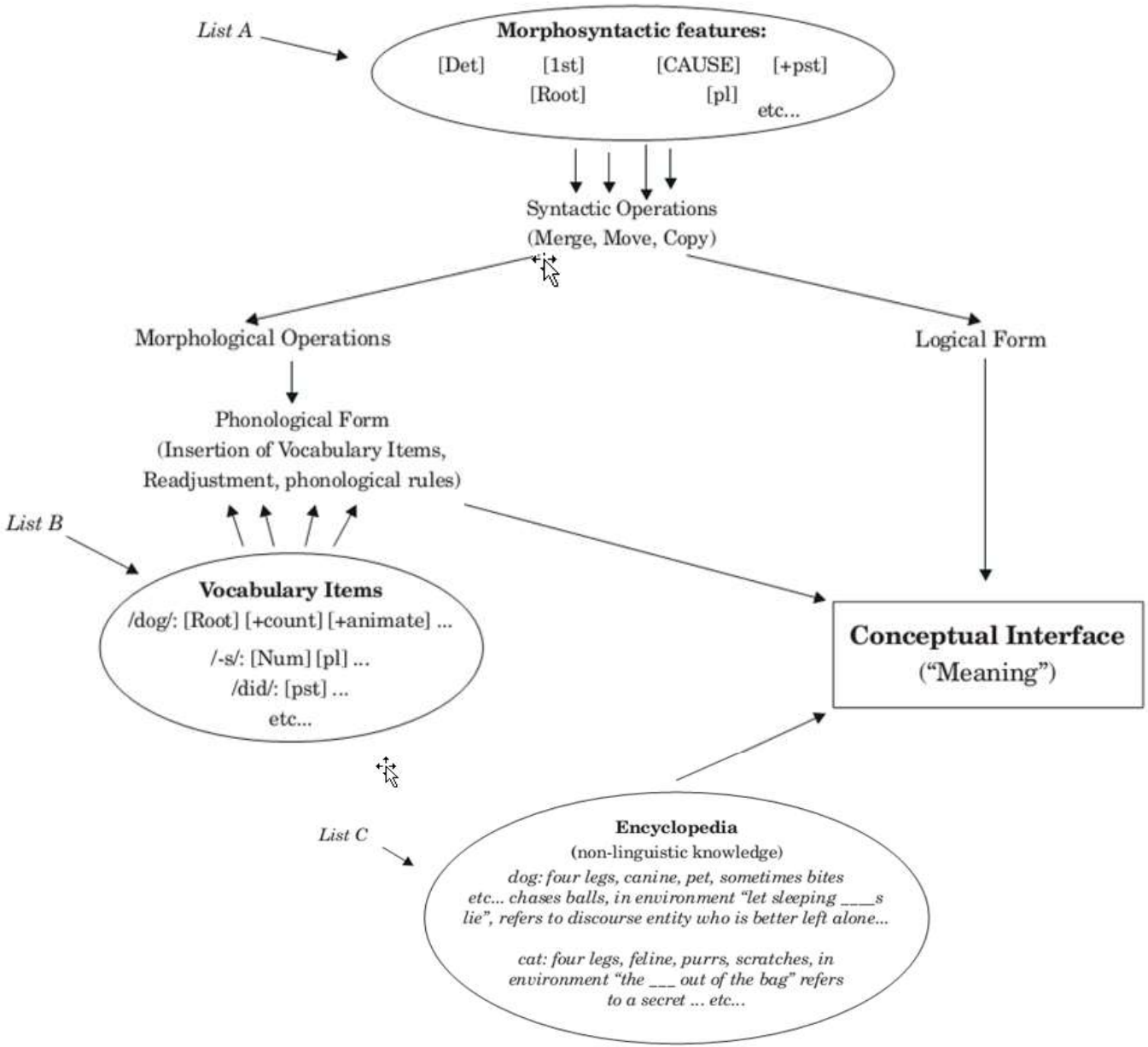
- ✓ **Late Insertion:** syntactic categories are abstract. Spell-Out is after syntax
- ✓ **Underspecification:** Vocabulary Items do not need to be fully specified
- ✓ **Syntactic Hierarchical Structure All the Way Down:** syntax and morphology are the same type of constituents





[4]Halle & Marantz. (1993). Distributed morphology and the pieces of inflection.
[5]Harley & Noyer. (1999). State-of-the-Article: Distributed Morphology

Distributed Morphology Schema

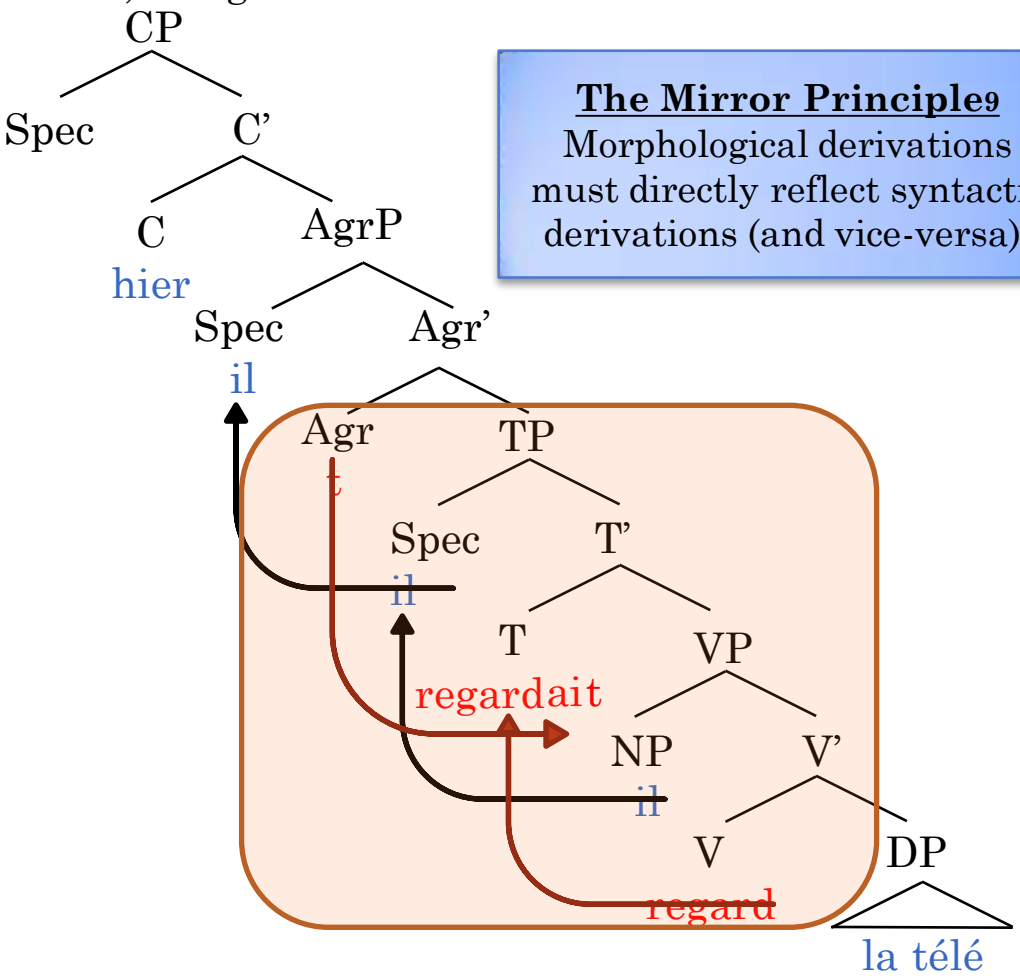


Harley & Noyer, 1999

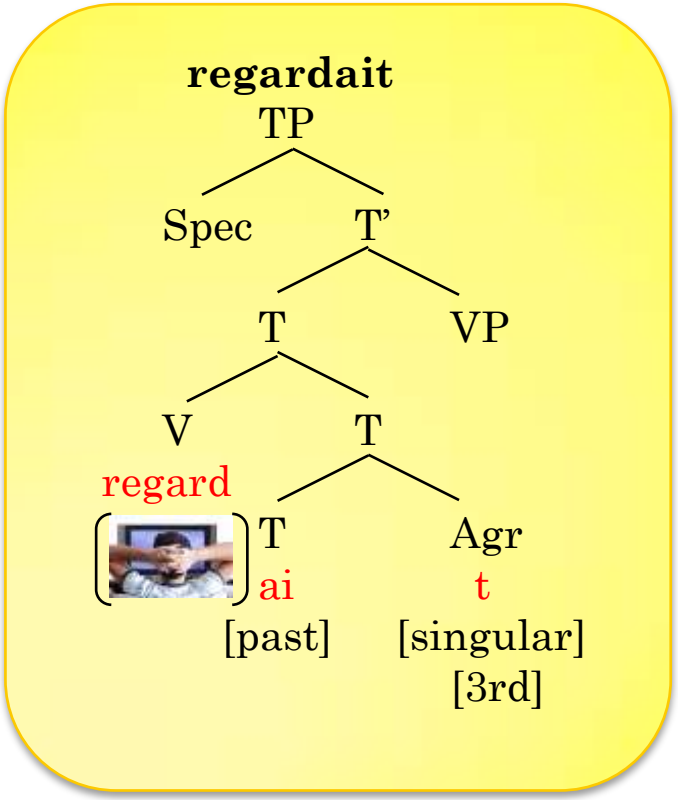
Syntactic Operations[7]

Morphological Operations[8]

Hier, il regardait la télé.



The Mirror Principle[9]
Morphological derivations must directly reflect syntactic derivations (and vice-versa).



[7]Emonds. (1978). The complex V' - V in French.
[8]Arregi, (2000). *How the Spanish verb works*.
[9]Baker. (1985). The mirror principle and morphosyntactic explanation

Objectives and Criteria

- **Mimicry** plausibility of the human cognitive system: visual recognition of French verbs
- **Recognize real verbs** from the 1st Class [-er] specifying: tense, number and person

- **Recognize pseudoverbs** specifying: tense, number and person
- Simulate visual recognition, frequency effects and computational effects[10,11]
- **Occam's Razor:** among competing hypothesis, the one with fewest assumptions should be selected

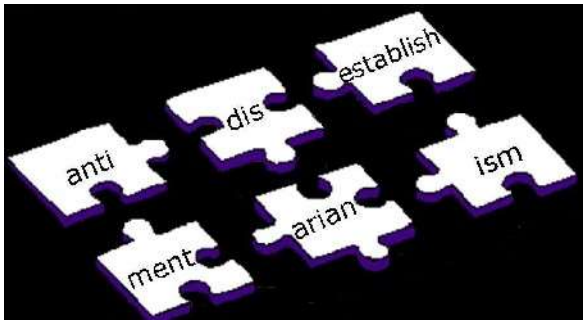
[10]Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition

[11]Coltheart et al. (2001). DRC: a Dual Route Cascade model of visual word recognition and reading aloud.

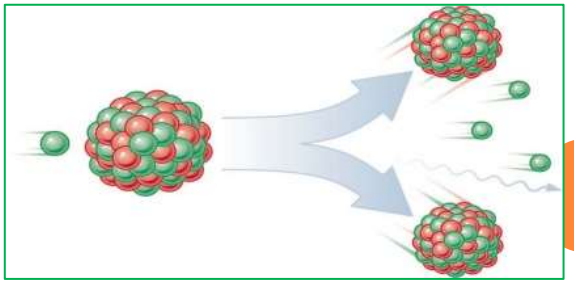
Operations and Functions

- **strcmp = Satisfy:** "selects an array of items from the lexicon and presents it in a format satisfying the conditions of X-bar theory" (Chomsky, 1995, p. 187)

- **strcat = Merge:** "joins terminal nodes under a category node of a head but maintain two independent terminal nodes under this category node" (Halle & Marantz, 1993, p. 116)



- **strsplit = Fission:** "cut one part of the word for future processing or comparison" (Harley & Noyer, 1999, p. 7)



Lexicon and Input

```
%% Close and clear all
close all
clear all

% Make the lexicon
vocabularyitems = dataset('XLSFile', 'dm_model', 'ReadVarNames', false,
vocabularyitems = dataset2cell(vocabularyitems);
vocabularyitems = vocabularyitems(2:56,1:7);

%% Visual recognition
clearvars -except vocabularyitems list k generalresult
input = 'parlerions';
word = input;
output = {''};
i = 1;
%% Compare input and output
while strcmp(input, output(1,1)) == 0
%%% Search and activation
    for j = 1:length(vocabularyitems)
        activation(j,:) = strncmp(word, vocabularyitems(j,1), length(vo
    end
    activated = vocabularyitems(activation,:);
%%% There is in lexicon
    if isempty(activated) == 0
%%%%%%% Morphemes features
        features = cellfun(@isvector, activated);
        sumfeatures = sum(features')';
        [maxfeatures, indexfeatures] = max(sumfeatures);
%%%%%%% Frequency
%%%%%%% Item selection
        morphemes(i,:) = activated(indexfeatures,:);
%%%%%%% Recomposition
        output = strcat(output, morphemes(i,:));
%%%%%%% Split stem
        splitroot = strsplit(word, morphemes(i,1));
        word = splitroot(1,2);
%%% There is no in the lexicon
    else
%%%%%%% Pseudoword selection
        word = word(2:end);
        splitsuffix = strsplit(input, word);
%%%%%%% Recomposition
        morphemes(1,:) = {splitsuffix{1,1} '@' '[pseudo]' '' '' '' '10(
        output = morphemes(1,:);
        i = i + 1;
    end
end
% Definitions
root = morphemes(1,:);
suffix = strsplit(input, root(1,1));
suffix = suffix{1,2};

if strcmp(output(1,4), '[future][past]')
   output{1,4} = '[conditional]';
end
%% Result
if strcmp(input, output{1,1}) == 1
   result = {'Good!' input output(1,1) list(k,2) morphemes(1,3) list(k,
else
   result = {'Bad!' input output(1,1) list(k,2) morphemes(1,3) list(k,3
end
generalresult(k,1:9) = result;
```

Morphological Operations

Result

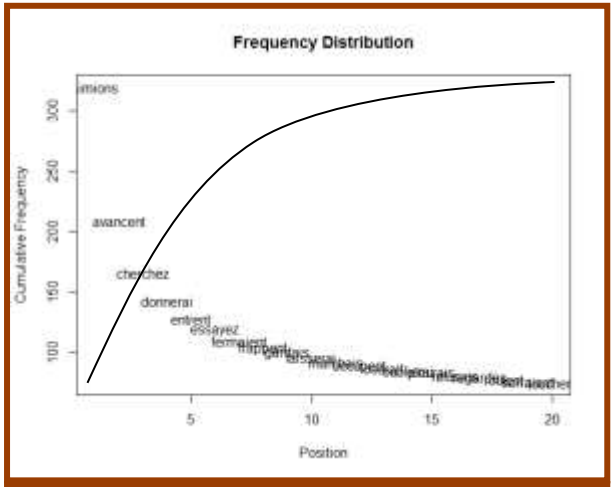
	A	B	C	D	E	F	G
1	ai	E	[t]	[past]			0,05
2	i	i	[t]	[past]			0,10
3	er	°R	[t]	[future]			0,15
4	r	R	[t]	[future]			0,20
5	dr	dR	[t]	[future]			0,25
6	e	º	[agr]				0,30
7	es	ºz	[agr]			[2]	0,35
8	s	z	[agr]			[1,2]	0,40
9	t	t	[agr]				0,45
10	ai	E	[agr]			[1]	0,50
11	as	az	[agr]			[2]	0,55
12	a	a	[agr]				0,60
13	ons	§z	[agr]		[plural]	[1]	0,65
14	ez	ez	[agr]		[plural]	[2]	0,70
15	ent	@t	[agr]		[plural]		0,75
16	ont	§t	[agr]		[plural]		0,80
17	donn	donn	[root]				1,00
18	regard	regard	[root]				1,28
19	aim	aim	[root]				1,58
20	laiss	laiss	[root]				2,00
21	cherch	cherch	[root]				2,32
22	entr	entr	[root]				2,58
23	essay	essay	[root]				2,81
24	march	march	[root]				3,00
25	gard	gard	[root]				3,17

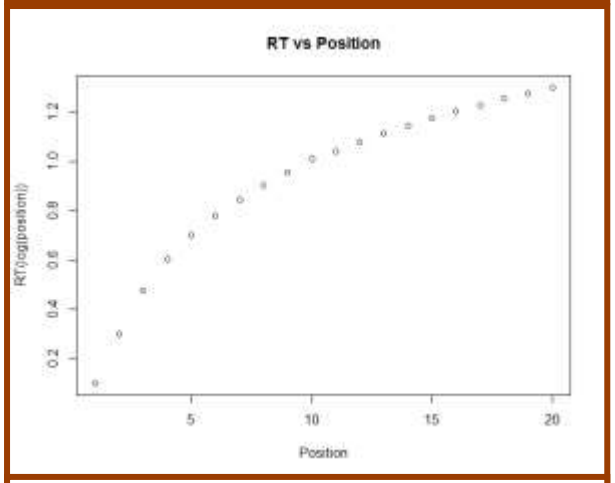
vocabularyitems | test | list | all_test | Feuil6 | data_matlab | Feuil5

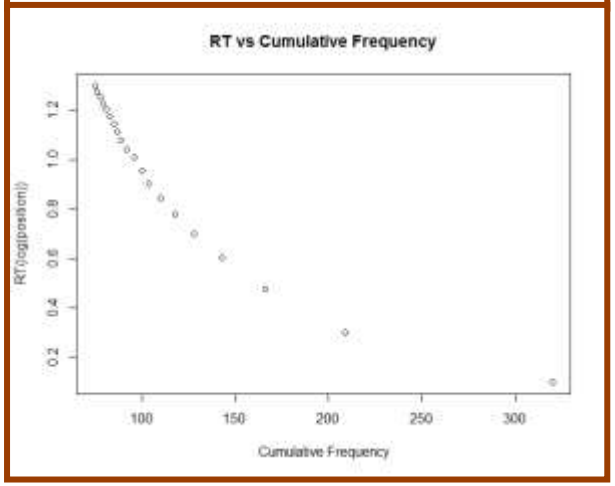
French Verb

Tense	Number	Person
[present]	[singular]	[1st],
[past]	[plural]	[2nd],
	[future]	[3rd]6

Kilani-Schoch & Dressler. (2005).
Morphologie naturelle et flexion du verbe français



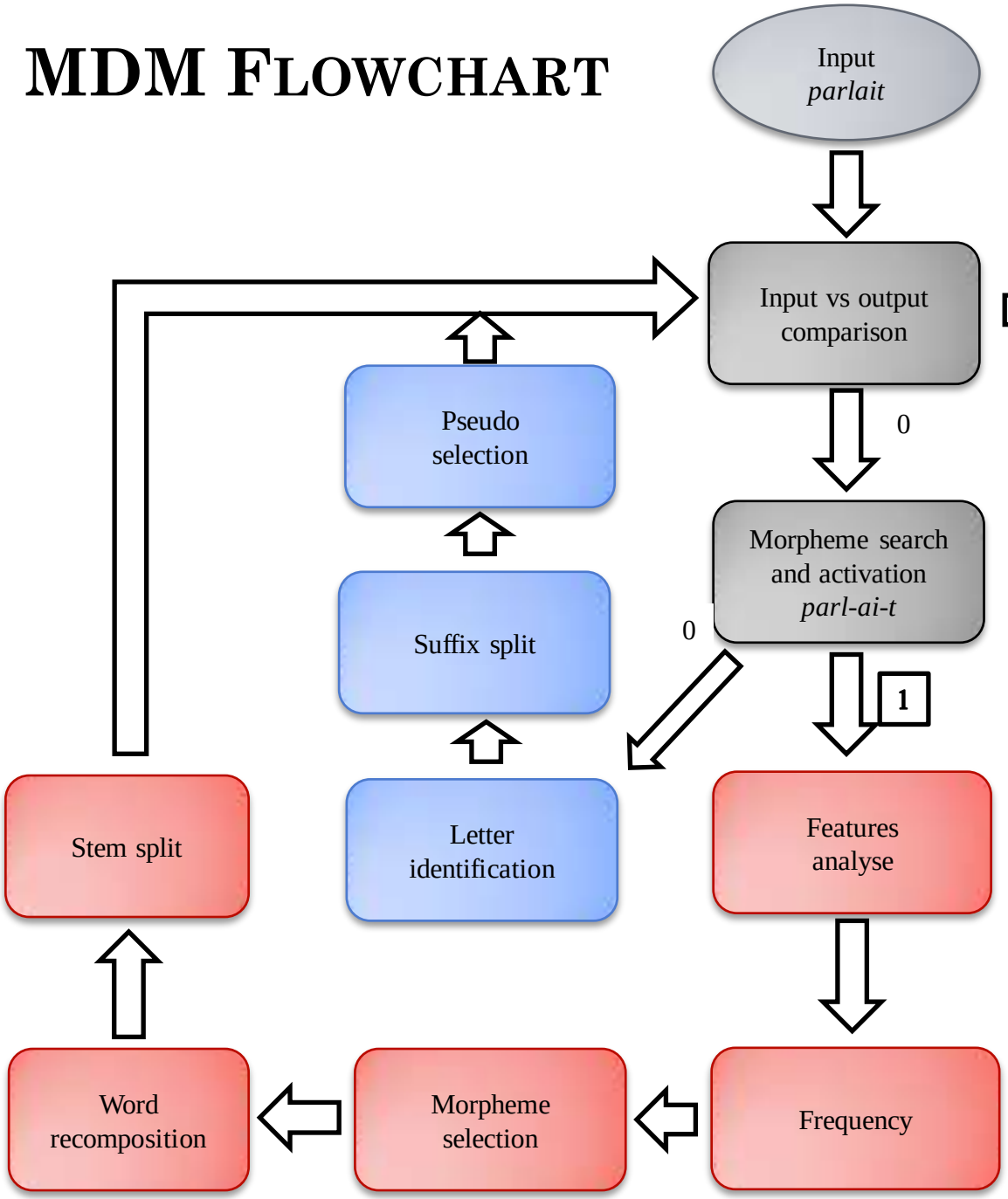




	A	B	C	D	E	F	G
1	ai	E	[t]	[past]			0,05
2	i	i	[t]	[past]			0,10
3	er	°R	[t]	[future]			0,15
4	r	R	[t]	[future]			0,20
5	dr	dR	[t]	[future]			0,25
6	e	º	[agr]				0,30
7	es	ºz	[agr]			[2]	0,35
8	s	z	[agr]			[1,2]	0,40
9	t	t	[agr]				0,45
10	ai	E	[agr]			[1]	0,50
11	as	az	[agr]			[2]	0,55
12	a	a	[agr]				0,60
13	ons	§z	[agr]		[plural]	[1]	0,65
14	ez	ez	[agr]		[plural]	[2]	0,70
15	ent	@t	[agr]		[plural]		0,75
16	ont	§t	[agr]		[plural]		0,80
17	donn	donn	[root]				1,00
18	regard	regard	[root]				1,28
19	aim	aim	[root]				1,58
20	laiss	laiss	[root]				2,00
21	cherch	cherch	[root]				2,32
22	entr	entr	[root]				2,58
23	essay	essay	[root]				2,81
24	march	march	[root]				3,00
25	gard	gard	[root]				3,17

vocabularyitems | test | list | all_test | Feuil6 | data_matlab | Feuil5

MDM Flowchart



verb = [root]+[t]+([t]+)[agr]

Verb
ΣFreq+operations

Pseudoword
ΣFreq+letterID+operations

Phonological rules?
Syntactic rules?
Word length?
Family size?
Neighbors frequency?
Surface frequency?

Results and Discussion

MDM simulation

output <2x9 cell>

	1	2	3	4	5	6	7	8	9
1	'regarderons'	'regard°R§z'	'[root][t][agr]'	'[future]'	'[plural]'	'[1]'	2.3400	'c1'	3
2	'regarde'	'regard°'	'[root][agr]'	[]	[]	[]	2.3400	'c1'	2

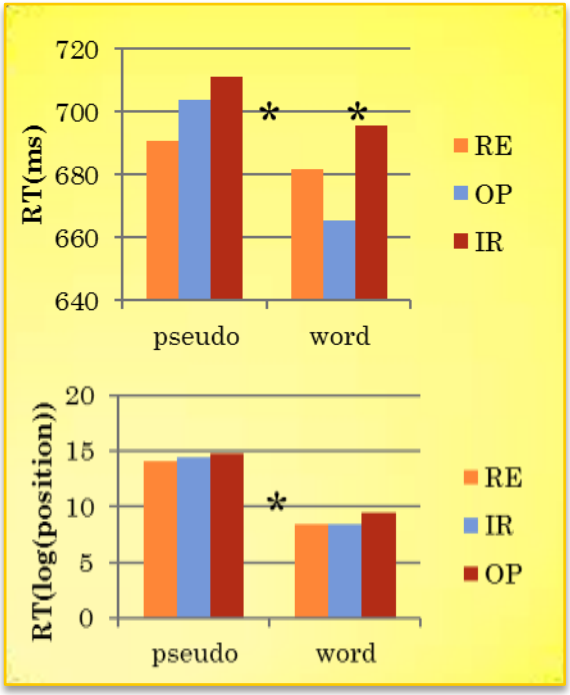
Psycholinguistic

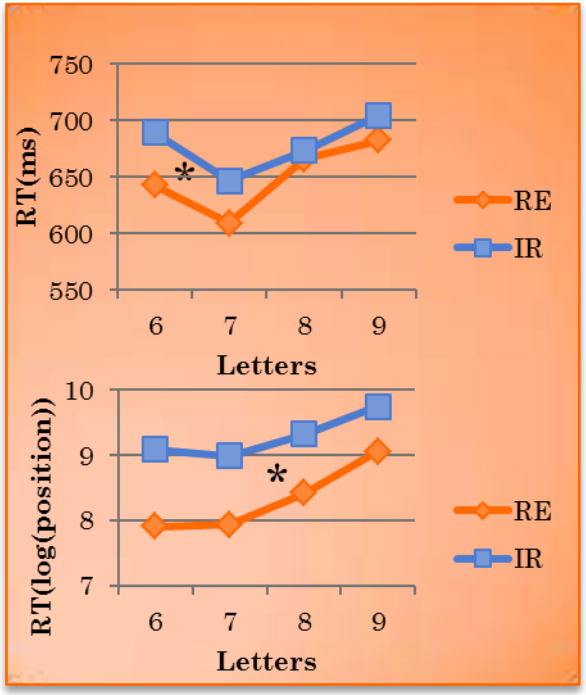
- 32 subjects, 16 women, right hand, French L1, mean age 20,1 years
- 300-1500ms (1,18% outliers)

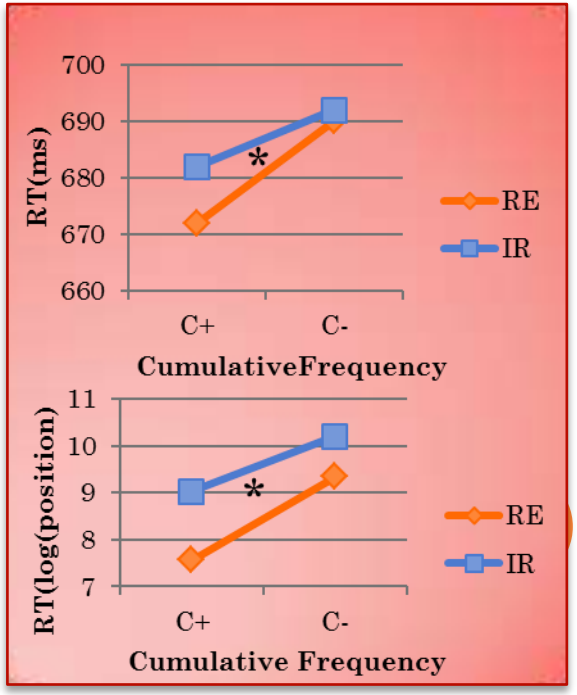
Accuracy/ Error	Psycholinguistic		MDM Simulation	
	Verb	Psuedo	Verb	Pseudo
Regulars	6,88%	4,71%	0%	5%
Irregulars	6,29%	6,85%	0%	3,75%
Operations	4,74%	5,19%	1,25%	2,50%

Psycholinguitic

MDM simulation

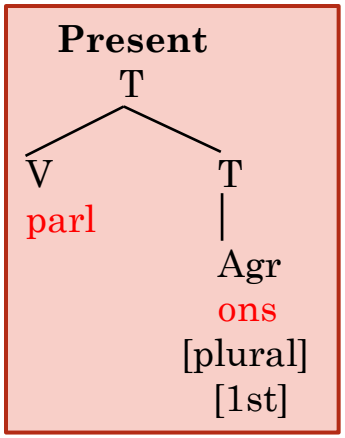


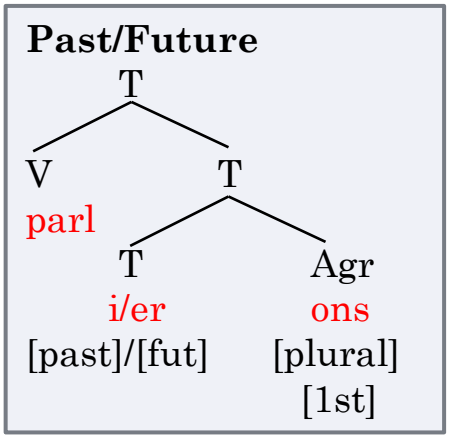


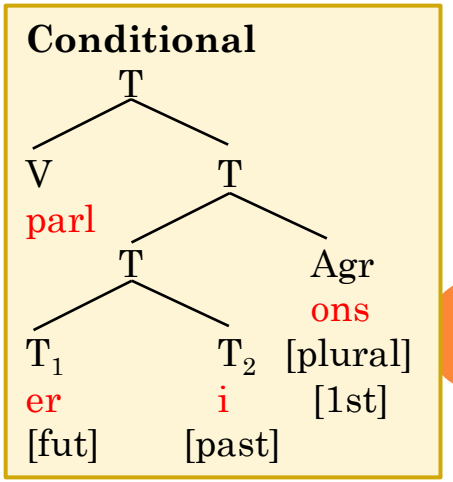


Results and Discussion

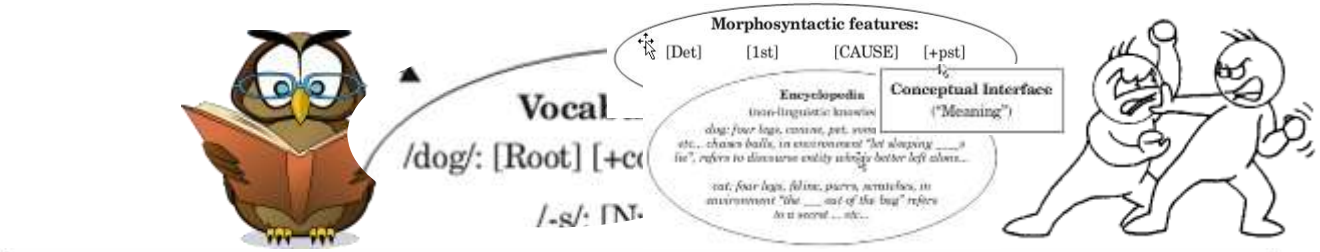
- French verbal structure
- The MDM1 can support 1st and 3rd French verbal classes (simulation: 240 inflected verbs, 2,08% error)
- Besides the root representation and allomorphy, the inflectional system is extremely regular and paradigmatic
- Allomorphy can be resolved by phonological rules and/or morphological rules
- Defragmented system
- ThV Argument:
  - 1. If we assume the first class productive "allomorphs" [e], [es] and [er] for [s], [s] and [r], the French verbal system does not need to have a thematic vowel
  - 2. The morphemes [e], [es] are the singular subjunctive morphemes for all classes







Discussion



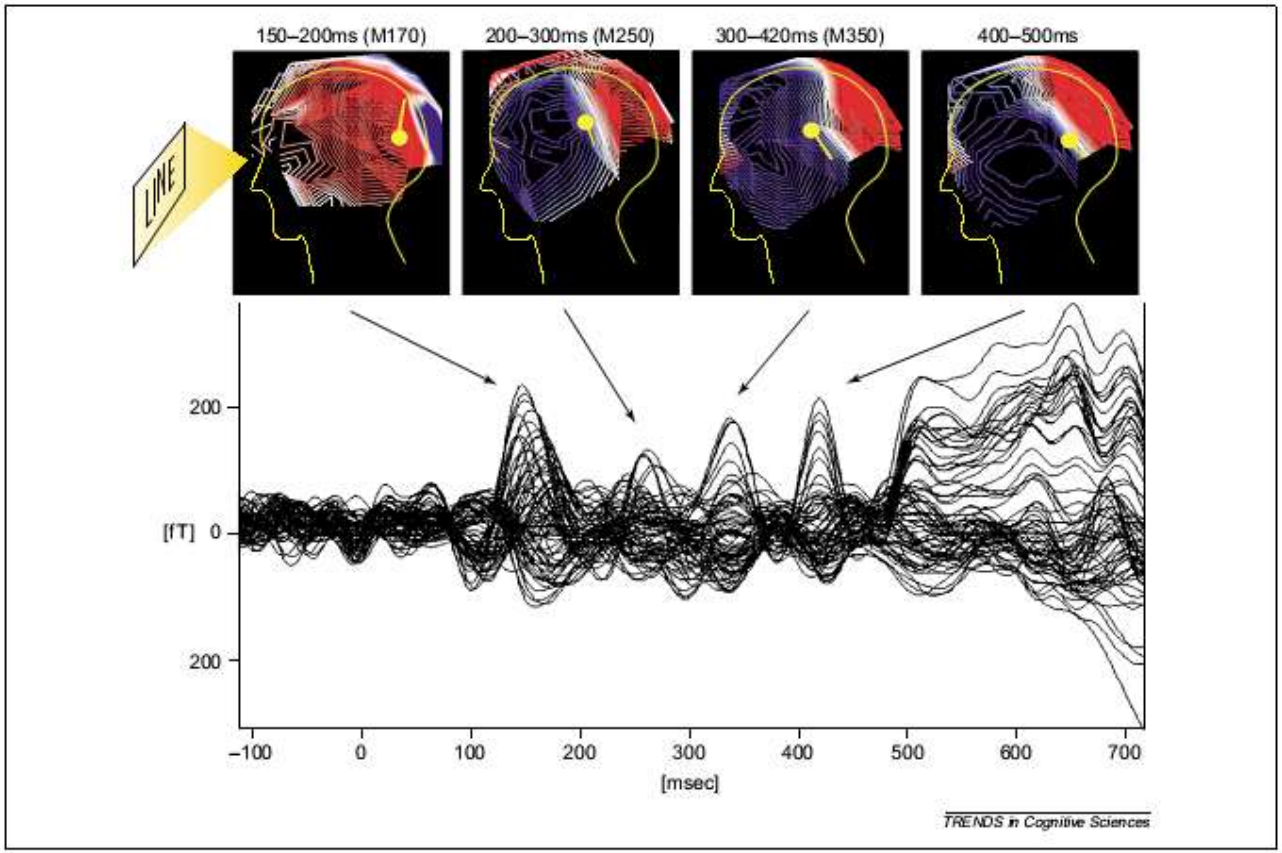


Fig. 1. A waveform illustrating the averaged MEG response of one subject to 69 presentations of visual words recorded with a 93-channel axial gradiometer whole-head system (Kanazawa Institute of Technology, Japan). Activity from all sensors is shown in the bottom panel; the magnetic field patterns of the major response components and their estimated current sources are in the top panel (single dipole modeling). The strong activity after 500 ms reflects the motor activity of the button press of the experimental task (lexical decision). Recent MEG results suggest that activity at 150–200 ms in the left hemisphere (M170) is associated with letter-string processing [14] and activity at 300–400 ms with lexical activation [11,12]. The functional significance of the activity in the 200–300 ms and 400–500 ms time windows is less clear. However, results obtained so far suggest that phonological stimulus factors, such as the frequency of the sounds in the word, affect the M250, and that activity at 400–500 ms is sensitive to the magnitude of interlexical competition [12]. The electrical N400, which is sensitive to both lexical and post-lexical stimulus factors [1], might be a summation of the M250, M350 and the post-M350 activity.

(M170)
Letter string processing

⇩

(M250)
Phonological factors

⇩

(M350)
Lexical access
freq., repet.

⇩

Interlexical competition

Pylkkänen & Marantz. (2003). Tracking the time course of word recognition with MEG.

Problems and Future

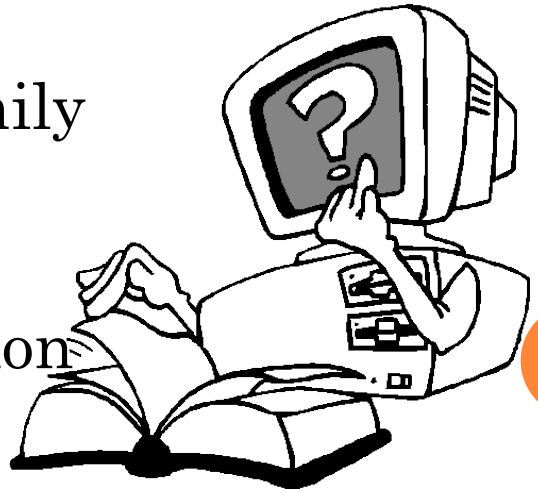
T O M C N A I R B
E Z A R T I E S R
G Y J F S A N D I
W E D A T N E R T
A E R D N A F A T
I H D I E E U F A
R N U G L W K I N
A Q B V E E U A Y
D E P T H L S W J

Problems

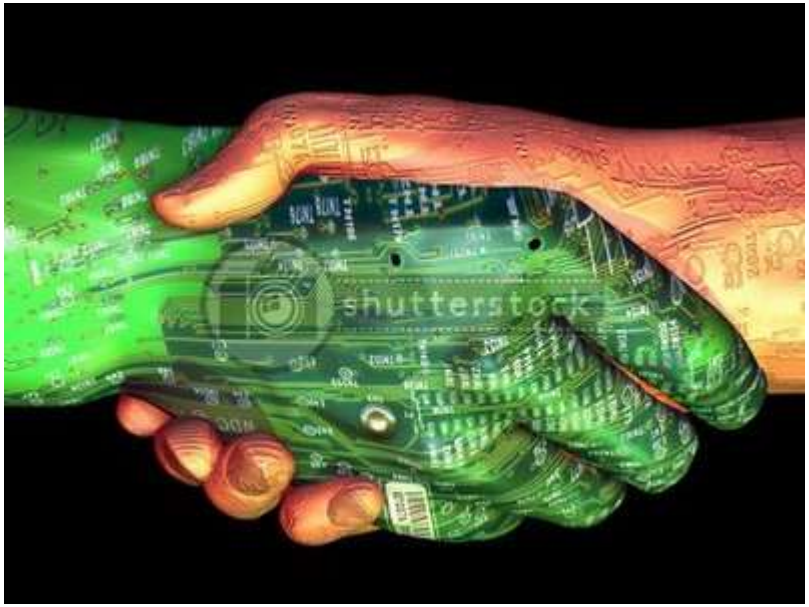
- Find more appropriate criteria measures to simulate the lexical activation, linguistic operations and RT simulation
- Pseudoword variability

Future

- 2nd class, subjunctive and prefix
- Incorporate more variables as: phonological rules, word length, family size, surface frequency
- Simulations: - thematic vowel
  - large corpus simulation
  - bilingual lexicon



THANK YOU!



References

- Arregi, K. (2000, 26 February 2000). *How the Spanish verb works*. LSRL 30, University o Florida, Gainesville.
- Baker, M. (1985). The mirror principle and morphosyntactic explanation. *Linguistic Inquiry, 16*(3), 373-415.
- Chomsky, A. N. (1995). *The Minimalist Program* (Vol. 28). Cambridge, MA: The MIT Press.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a Dual Route Cascade model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.
- Emonds, J. (1978). The complex V' - V in French. *Linguistic Inquiry, 9*(2), 151-175.
- Halle, M. (1973). Prolegomena to a theory of word formation. *Linguistic Inquiry, 4*(1), 3-16.
- Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building 20: essays in linguistics in honor of Sylvain Bromberger* (pp. 111-176). Cambridge, Massachusetts: The MIT Press.
- Harley, H., & Noyer, R. (1999). State-of-the-Article: Distributed Morphology. *Glot International, 4*(4), 3-9.
- Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition - sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1311-1334.

References

- Kilani-Schoch, M., & Dressler, W. U. (2005). *Morphologie naturelle et flexion du verbe français* (Vol. 488). Tübingen: Gunter Narr Verlag Tübingen.
- Kroll, J. F. (1993). Accessing conceptual representations for words in a second language. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (Vol. 6, pp. 53-81). Amsterdam: John Benjamins Publishing Company.
- Marslen-Wilson, W. D., & Tyler, L. K. (2007). Morphology, language and the brain: the decompositional substrate for language comprehension. *Philosophical Transactions of the Royal Society B, 362*, 823-836.
- Meunier, F., & Marslen-Wilson, W. D. (2004). Regularity and irregularity in French verbal inflection. *Language and Cognitive Processes, 19*(4), 561-580.
- New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: a new French lexical database. *Behavior Research Methods, Intruments, and Computers, 36*(3), 516-524.
- Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *TRENDS in Cognitive Sciences, 7*(5), 187-189.
- Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review, 11*(6), 1090-1098.

Regulars

	Regulars	C+					C-				
Nb	Pair	S+	S+Freq	S-	S-Freq	C+Freq	S+	S+Freq	S-	S-Freq	C-Freq
1	aimer/figurer	aimions	6,55	aimeront	0,34	795,61	figurez	6,15	figurera	0,34	57,23
2	avancer/briller	avancent	7,50	avancera	0,74	195,00	brillent	7,50	brillais	0,14	81,22
3	chercher/baigner	cherchez	7,36	cherchiez	0,74	448,99	baignait	6,62	baignons	0,14	41,42
4	donner/reculer	donnerai	6,15	donneriez	0,34	896,01	reculons	6,62	reculera	0,14	69,05
5	entrer/détacher	entrent	7,30	entriez	0,20	398,38	détache	7,23	détachez	0,41	65,47
6	essayer/inspirer	essayez	6,96	essayes	0,34	296,69	inspire	6,82	inspires	0,07	45,00
7	fermer/utiliser	fermaient	5,27	fermera	0,34	197,16	utilisait	5,41	utilisons	0,07	43,51
8	frapper/trembler	frappent	5,34	frappons	0,00	168,31	tremblent	5,68	trembliez	0,00	34,13
9	garder/discuter	gardais	5,74	garderas	0,61	257,50	discute	5,54	discutes	0,14	58,65
10	laisser/pardonner	laisserai	5,81	laisseras	0,74	851,55	pardonnez	5,74	pardonnes	0,54	44,59
11	marcher/insister	marchais	7,91	marcheras	0,27	325,61	insistait	7,97	insistons	0,20	67,03
12	occuper/habiller	occupent	7,30	occupiez	0,07	219,80	habille	7,97	habillez	0,20	67,36
13	oser/agiter	oserait	5,34	oserons	0,00	155,54	agitent	5,54	agitera	0,07	89,19
14	oublier/accuser	oublierai	6,55	oubliiez	0,07	286,96	accusait	6,55	accusons	0,14	39,93
15	pleurer/désirer	pleurais	5,07	pleurera	0,61	163,31	désirais	4,66	désireras	0,07	61,89
16	refuser/organiser	refusais	4,59	refuseras	0,27	152,77	organise	4,19	organisez	0,07	47,90
17	regarder/dépasser	regardes	5,34	regardiez	0,54	997,91	dépassent	5,74	dépassais	0,41	78,78
18	rouler/examiner	roulent	6,28	roulera	0,34	163,45	examine	6,28	examines	0,07	50,68
19	serrer/attaquer	serraient	5,81	serrerons	0,07	207,50	attaquait	6,01	attaquez	0,41	70,41
20	toucher/admirer	touchent	6,69	touchons	0,81	190,27	admirais	6,35	admires	0,41	68,18

New et al, 2004

IRREGULARS

	Irregulars	C+					C-				
Nb	infinitive	S+	S+Freq	S-	S-Freq	C+Freq	S+	S+Freq	S-	S-Freq	C-Freq
1	apercevoir	apercevait	25,68	apercevrai	0,20	85,43	aperçoit	21,82	aperçoives	0,14	42,49
2	apprendre	apprendra	5,07	apprendrez	0,74	107,50	apprenais	5,47	apprenons	0,68	31,97
3	boire	boirai	1,55	boiras	0,74	148,18	buvions	2,36	buviez	0,14	54,12
4	connaître	connaissiez	2,84	connaisses	0,74	335,26	connaîtrait	2,77	connaîtrez	0,74	160,28
5	craindre	craignait	20,14	craignes	0,00	49,66	crains	17,64	craindra	0,07	38,18
6	devenir	devenais	6,15	deveniez	0,14	376,50	deviendrait	8,58	deviendras	0,95	89,26
7	envoyer	envoyaient	2,91	envoyions	0,07	131,57	enverrai	2,70	enverras	0,47	10,27
8	mourir	mourait	11,35	mouriez	0,00	169,60	meurent	10,47	meures	0,14	42,51
9	obtenir	obtenait	2,36	obtenons	0,41	62,71	obtient	3,18	obtiendra	0,41	7,31
10	parvenir	parvenais	5,07	parveniez	0,07	95,42	parviens	5,88	parviendra	1,22	7,31
11	prévenir	prévenait	2,23	prévenais	0,07	58,46	prévient	2,03	préviendra	0,20	10,21
12	recevoir	recevaient	4,93	recevions	0,81	94,33	reçoivent	4,19	reçoives	0,00	27,36
13	rejoindre	rejoignent	5,07	rejoignais	0,34	40,97	rejoins	3,65	rejointes	0,54	29,59
14	reprendre	reprends	7,16	reprendras	0,07	114,73	reprenaient	5,88	reprenions	0,68	54,40
15	retenir	retenait	15,34	retenions	0,34	103,52	retient	13,11	retiendra	0,54	21,70
16	revoir	revoyais	4,73	reverront	0,74	87,31	reverrai	4,05	revoyons	0,41	18,51
17	souvenir	souvenaient	2,16	souvenions	0,27	94,33	souvienne	3,38	souviennes	0,27	6,69
18	surprendre	surprend	7,09	surprenons	0,14	28,46	surprenait	7,30	surprendra	0,61	13,53
19	tenir	tenions	4,80	teniez	0,95	525,82	tiendra	4,59	tiendrez	0,74	193,04
20	valoir	valaient	5,27	valais	0,47	82,31	vaille	5,34	vaillent	0,20	5,54

New et al, 2004

Operation

	Operations		C+							C-				
Nb	Operation	Pair	Infinitive	S+	S+Freq	S-	S-Freq	C+Freq	infinitive	S+	S+Freq	S-	S-Freq	C-Freq
1	O-	traverser/adresser	traverser	traversons	2,84	traversez	0,54	200,81	adresser	adressent	1,55	adresso	0,47	98,04
2	O-	tourner/dépêcher	tourner	tournons	1,82	tournes	0,95	377,09	dépêcher	dépêchons	1,55	dépêch	0,47	22,84
3	O-	raconter/dîner	raconter	racontez	5,74	racontons	0,27	261,89	dîner	dînons	1,89	dînent	0,47	59,66
4	O-	poser/estimer	poser	posent	7,50	posons	0,34	409,05	estimer	estiment	1,49	estimo	0,54	37,64
5	O-	expliquer/fatiguer	expliquer	expliquent	2,09	expliquons	0,20	233,92	fatiguer	fatiguent	1,55	fatiguo	0,00	49,05
6	O-	montrer/presser	montrer	montrez	2,57	montrons	0,27	276,55	presser	pressons	1,89	presser	0,81	71,01
7	O-	décider/risquer	décider	décident	2,57	décidez	0,88	214,19	risquer	risquent	3,85	risquon	0,95	99,32
8	O-	coucher/sauver	coucher	couchent	3,38	couchons	0,41	196,22	sauver	sauvent	1,28	sauvons	0,27	99,05
9	O-	changer/supposer	changer	changent	5,34	changeons	0,68	246,49	supposer	supposons	1,62	suppos	0,88	62,50
10	O-	ajouter/tromper	ajouter	ajoutent	1,89	ajoutes	0,61	224,66	tromper	trompent	3,11	trompo	0,34	97,30
11	O+	traverser/adresser	traverser	traversais	2,46	traversera	0,68	200,81	adresser	adressais	1,49	adresse	0,34	98,04
12	O+	tourner/dépêcher	tourner	tournais	3,58	tournera	0,68	377,09	dépêcher	dépêchait	1,15	dépêch	0,07	22,84
13	O+	raconter/dîner	raconter	racontais	5,00	racontiez	0,61	261,89	dîner	dînait	3,24	dîniez	0,00	59,66
14	O+	poser/estimer	poser	posais	5,41	posiez	0,27	409,05	estimer	estimais	2,16	estimie	0,07	37,64
15	O+	expliquer/fatiguer	expliquer	expliquera	1,89	expliquiez	0,27	233,92	fatiguer	fatiguait	2,36	fatigue	0,14	49,05
16	O+	montrer/presser	montrer	montrais	2,43	montriez	0,41	276,55	presser	pressais	1,22	presser	0,07	71,01
17	O+	décider/risquer	décider	décidera	2,16	décidiez	0,14	214,19	risquer	risquais	3,51	risquie	0,61	99,32
18	O+	coucher/sauver	coucher	couchais	2,57	couchiez	0,20	196,22	sauver	sauvera	1,49	sauvais	0,54	99,05
19	O+	changer/supposer	changer	changera	5,07	changions	0,27	246,49	supposer	supposais	2,30	supposi	0,07	62,50
20	O+	ajouter/tromper	ajouter	ajoutais	1,42	ajoutera	0,68	224,66	tromper	trompais	4,12	trompie	0,07	97,30

New et al, 2004