

L'anti-correcteur, pourquoi ?

- Porter un regard différent sur l'erreur (Gonac'h 2015)
- Proposer une évaluation positive de l'orthographe (Brissaud & Mortamet (Eds) 2015) : souligner les réussites plutôt que les erreurs
- Mesurer l'acquisition de compétences d'écriture tout en constituant une aide à l'apprentissage (Cogis et al. 2015, Anxionnaz 2015)
- Se concentrer sur des « zones de résistances » (Roubaud 2014)

Sélection de zones de résistances pertinentes

État de l'art – linguistique et didactique du français

- Erreurs fréquentes liées à une règle complexe (Roubaud 2014, Anxionnaz 2015)
- Erreurs graves car règle a priori évidente (Groupe R.O 2012)

Cinq secteurs propices à la résistance

- Segmentation en phrases
- Accord verbal
- Accord au sein du groupe nominal
- Orthographe lexicale – doubles consonnes, noms composés, finales muettes, accentuation
- Orthographe lexicale – homonymie/phonie, e.g. à/a, on/ont, et/est

État en corpus – CeL*, 3 mois de récolte : 8 487 257 erreurs

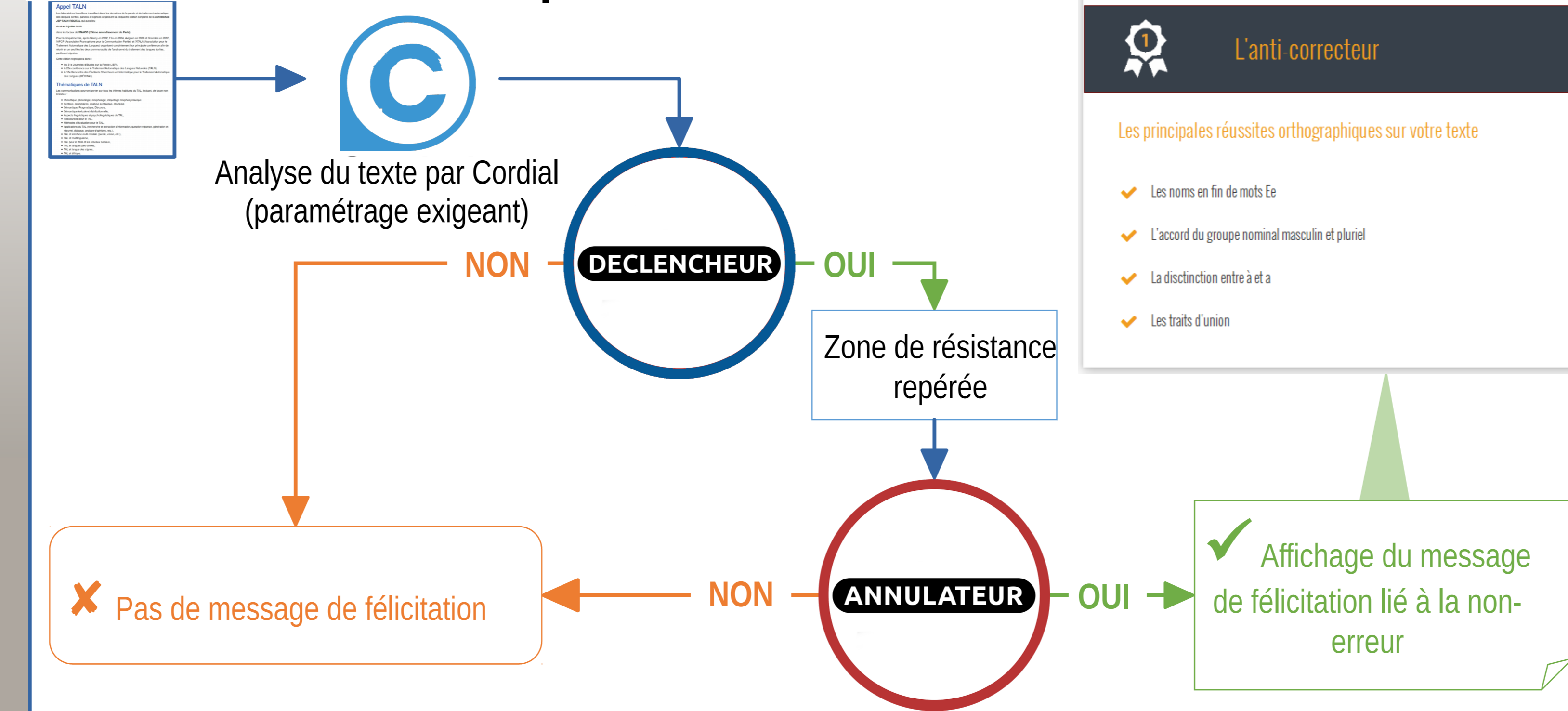
- 78,4 % d'erreurs non liées à un secteur propice à la résistance
- 21,6 % des erreurs liées à des zones de résistance

Constitution de listes de mots fréquemment erronés :

Zone	Description	Nb
L_TDU	mots entre lesquels un tiret est fréquemment ajouté à tort	9 948
L_NCP	noms composés pluriels qui sont fréquemment mal orthographiés	449
L_ACC	mots fréquemment mal accentués	19 331
L_CON	mots fréquemment mal orthographiés à cause d'une double consonne, soit par oubli de la double consonne, soit par dédoublement erroné d'une consonne	10 438
L_EES	noms se terminant en -é(e)s fréquemment mal orthographiés	339
L_SIM	mots homonymes et paronymes fréquemment confondus	360 957

*CeL : <http://cordial-enligne.fr/>

L'anti-correcteur c'est quoi ?



Zones actuellement détectées

Zone	Description (exemple)	
	Déclencheur	Annulateur
Z_PHR	Phrase complète (<i>c'est pas possible!</i>) Par défaut	<ul style="list-style-type: none"> absence de verbe négation incomplète absence des COD et/ou COI d'un verbe (bi)transitif absence d'un déterminant devant les noms sujets
Z_TDU	Trait d'union (<i>Que fais tu là princesse ?</i>) <ul style="list-style-type: none"> mot contenu dans L_TDU séquence /.-/ 	<ul style="list-style-type: none"> présence erronée ou absence d'un trait d'union dans : un mot inconnu du dictionnaire une construction syntaxique en nécessitant un
Z_NCP	Noms composés pluriels (<i>elle baissa les yeux sur les chou-fleurs.</i>) mot inclus dans la liste L_NCP	mot inconnu dont la suggestion de correction est contenue dans la liste L_NCP
Z_ACC	Accentuation (<i>etre</i>) mot contenu dans L_ACC	<ul style="list-style-type: none"> erreur d'accentuation dans un mot inconnu du dictionnaire confusion entre formes verbales problème d'accentuation d'une majuscule en début de mot
Z_GNP	Accord du groupe nominal masculin pluriel (<i>que de bon films.</i>) groupe nominal (analyseur syntaxique Cordial)	erreur d'accord en nombre et/ou genre dans un groupe nominal pluriel
Z_APP	Accord avoir + participe passé (<i>elle a amenée.</i>) temps composé (analyseur syntaxique Cordial)	erreur d'accord d'un participe passé précédé de l'auxiliaire avoir conjugué.
Z_AvA	Mot similaire à vs. a (<i>Elle apprenait à ce battre, a manier les armes.</i>) caractère A, a, À ou à	confusion entre "à" et "a" (également détecté comme Z_SIM)
Z_CON	Doubles consonnes (<i>apelle</i>) <ul style="list-style-type: none"> mot contenu dans L_CON double consonne 	<ul style="list-style-type: none"> doublement erroné de consonne oubli d'une consonne double
Z_EES	Noms se terminant en -é(e)s (<i>activités</i>) mot contenu dans L_EES	Déclencheur
Z_SIM	Mots similaires (<i>Désoler, j'ai états un peu souffrante la semaine dernière.</i>) mot contenu dans L_SIM	confusion entre des homophones/homographes

Mise à l'épreuve

- Répartition des zones et taux de réussite dans un corpus diversifié

Source	Genre	Auteurs	# Textes	# Phrases	# Mots
Corpus total			44 370	87 014	1 867 063
littéracie avancée (Rinck & Jacques 2015)	dissertation	L2	10	1 333	27 152
	note de lecture	M2	20	1 802	39 843
	lettre de motivation	M1	20	299	6 552
wikipedia (Ho-Dac & Laippala 2015)	rapport de stage	L3	15	3 249	66 857
	internaute		1 000	3 957	64 425
vikidia	post de discussion	" (7-14 ans)	529	2 941	32 861
enceinte.com	internaute		843	6 972	133 264
rue89.com	article de presse	professionnel	1 000	26 845	566 282
agoravox.fr	internaute		1 000	39 616	929 827

- Proportion de phrases contenant au moins une zone de résistance
- Corpus affichant significativement plus/moins de zones (z>2)
- Taux de réussite sur la zone
- Corpus affichant le plus mauvais taux de réussite (aucun z>2)

	Phrases contenant au moins une zone				Taux de réussite	
	Nb	%	Min (%)	Max (%)	%	Min (%)
Z_PHR	86 761	100			87	
Z_TDU	70 046	80	Vikidia (55)		96	
Z_NCP	2 935	3		Rapport (10)	95	Forum (91)
Z_ACC	82 839	95			92	Forum (68)
Z_GNP	47 380	54			98	Forum (77)
Z_APP	15 590	18		Rapport (35)	100	Forum (93)
Z_AvA	34 954	40			98	Forum (77)
Z_CON	84 190	97			98	
Z_EES	15 404	18			99	
Z_SIM	84 412	97			96	Forum (79)

Premiers résultats

- Des zones plus fréquentes : Orthographe lexicale (1) – Z_ACC, Z_SIM, Z_CON
- Des zones très rares et généralement sans fautes : Z_NCP, Z_AAP et Z_EES
- Un taux de réussite global plutôt bon
- Des variations entre corpus

Forums de santé	E	De grosses difficultés. De gros efforts à fournir
Lettres de motivation	B	Bon dans l'ensemble mais pourrait davantage s'investir
Articles de presse Rapports de stage	A	Bonnes prises de risque, continuez comme ça !

« ta accepter la vie qu'on ta offert comme une papillote de l'enfer tes patriote et t'en est fière ta pas compris que se monde est t'a refaire »
Anonyme, CeL