

L'expression	dénote	remarque
.	$\Sigma$	. vaut pour n'importe quel symbole
Répétitions		
$e^*$	$e^*$	
$e^+$	$ee^*$	
$e?$	$e + \varepsilon$	
$e\{n\}$	$(e^n)$	
$e\{n,m\}$	$(e^n + e^{n+1} \dots + e^m)$	à condition que $n \leq m$
Regroupements		
$[abc]$	$(a + b + c)$	$a, b, c$ sont des caractères
$[a-z]$	$(a + b + c \dots z)$	utilise l'ordre des caractères ASCII
$[^a-z]$	$\Sigma \setminus \{a, b, c\}$	n'inclut pas le symbole de fin de ligne $\backslash n$
Ancres		
$\backslash < e$	$e$	$e$ doit apparaître en début de mot, ie. précédé d'un séparateur (espace, virgule...)
$e \backslash >$	$e$	$e$ doit apparaître en fin de mot, ie. suivi d'un séparateur (espace, virgule...)
$^ e$	$e$	$e$ doit apparaître en début de ligne
$e \$$	$e$	$e$ doit apparaître en fin de ligne
Caractères spéciaux		
$\backslash .$	.	
$\backslash *$	*	
$\backslash +$	+	
$\backslash n$		dénote une fin de ligne
...	+	

Tab. 2.2 – Définition des motifs pour `grep`

```
> grep -E '([îâû]n?t|[îâû]mes|[iau]n?ss(e|es|ions|iez|ent))' Balzac.txt
```

Cette expression est un peu trop générale, puisqu'elle inclut des séquences comme `unssiez` ; pour l'instant on s'en contentera. Pour continuer, revenons à notre ambition initiale : chercher des verbes. Il importe donc que les terminaisons que nous avons définies apparaissent bien comme des suffixes. Comment faire pour cela ? Imposer, par exemple, que ces séquences soient suivies par un caractère de ponctuation parmi : `[ , ; ! : ? ]`. On pourrait alors écrire :

```
> grep -E '([îâû]n?t|[îâû]mes|[iau]n?ss(e|es|ions|iez|ent))[ , ; ! : ? ]' \
Balzac.txt
```

indiquant que la terminaison verbale doit être suivie d'un des séparateurs. `grep` connaît même une notation un peu plus générale, utilisant : `[ :punct : ]`, qui comprend toutes les ponctuations et `[ :space : ]`, qui inclut tous les caractères d'espacement (blanc, tabulation...). Ce n'est pourtant pas cette notation que nous allons utiliser, mais la notation `\>`, qui est une notation pour `elorsque` celui-ci est trouvé à la fin d'un mot. La condition que la terminaison est bien en fin de mot s'écrit alors :