

TP. Transformer des tables avec l'éditeur VI et les commandes Unix / cdspack

– 3 juin 2022 –

Cas pratiques :

I. Transformer les tables 3 et 4 de 2020AJ....159..190L afin que chaque mesure soit contenue dans un champs utilisable par un programme.

1. Récupérer les tables, supprimer les lignes d'en-tête et notes mais inclure les champs « COSMOS », « GOODS-N » et « EGS » dans les sous-ensembles.
2. Mettre en forme les coordonnées
3. Comparer les deux tables pour savoir si elles contiennent les mêmes objets.

II. Transformer la table 1 de 2021AJ....161..111D et supprimer les lignes où les colonnes 2MASS ET OH sont vides.

III. Transformer les tables 1 et 2 de 2019MNRAS.483.4160 et les fusionner

IV. Transformer la table 1 de 2021AJ....162..276Z pour avoir les noms et coo et y ajouter la colonne Classification de la table 2 (ou 3).

V. Récupérer la Table 1 (PDF) de 2019MNRAS.488.1790L

VI. Transformer la Table A1 de 2020MNRAS.492.3728B

VII. Table 1 de 2022ApJS..258....8M : Ajouter les noms Gaia DR2 dans la col. Name pour avoir une colonne de noms complète sans modifier l'alignement de la table.

Solutions :

I. Transformer les tables 3 et 4 de 2020AJ....159..190L

A. Récupération des Tables 3 et 4 dans un répertoire :

wget -c « URL » -O t3_ascii.txt (idem pour t4)

L'URL étant récupérée via clic droit « copier le lien » sur l'hyperlien « ASCII » du journal.

Option -O : permet de donner le nom qu'on souhaite au fichier récupéré.

man wget pour les différentes options.

On a le fichier **t3_ascii.txt**

B. Transformation de la table en .tsv pour chaque mesure

** Ouvrir la table dans VI*

vi t3_ascii.txt (c'est en fait un alias pour vim)

** Éliminer l'en-tête :*

d1G

** Eliminer les notes de bas de page :*

dG

** Récupérer les sous-ensembles :*

:3,28!acut -i'COSMOS' -i -c1- pour ajouter COSMOS devant les lignes concernées.

N.B. : Le second « -i » tout seul permet d'insérer le caractère défini comme séparateur de champ par défaut (en l'occurrence la tabulation).

d1G pour enlever l'en-tête COSMOS (ou **d1j** ou **d1k** selon comme on est tourné)

:30,57!acut -i'GOODS-N' -i -c1- pour ajouter GOODS-N devant les lignes concernées

d2j pour enlever l'en-tête GOODS-N

!G lorsque l'on est sur la bonne ligne (pour avoir le prompt **:[ligne-courante],\$!**) puis

acut -i'EGS' -i -c1- pour ajouter EGS devant les lignes concernées

d2j pour enlever l'en-tête EGS.

** Transformer les caractères qu'on ne veut pas conserver*

:%s/^h/ / pour transformer les ^h en espace dans tout le fichier

-idem pour les autres ou bien pour tout faire d'un coup :

:set magic pour activer les expressions régulières

:%s/^[d-m]/ /g pour transformer les ^h, ^m et ^d en 1 espace – ne pas oublier « g » pour l'exécution sur toute la ligne.

:%s/s// pour transformer les « s » en rien.

:%s/ +or- /t/g pour transformer les « +or- » en tabulations

On pourrait s'arrêter là puisqu'on a désormais un fichier tsv avec chaque champs séparé par un séparateur de champs (la tabulation – que l'on peut transformer par un « | » ou n'importe quel autre séparateur que l'on souhaiterait).

* Utiliser anafile pour réaligner avec les bons formats en une seule commande

Si on souhaite également réaligner les coordonnées pour avoir 2 chiffres en RA_m, DE_d, etc. Le plus simple à ce stade c'est **anafile** (voir **man anafile**).

Enregistrer votre nouveau fichier tsv (appelons le « **t3.tsv** ») et quitter.

:w t3.tsv

:q! pour quitter sans enregistrer et conserver votre table d'origine

anafile -d -ccg t3.tsv > prog_t3 pour créer la commande qui va reformater la table.

N.B. : **-d** indique que le séparateur de champs est celui considéré par défaut (toujours la tabulation, ça tombe bien)...

* Dans le fichier **prog_t3** :

d1G et **dG** pour ne conserver que la commande acut.

Ajouter « > table3.dat » pour re-diriger directement le résultat de la commande dans une table toute propre (passer en mode « insertion » pour écrire dans le fichier).

On peut faire :

:%s/i' 'i'/'g pour remplacer les séparateurs (espaces) par des « | »

On a alors un fichier « **prog_t3** » avec la commande :

**fcut t3.tsv | acut -d **

```
    -f1%-7j -i'|' -f2%-7j -i'|' -f3%11.2j -i'|' -f4%11.3j\  
-i'|' -f5%7.3j -i'|' -f6%5.3j -i'|' -f7%5.3j -i'|' -f8%3j \  
-i'|' -f9%3j -i'|' -f10%4.2j -i'|' -f11%6.4j -i'|' -f12%4.2j > table3.dat
```

:wq pour enregistrer et quitter

source prog_t3 pour exécuter la commande du fichier ci-dessus.

Là, petit problème pour les champs f3 et f4, nos fameuses coo => le résultat d'anafile est quelque peu effrayant mais on arrive à reconnaître que le problème concerne ces deux champs-là. En fait, en remplaçant « -f3%11.2j -i'|' -f4%11.3j » par « -f3%11j -i'|' -f4%11j » dans la commande ci-dessus (**x** pour supprimer les caractères indésirables), le programme ne râle plus (11j signifie 11 caractères alpha-numériques quelconques) et si on a remplacé « > table3.dat » par >| table3.dat, lorsqu'on relance la commande, on obtient une jolie table sans message d'erreur.

* Réalignement des coordonnées

Les coo de la **table3.dat** sont néanmoins toujours désalignées... On va pouvoir les réaligner en partant de la gauche de la colonne des RA :

!G puis

acut -c1-16 -c17-19l -c20-l pour réaligner les RA_h à gauche

!G puis on rappelle la commande précédente avec flèche vers le haut pour la transformer en

.,,\$!acut -c1-19 -c20-22l -c23- pour réaligner les RA_m à gauche puis

.,,\$!acut -c1-19 -c20-21r -c22- pour réaligner les RA_m à droite et

.,,\$!trcol -t20 ' ' '0' pour ajouter le fameux zéro manquant...

A peu près idem pour les DEC.

.,,\$!acut -c1-28 -c29-31l -c32- pour réaligner DE_d à gauche

.,,\$!acut -c1-28 -c29-30r -c31- pour réaligner DE_d à droite

Pour VizieR on prendra soin d'ajouter un signe « + » devant DEd sinon le programme n'est pas content.

```
.,,$!acut -c1-28 -i+' -c29-
```

```
.,,$!acut -c1-32 -c33-35l -c36- pour réaligner DEm
```

```
.,,$!trcol -t23,30,36 ' ' '0' pour ajouter les zéros manquants
```

N.B. : dans tous les cas,

u pour défaire

ctrl+r pour refaire

Flèche vers le haut pour rappeler une commande (fonctionne aussi dans le Shell)

Cette fois-ci la table contient à la fois un séparateur de champs pour mes colonnes et est également alignée correctement pour une lecture humaine.

Ce sera sensiblement la même chose pour la Table 4.

** Voir si les objets de la table 3 sont dans la table 4*

Dans les deux tables, les identificateurs sont sur les colonnes 9-15

```
acut -c9-15 table3.dat|sort|uniq > list_idt3
```

```
acut -c9-15 table4.dat|sort|uniq > list_idt4
```

vimdiff list_idt3 list_idt4 permet de comparer les identificateurs présents dans une table et pas l'autre.

Il faut trier les identificateurs pour que cela soit plus clair (d'où la commande « **sort** ») et pour comparer les identificateurs, on peut faire un « **uniq** » pour avoir le nombre exact d'identificateur uniques sans les doublons.

Le **wc -l** vous indique alors que le nombre de lignes pour list_idt4 est 43 et pas 54 comme dans la table.

En l'occurrence, il y a des objets en commun mais ce ne sont pas exactement les mêmes d'une table à l'autre.

III. Tables 1 et 2 de 2019MNRAS.483.4160 – Une méthode possible

Table 1.

A-type stars from the FORS2 observations.

ID	RA (J2000.0)	Dec. (J2000.0)	<i>g</i>	V_{GSR}	<i>D</i>	<i>EW</i> Ca	Class	Group
	(deg)	(deg)	(mag)	(km s^{-1})	(kpc)	(Å)		
S1 01	69.47791	-41.67166	19.12	-85.7 ± 6.1	48 ± 2	0.4 ± 0.2	BHB	MCs-M1
S1 02	75.92958	-44.00163	19.02	17.9 ± 4.7	50 ± 2	0.7 ± 0.2	BHB	MCs-M1
S1 03	75.24083	-40.96847	19.30	-42.5 ± 5.2	53 ± 2	0.3 ± 0.1	BHB	MCs-M1
S1 16	98.09583	-55.92491	19.39	140.6 ± 8.5	55 ± 3/ 35 ± 8	1.2 ± 0.3	BHB/BS	MCs-M1
S1 28	85.51833	-56.58208	19.30	107.7 ± 7.8	53 ± 2/ 33 ± 8	1.8 ± 0.3	BHB/BS	MCs-M1
S1 30	85.39375	-54.35069	19.07	-55.0 ± 7.0	27 ± 6	0.5 ± 0.2	BS	MW
S1 32	82.94583	-56.78797	19.12	165.8 ± 6.8	53 ± 2/ 26 ± 6	1.7 ± 0.3	BHB/BS	MCs-M1
S1 38	77.15333	-55.72430	19.13	57.4 ± 5.4	52 ± 2	1.4 ± 0.4	BHB	MCs-M1
S1 43	73.76208	-55.52858	19.23	69.3 ± 6.4	53 ± 2	1.1 ± 0.5	BHB	MCs-M1
S1 50	83.41541	-51.45541	19.11	92.7 ± 5.6	52 ± 2	1.0 ± 0.4	BHB	MCs-M1
S1 57	84.92083	-47.85800	19.22	-148.2 ± 8.3	31 ± 7	0.8 ± 0.3	BS	MW
S1 63	75.74666	-49.33027	19.10	101.8 ± 5.0	52 ± 2	0.5 ± 0.2	BHB	MCs-M1

Très jolie table HTML mais pas très pratique pour nos usages...

A. Récupération Table 1

=> sélection à la souris de la version HTML pour copier/coller dans un fichier. Donne [table1.tsv](#) :

S1 01	69.477	91	-41.671	66	19.12	-85.7 ± 6.1	48 ± 2	0.4 ± 0.2	BHB	MCs-M1
S1 02	75.929	58	-44.001	63	19.02	17.9 ± 4.7	50 ± 2	0.7 ± 0.2	BHB	MCs-M1
S1 03	75.240	83	-40.968	47	19.30	-42.5 ± 5.2	53 ± 2	0.3 ± 0.1	BHB	MCs-M1
S1 16	98.095	83	-55.924	91	19.39	140.6 ± 8.5	55 ± 3/ 35 ± 8	1.2 ± 0.3	BHB/BS	MCs-M1
S1 28	85.518	33	-56.582	08	19.30	107.7 ± 7.8	53 ± 2/ 33 ± 8	1.8 ± 0.3	BHB/BS	MCs-M1
S1 30	85.393	75	-54.350	69	19.07	-55.0 ± 7.0	27 ± 6	0.5 ± 0.2	BS	MW
S1 32	82.945	83	-56.787	97	19.12	165.8 ± 6.8	53 ± 2/ 26 ± 6	1.7 ± 0.3	BHB/BS	MCs-M1
S1 38	77.153	33	-55.724	30	19.13	57.4 ± 5.4	52 ± 2	1.4 ± 0.4	BHB	MCs-M1
S1 43	73.762	08	-55.528	58	19.23	69.3 ± 6.4	53 ± 2	1.1 ± 0.5	BHB	MCs-M1
S1 50	83.415	41	-51.455	41	19.11	92.7 ± 5.6	52 ± 2	1.0 ± 0.4	BHB	MCs-M1
S1 57	84.920	83	-47.858	00	19.22	-148.2 ± 8.3	31 ± 7	0.8 ± 0.3	BS	MW
S1 63	75.746	66	-49.330	27	19.10	101.8 ± 5.0	52 ± 2	0.5 ± 0.2	BHB	MCs-M1
S2 01	30.750	12	-57.107	19	19.49	99.4 ± 8.2	57 ± 3	0.4 ± 0.3	BHB	MCs-M2
S2 02	32.343	49	-55.391	11	19.30	-116.7 ± 7.9	28 ± 6	1.3 ± 0.3	BS	MW
S2 04	34.099	58	-53.908	25	19.28	164.6 ± 7.3	21 ± 5	2.2 ± 0.3	BS	MW
S2 05	35.289	41	-53.591	27	19.20	29.6 ± 10.5	23 ± 5	1.4 ± 0.2	BS	MW
S2 08	32.283	66	-53.117	11	19.19	99.6 ± 5.5	55 ± 3	1.3 ± 0.6	BHB	MCs-M2
S2 09	33.148	50	-52.254	47	19.25	116.5 ± 5.5	56 ± 3	0.3 ± 0.2	BHB	MCs-M2
S2 10	30.539	04	-51.972	69	19.22	-92.4 ± 7.0	22 ± 5	1.1 ± 0.3	BS	MW
S2 12	31.980	20	-50.742	77	19.15	96.9 ± 6.3	56 ± 3	0.8 ± 0.2	BHB	MCs-M2
S2 13	31.937	20	-50.548	33	18.99	11.6 ± 6.7	18 ± 4	2.5 ± 0.3	BS	MW
S2 14	30.024	99	-49.506	63	19.50	-25.2 ± 5.2	52 ± 2	0.5 ± 0.2	BHB	MCs-M1
S2 15	33.943	95	-50.195	63	19.34	-22.4 ± 5.5	23 ± 5	0.7 ± 0.2	BS	MW
S2 17	31.462	58	-46.860	97	19.07	76.7 ± 8.7	21 ± 5	0.2 ± 0.5	BS	MW
S2 18	30.237	41	-46.615	47	19.19	157.9 ± 6.3	26 ± 6	0.9 ± 0.3	BS	MW
S2 19	31.680	04	-46.010	44	19.10	24.9 ± 6.7	21 ± 5	2.2 ± 0.3	BS	MW
S2 20	31.359	33	-45.651	91	19.42	82.8 ± 6.4	35 ± 8	0.6 ± 0.2	BS	MCs-M2
S2 21	32.336	58	-44.162	47	19.14	-76.8 ± 6.6	22 ± 5	0.6 ± 0.2	BS	MW
S2 25	32.249	87	-42.476	61	19.30	-101.8 ± 5.6	21 ± 5	0.6 ± 0.2	BS	MW
S2 27	33.658	83	-40.943	58	19.07	-39.7 ± 6.1	20 ± 5	1.7 ± 0.3	BS	MW
S2 28	30.146	99	-55.768	55	19.07	-87.7 ± 6.4	28 ± 6	0.7 ± 0.2	BS	MW
S2 29	36.179	20	-52.505	97	19.19	61.3 ± 4.4	53 ± 2	1.0 ± 0.3	BHB	MCs-M2
S2 30	33.879	25	-51.799	97	19.17	-9.4 ± 6.4	30 ± 7	0.2 ± 0.2	BS	MW
S2 31	30.082	87	-50.651	88	19.02	-56.1 ± 6.3	21 ± 5	1.1 ± 0.2	BS	MW
S2 32	32.859	58	-50.578	58	18.93	16.6 ± 5.2	51 ± 2/ 18 ± 4	1.5 ± 0.2	BHB/BS	MCs-M2

B. Transformation en colonnes séparées facilement utilisables par un script

Commande file (man file) :

La commande **file table1.tsv** informe que le fichier obtenu est en UTF-8 Unicode text. Ce n'est pas terrible pour les commandes de substitutions/transformations que l'on voudrait faire dans le fichier. Il faut donc commencer par obtenir un fichier ASCII.

Première solution :

* **transformer les signes « - » par des « - » au format ASCII...** ATTENTION, il est important de le faire avant toute chose car sinon on risque d'avoir des surprises lors de certaines transformations (ça se voit assez vite).

`:%s/-/ /g`

(Le premier signe « - » est pris à la souris). Ne pas oublier « g » pour faire la substitution sur toute la ligne.

* **transformer les ± en \t** (= tabulation ; la tabulation permet d'avoir un séparateur de champs identique à celui utilisé pour les autres colonnes).

Attention : à ne pas faire si il y a à la fois des ± et des erreurs positives et négatives ou nulles (sinon on va se retrouver avec un nombre de champs différents ; ça se verra assez vite aussi mais c'est plus risqué). Ici, je le fais pour Vgsr :

`:%s/ ± /\t/`

En prenant le signe ± dans la souris.

`:%s` modifie les premières occurrences rencontrées sur une ligne dans tout le fichier ; en l'occurrence, ma colonne Vgsr qui contient bien un ± sur toutes les lignes.

La même commande est faite une deuxième fois pour les premières mesures de distance.

Pour rappeler une commande (: **et flèche vers le haut**).

Par contre, je n'ai pas fait la commande `:%s/ ± /\t/g` qui substitue toute la ligne car je n'aurais pas le bon nombre de champs par rapport aux mesures simple/double sinon. Il reste donc quelques ± qui ne sont pas embêtants pour la méthode que je vais utiliser.

Commande anafile (anafile -help)

* utiliser **anafile** pour transformer la table en une fois

`anafile -d -ccg table1.test > prog_t1`

Récupérer la ligne de commande **acut** du fichier **prog_t1** :

`vi prog_t1`

`d1G et dG` pour supprimer les parties du fichier qui ne nous intéressent pas.

Commande globale :

```
fcat table1.tsv | acut -d \  
    -f1%-5j -i'|' -f2%-11j -i'|' -f3%-14j -i'|' -f4%5.2j \  
    -i'|' -f5%-8j -i'|' -f6%4.1j -i'|' -f7%-17j -i'|' -f8%-10j \  
    -i'|' -f9%-6j -i'|' -f10%-6j >| table1.dat
```

—

fcat appelle la table.

acut : commande acut que vous connaissez (**acut --help** pour les options) vous auriez pu transformer la table avec différentes commandes acut par petits bouts). -d => par défaut, le séparateur de champs est la tabulation.

-fNN = les différents champs (fields ou colonnes de la table)

4.1j => format sur 4 caractères numériques avec 1 décimale

5j => format sur 5 caractères alpha-numériques.

-i'|' => Mettre '|' comme séparateur de champs dans la table en sortie.

Pour exécuter cette commande

source prog_t1 qui donne la table1.dat :

S1 01 69.477 91 -41.671 66 19.12 -85.7 6.1 48 2	[0.4 ± 0.2 BHB	MCS-M1	S1 01 69.477 91 -41.671 66 19.12 -85.7 6.1 48 2	[0.4 ± 0.2 BHB	MCS-M1
S1 02 75.929 58 -44.001 63 19.02 17.9 4.7 50 2	[0.7 ± 0.2 BHB	MCS-M1	S1 02 75.929 58 -44.001 63 19.02 17.9 4.7 50 2	[0.7 ± 0.2 BHB	MCS-M1
S1 03 75.240 83 -40.968 47 19.30 -42.5 5.2 53 2	[0.3 ± 0.1 BHB	MCS-M1	S1 03 75.240 83 -40.968 47 19.30 -42.5 5.2 53 2	[0.3 ± 0.1 BHB	MCS-M1
S1 16 98.095 83 -55.924 91 19.39 140.6 8.5 55 3/ 35 ± 8	[1.2 ± 0.3 BHB/BS	MCS-M1	S1 16 98.095 83 -55.924 91 19.39 140.6 8.5 55 3/ 35 ± 8	[1.2 ± 0.3 BHB/BS	MCS-M1
S1 28 85.518 33 -56.582 08 19.30 107.7 7.8 53 2/ 33 ± 8	[1.8 ± 0.3 BHB/BS	MCS-M1	S1 28 85.518 33 -56.582 08 19.30 107.7 7.8 53 2/ 33 ± 8	[1.8 ± 0.3 BHB/BS	MCS-M1
S1 30 85.393 75 -54.350 69 19.07 -55.0 7.0 27 6	[0.5 ± 0.2 BS	MW	S1 30 85.393 75 -54.350 69 19.07 -55.0 7.0 27 6	[0.5 ± 0.2 BS	MW
S1 32 82.945 83 -56.787 97 19.12 165.8 6.8 53 2/ 26 ± 6	[1.7 ± 0.3 BHB/BS	MCS-M1	S1 32 82.945 83 -56.787 97 19.12 165.8 6.8 53 2/ 26 ± 6	[1.7 ± 0.3 BHB/BS	MCS-M1
S1 38 77.153 33 -55.724 30 19.13 57.4 5.4 52 2	[1.4 ± 0.4 BHB	MCS-M1	S1 38 77.153 33 -55.724 30 19.13 57.4 5.4 52 2	[1.4 ± 0.4 BHB	MCS-M1
S1 43 73.762 08 -55.528 58 19.23 69.3 6.4 53 2	[1.1 ± 0.5 BHB	MCS-M1	S1 43 73.762 08 -55.528 58 19.23 69.3 6.4 53 2	[1.1 ± 0.5 BHB	MCS-M1
S1 50 83.415 41 -51.455 41 19.11 92.7 5.6 52 2	[1.0 ± 0.4 BHB	MCS-M1	S1 50 83.415 41 -51.455 41 19.11 92.7 5.6 52 2	[1.0 ± 0.4 BHB	MCS-M1
S1 57 84.920 83 -47.858 00 19.22 -148.2 8.3 31 7	[0.8 ± 0.3 BS	MW	S1 57 84.920 83 -47.858 00 19.22 -148.2 8.3 31 7	[0.8 ± 0.3 BS	MW
S1 63 75.746 66 -49.330 27 19.10 101.8 5.0 52 2	[0.5 ± 0.2 BHB	MCS-M1	S1 63 75.746 66 -49.330 27 19.10 101.8 5.0 52 2	[0.5 ± 0.2 BHB	MCS-M1
S2 01 30.750 12 -57.107 19 19.49 99.4 8.2 57 3	[0.4 ± 0.3 BHB	MCS-M2	S2 01 30.750 12 -57.107 19 19.49 99.4 8.2 57 3	[0.4 ± 0.3 BHB	MCS-M2
S2 02 32.343 49 -55.391 11 19.30 -116.7 7.9 28 6	[1.3 ± 0.3 BS	MW	S2 02 32.343 49 -55.391 11 19.30 -116.7 7.9 28 6	[1.3 ± 0.3 BS	MW
S2 04 34.099 58 -53.908 25 19.28 164.6 7.3 21 5	[2.2 ± 0.3 BS	MW	S2 04 34.099 58 -53.908 25 19.28 164.6 7.3 21 5	[2.2 ± 0.3 BS	MW
S2 05 35.289 41 -53.591 27 19.20 29.6 10.5 23 5	[1.4 ± 0.2 BS	MW	S2 05 35.289 41 -53.591 27 19.20 29.6 10.5 23 5	[1.4 ± 0.2 BS	MW
S2 08 32.283 66 -53.117 11 19.19 99.6 5.5 55 3	[1.3 ± 0.6 BHB	MCS-M2	S2 08 32.283 66 -53.117 11 19.19 99.6 5.5 55 3	[1.3 ± 0.6 BHB	MCS-M2
S2 09 33.148 50 -52.254 47 19.25 116.5 5.5 56 3	[0.3 ± 0.2 BHB	MCS-M2	S2 09 33.148 50 -52.254 47 19.25 116.5 5.5 56 3	[0.3 ± 0.2 BHB	MCS-M2
S2 10 30.539 04 -51.972 69 19.22 -92.4 7.0 22 5	[1.1 ± 0.3 BS	MW	S2 10 30.539 04 -51.972 69 19.22 -92.4 7.0 22 5	[1.1 ± 0.3 BS	MW
S2 12 31.980 20 -50.742 77 19.15 96.9 6.3 56 3	[0.8 ± 0.2 BHB	MCS-M2	S2 12 31.980 20 -50.742 77 19.15 96.9 6.3 56 3	[0.8 ± 0.2 BHB	MCS-M2
S2 13 31.937 20 -50.548 33 18.99 11.6 6.7 18 4	[2.5 ± 0.3 BS	MW	S2 13 31.937 20 -50.548 33 18.99 11.6 6.7 18 4	[2.5 ± 0.3 BS	MW
S2 14 30.024 99 -49.506 63 19.50 -25.2 5.2 52 2	[0.5 ± 0.2 BHB	MCS-M1	S2 14 30.024 99 -49.506 63 19.50 -25.2 5.2 52 2	[0.5 ± 0.2 BHB	MCS-M1
S2 15 33.943 95 -50.195 63 19.34 -22.4 5.5 23 5	[0.7 ± 0.2 BS	MW	S2 15 33.943 95 -50.195 63 19.34 -22.4 5.5 23 5	[0.7 ± 0.2 BS	MW
S2 17 31.462 58 -46.860 97 19.07 76.7 8.7 21 5	[0.2 ± 0.5 BS	MW	S2 17 31.462 58 -46.860 97 19.07 76.7 8.7 21 5	[0.2 ± 0.5 BS	MW
S2 18 30.237 41 -46.615 47 19.19 157.9 6.3 26 6	[0.9 ± 0.3 BS	MW	S2 18 30.237 41 -46.615 47 19.19 157.9 6.3 26 6	[0.9 ± 0.3 BS	MW
S2 19 31.680 04 -46.010 44 19.10 24.9 6.7 21 5	[2.2 ± 0.3 BS	MW	S2 19 31.680 04 -46.010 44 19.10 24.9 6.7 21 5	[2.2 ± 0.3 BS	MW
S2 20 31.359 33 -45.651 91 19.42 82.8 6.4 35 8	[0.6 ± 0.2 BS	MCS-M2	S2 20 31.359 33 -45.651 91 19.42 82.8 6.4 35 8	[0.6 ± 0.2 BS	MCS-M2
S2 21 32.336 58 -44.162 47 19.14 -76.8 6.6 22 5	[0.6 ± 0.2 BS	MW	S2 21 32.336 58 -44.162 47 19.14 -76.8 6.6 22 5	[0.6 ± 0.2 BS	MW
S2 25 32.249 87 -42.476 61 19.30 -101.8 5.6 21 5	[0.6 ± 0.2 BS	MW	S2 25 32.249 87 -42.476 61 19.30 -101.8 5.6 21 5	[0.6 ± 0.2 BS	MW
S2 27 33.658 83 -40.943 58 19.07 -39.7 6.1 20 5	[1.7 ± 0.3 BS	MW	S2 27 33.658 83 -40.943 58 19.07 -39.7 6.1 20 5	[1.7 ± 0.3 BS	MW
S2 28 30.146 99 -55.768 55 19.07 -87.7 6.4 28 6	[0.7 ± 0.2 BS	MW	S2 28 30.146 99 -55.768 55 19.07 -87.7 6.4 28 6	[0.7 ± 0.2 BS	MW
S2 29 36.179 20 -52.585 97 19.19 61.3 4.4 53 2	[1.0 ± 0.3 BHB	MCS-M2	S2 29 36.179 20 -52.585 97 19.19 61.3 4.4 53 2	[1.0 ± 0.3 BHB	MCS-M2
S2 30 33.879 25 -51.799 97 19.17 -9.4 6.4 30 7	[0.2 ± 0.2 BS	MW	S2 30 33.879 25 -51.799 97 19.17 -9.4 6.4 30 7	[0.2 ± 0.2 BS	MW
S2 31 30.082 87 -50.651 88 19.02 -56.1 6.3 21 5	[1.1 ± 0.2 BS	MW	S2 31 30.082 87 -50.651 88 19.02 -56.1 6.3 21 5	[1.1 ± 0.2 BS	MW
S2 32 32.859 58 -50.578 58 18.93 16.6 5.2 51 2/ 18 ± 4	[1.5 ± 0.2 BHB/BS	MCS-M2	S2 32 32.859 58 -50.578 58 18.93 16.6 5.2 51 2/ 18 ± 4	[1.5 ± 0.2 BHB/BS	MCS-M2
	1,13	Haut		1,16-14	Haut

Et là, c'est le drame. On note qu'il y a un blanc en colonne 13 qui n'est pas un vrai blanc. Si on regarde le bas du fichier, après « 13 », on passe à « 16-14 » ce qui signifie que ce caractère non-ASCII va nous embêter.

D'ailleurs, si on essaye de faire :

trcol -t13 ' ' |, on obtient un truc bizarre...

S1 01 69.477 <80><89>91 -41.671<80><89>66 19.12 -85.7 6.1 48 2	[0.4 Â± 0.2 BHB	MCS-M1
S1 02 75.929 <80><89>58 -44.001<80><89>63 19.02 17.9 4.7 50 2	[0.7 Â± 0.2 BHB	MCS-M1
S1 03 75.240 <80><89>83 -40.968<80><89>47 19.30 -42.5 5.2 53 2	[0.3 Â± 0.1 BHB	MCS-M1
S1 16 98.095 <80><89>83 -55.924<80><89>91 19.39 140.6 8.5 55 3/ 35 Â± 8	[1.2 Â± 0.3 BHB/BS	MCS-M1
S1 28 85.518 <80><89>33 -56.582<80><89>08 19.30 107.7 7.8 53 2/ 33 Â± 8	[1.8 Â± 0.3 BHB/BS	MCS-M1
S1 30 85.393 75 -54.350 69 19.07 -55.0 7.0 27 6	[0.5 Â± 0.2 BS	MW
S1 32 82.945 <80><89>83 -56.787<80><89>97 19.12 165.8 6.8 53 2/ 26 Â± 6	[1.7 Â± 0.3 BHB/BS	MCS-M1
S1 38 77.153 33 -55.724<80><89>30 19.13 57.4 5.4 52 2	[1.4 Â± 0.4 BHB	MCS-M1
S1 43 73.762 <80><89>08 -55.528<80><89>58 19.23 69.3 6.4 53 2	[1.1 Â± 0.5 BHB	MCS-M1
S1 50 83.415 <80><89>41 -51.455<80><89>41 19.11 92.7 5.6 52 2	[1.0 Â± 0.4 BHB	MCS-M1
S1 57 84.920 <80><89>83 -47.858<80><89>00 19.22 -148.2 8.3 31 7	[0.8 Â± 0.3 BS	MW
S1 63 75.746 <80><89>66 -49.330<80><89>27 19.10 101.8 5.0 52 2	[0.5 Â± 0.2 BHB	MCS-M1
S2 01 30.750 <80><89>12 -57.107<80><89>19 19.49 99.4 8.2 57 3	[0.4 Â± 0.3 BHB	MCS-M2
S2 02 32.343 49 -55.391<80><89>11 19.30 -116.7 7.9 28 6	[1.3 Â± 0.3 BS	MW
S2 04 34.099 <80><89>58 -53.908<80><89>25 19.28 164.6 7.3 21 5	[2.2 Â± 0.3 BS	MW
S2 05 35.289 <80><89>41 -53.591<80><89>27 19.20 29.6 10.5 23 5	[1.4 Â± 0.2 BS	MW
S2 08 32.283 <80><89>66 -53.117<80><89>11 19.19 99.6 5.5 55 3	[1.3 Â± 0.6 BHB	MCS-M2
S2 09 33.148 <80><89>50 -52.254<80><89>47 19.25 116.5 5.5 56 3	[0.3 Â± 0.2 BHB	MCS-M2
S2 10 30.539 <80><89>04 -51.972<80><89>69 19.22 -92.4 7.0 22 5	[1.1 Â± 0.3 BS	MW
S2 12 31.980 <80><89>20 -50.742<80><89>77 19.15 96.9 6.3 56 3	[0.8 Â± 0.2 BHB	MCS-M2
S2 13 31.937 <80><89>20 -50.548<80><89>33 18.99 11.6 6.7 18 4	[2.5 Â± 0.3 BS	MW
S2 14 30.024 99 -49.506<80><89>63 19.50 -25.2 5.2 52 2	[0.5 Â± 0.2 BHB	MCS-M1
S2 15 33.943 95 -50.195<80><89>63 19.34 -22.4 5.5 23 5	[0.7 Â± 0.2 BS	MW
S2 17 31.462 58 -46.860<80><89>97 19.07 76.7 8.7 21 5	[0.2 Â± 0.5 BS	MW
S2 18 30.237 41 -46.615<80><89>47 19.19 157.9 6.3 26 6	[0.9 Â± 0.3 BS	MW
S2 19 31.680 04 -46.010<80><89>44 19.10 24.9 6.7 21 5	[2.2 Â± 0.3 BS	MW
S2 20 31.359 33 -45.651<80><89>91 19.42 82.8 6.4 35 8	[0.6 Â± 0.2 BS	MCS-M2
S2 21 32.336 58 -44.162<80><89>47 19.14 -76.8 6.6 22 5	[0.6 Â± 0.2 BS	MW
S2 25 32.249 87 -42.476<80><89>61 19.30 -101.8 5.6 21 5	[0.6 Â± 0.2 BS	MW
S2 27 33.658 83 -40.943<80><89>58 19.07 -39.7 6.1 20 5	[1.7 Â± 0.3 BS	MW
S2 28 30.146 99 -55.768<80><89>55 19.07 -87.7 6.4 28 6	[0.7 Â± 0.2 BS	MW
S2 29 36.179 20 -52.585<80><89>97 19.19 61.3 4.4 53 2	[1.0 Â± 0.3 BHB	MCS-M2
S2 30 33.879 25 -51.799<80><89>97 19.17 -9.4 6.4 30 7	[0.2 Â± 0.2 BS	MW
S2 31 30.082 87 -50.651<80><89>88 19.02 -56.1 6.3 21 5	[1.1 Â± 0.2 BS	MW
S2 32 32.859 58 -50.578<80><89>58 18.93 16.6 5.2 51 2/ 18 Â± 4	[1.5 Â± 0.2 BHB/BS	MCS-M2

Heureusement, avec « u », on revient à l'état précédent.

En faisant :

file **table1.dat**, on voit que le fichier est toujours en UTF-8 Unicode text.
On voulait de l'ASCII...

Commande iconv (man iconv)

La commande **iconv** est faite pour ça en fait .

iconv -f UTF-8 -t ASCII//TRANSLIT table1.dat -o table1.out donne un fichier **table1.out** en ASCII

On va maintenant pouvoir continuer avec nos commandes pour transformer la **table1.out** tranquillement.

Cette fois, la commande :

trcol -t51 ' ' |' donne la table :

```
S1 01|69.477 91|-41.671 66|19.12| -85.7| 6.1|48|2 | | |0.4 +- 0.2|BHB |MCS-M1
S1 02|75.929 58|-44.001 63|19.02| 17.9| 4.7|50|2 | | |0.7 +- 0.2|BHB |MCS-M1
S1 03|75.240 83|-40.968 47|19.30| -42.5| 5.2|53|2 | | |0.3 +- 0.1|BHB |MCS-M1
S1 16|98.095 83|-55.924 91|19.39| 140.6| 8.5|55|3|35 +- 8 |1.2 +- 0.3|BHB/BS|MCS-M1
S1 28|85.518 33|-56.582 08|19.30| 107.7| 7.8|53|2|33 +- 8 |1.8 +- 0.3|BHB/BS|MCS-M1
S1 30|85.393 75|-54.350 69|19.07| -55.0| 7.0|27|6 | | |0.5 +- 0.2|BS |MW
S1 32|82.945 83|-56.787 97|19.12| 165.8| 6.8|53|2|26 +- 6 |1.7 +- 0.3|BHB/BS|MCS-M1
S1 38|77.153 33|-55.724 30|19.13| 57.4| 5.4|52|2 | | |1.4 +- 0.4|BHB |MCS-M1
S1 43|73.762 08|-55.528 58|19.23| 69.3| 6.4|53|2 | | |1.1 +- 0.5|BHB |MCS-M1
S1 50|83.415 41|-51.455 41|19.11| 92.7| 5.6|52|2 | | |1.0 +- 0.4|BHB |MCS-M1
S1 57|84.920 83|-47.858 00|19.22|-148.2| 8.3|31|7 | | |0.8 +- 0.3|BS |MW
S1 63|75.746 66|-49.330 27|19.10| 101.8| 5.0|52|2 | | |0.5 +- 0.2|BHB |MCS-M1
S2 01|30.750 12|-57.107 19|19.49| 99.4| 8.2|57|3 | | |0.4 +- 0.3|BHB |MCS-M2
S2 02|32.343 49|-55.391 11|19.30|-116.7| 7.9|28|6 | | |1.3 +- 0.3|BS |MW
S2 04|34.099 58|-53.908 25|19.28| 164.6| 7.3|21|5 | | |2.2 +- 0.3|BS |MW
S2 05|35.289 41|-53.591 27|19.20| 29.6|10.5|23|5 | | |1.4 +- 0.2|BS |MW
S2 08|32.283 66|-53.117 11|19.19| 99.6| 5.5|55|3 | | |1.3 +- 0.6|BHB |MCS-M2
S2 09|33.148 50|-52.254 47|19.25| 116.5| 5.5|56|3 | | |0.3 +- 0.2|BHB |MCS-M2
S2 10|30.539 04|-51.972 69|19.22|-92.4| 7.0|22|5 | | |1.1 +- 0.3|BS |MW
S2 12|31.980 20|-50.742 77|19.15| 96.9| 6.3|56|3 | | |0.8 +- 0.2|BHB |MCS-M2
S2 13|31.937 20|-50.548 33|18.99| 11.6| 6.7|18|4 | | |2.5 +- 0.3|BS |MW
S2 14|30.024 99|-49.506 63|19.50|-25.2| 5.2|52|2 | | |0.5 +- 0.2|BHB |MCS-M1
S2 15|33.943 95|-50.195 63|19.34|-22.4| 5.5|23|5 | | |0.7 +- 0.2|BS |MW
S2 17|31.462 58|-46.860 97|19.07| 76.7| 8.7|21|5 | | |0.2 +- 0.5|BS |MW
S2 18|30.237 41|-46.615 47|19.19| 157.9| 6.3|26|6 | | |0.9 +- 0.3|BS |MW
S2 19|31.680 04|-46.010 44|19.10| 24.9| 6.7|21|5 | | |2.2 +- 0.3|BS |MW
S2 20|31.359 33|-45.651 91|19.42| 82.8| 6.4|35|8 | | |0.6 +- 0.2|BS |MCS-M2
S2 21|32.336 58|-44.162 47|19.14|-76.8| 6.6|22|5 | | |0.6 +- 0.2|BS |MW
S2 25|32.249 87|-42.476 61|19.30|-101.8| 5.6|21|5 | | |0.6 +- 0.2|BS |MW
S2 27|33.658 83|-40.943 58|19.07|-39.7| 6.1|20|5 | | |1.7 +- 0.3|BS |MW
S2 28|30.146 99|-55.768 55|19.07|-87.7| 6.4|28|6 | | |0.7 +- 0.2|BS |MW
S2 29|36.179 20|-52.505 97|19.19| 61.3| 4.4|53|2 | | |1.9 +- 0.3|BHB |MCS-M2
S2 30|33.879 25|-51.799 97|19.17|-9.4| 6.4|30|7 | | |0.2 +- 0.2|BS |MW
S2 31|30.082 87|-50.651 88|19.02|-56.1| 6.3|21|5 | | |1.1 +- 0.2|BS |MW
S2 32|32.859 58|-50.578 58|18.93| 16.6| 5.2|51|2|18 +- 4 |1.5 +- 0.2|BHB/BS|MCS-M2
:.,$!trcol -t51 ' ' |'
```

trcol -t50 '/' ' ' pour enlever les « / »

acut -c1-53 -i'|' -c58- et

.,,\$!acut -c1-12 -c14-23 -c25- (on enlève les colonnes 13 et 24 pour avoir des coo propres) et on obtient une table avec tous les champs, quasiment correctement alignée : **table1.dat**

```
S1 01|69.47791|-41.67166|19.12| -85.7| 6.1|48|2 | | |0.4|0.2|BHB |MCS-M1
S1 02|75.92958|-44.00163|19.02| 17.9| 4.7|50|2 | | |0.7|0.2|BHB |MCS-M1
S1 03|75.24083|-40.96847|19.30| -42.5| 5.2|53|2 | | |0.3|0.1|BHB |MCS-M1
S1 16|98.09583|-55.92491|19.39| 140.6| 8.5|55|3|35|8 |1.2|0.3|BHB/BS|MCS-M1
S1 28|85.51833|-56.58208|19.30| 107.7| 7.8|53|2|33|8 |1.8|0.3|BHB/BS|MCS-M1
S1 30|85.39375|-54.35069|19.07| -55.0| 7.0|27|6 | | |0.5|0.2|BS |MW
S1 32|82.94583|-56.78797|19.12| 165.8| 6.8|53|2|26|6 |1.7|0.3|BHB/BS|MCS-M1
S1 38|77.15333|-55.72430|19.13| 57.4| 5.4|52|2 | | |1.4|0.4|BHB |MCS-M1
S1 43|73.76208|-55.52858|19.23| 69.3| 6.4|53|2 | | |1.1|0.5|BHB |MCS-M1
S1 50|83.41541|-51.45541|19.11| 92.7| 5.6|52|2 | | |1.0|0.4|BHB |MCS-M1
S1 57|84.92083|-47.85800|19.22|-148.2| 8.3|31|7 | | |0.8|0.3|BS |MW
S1 63|75.74666|-49.33027|19.10| 101.8| 5.0|52|2 | | |0.5|0.2|BHB |MCS-M1
S2 01|30.75012|-57.10719|19.49| 99.4| 8.2|57|3 | | |0.4|0.3|BHB |MCS-M2
S2 02|32.34349|-55.39111|19.30|-116.7| 7.9|28|6 | | |1.3|0.3|BS |MW
S2 04|34.09958|-53.90825|19.28| 164.6| 7.3|21|5 | | |2.2|0.3|BS |MW
S2 05|35.28941|-53.59127|19.20| 29.6|10.5|23|5 | | |1.4|0.2|BS |MW
S2 08|32.28366|-53.11711|19.19| 99.6| 5.5|55|3 | | |1.3|0.6|BHB |MCS-M2
S2 09|33.14850|-52.25447|19.25| 116.5| 5.5|56|3 | | |0.3|0.2|BHB |MCS-M2
S2 10|30.53904|-51.97269|19.22|-92.4| 7.0|22|5 | | |1.1|0.3|BS |MW
S2 12|31.98020|-50.74277|19.15| 96.9| 6.3|56|3 | | |0.8|0.2|BHB |MCS-M2
S2 13|31.93720|-50.54833|18.99| 11.6| 6.7|18|4 | | |2.5|0.3|BS |MW
S2 14|30.02499|-49.50663|19.50|-25.2| 5.2|52|2 | | |0.5|0.2|BHB |MCS-M1
```

C. Quant à la Table 2.

S1 01	0.4 ± 0.2	BHB	-1.84 ± 0.41	
S1 02	0.7 ± 0.2	BHB	-1.73 ± 0.43	
S1 03	0.3 ± 0.1	BHB	-2.03 ± 0.37	
S1 16	1.2 ± 0.3	BHB/BS	-0.99 ± 0.59	/ -0.63 ± 0.42
S1 28	1.8 ± 0.3	BHB/BS	-0.56 ± 0.77	/ -0.15 ± 0.47
S1 32	1.7 ± 0.3	BHB/BS	-0.81 ± 0.72	/ -0.45 ± 0.45
S1 38	1.4 ± 0.4	BHB	-0.95 ± 0.65	
S1 43	1.1 ± 0.5	BHB	-1.17 ± 0.69	
S1 50	1.0 ± 0.4	BHB	-1.39 ± 0.63	
S1 63	0.5 ± 0.2	BHB	-1.95 ± 0.42	
S2 01	0.4 ± 0.3	BHB	-1.85 ± 0.49	
S2 08	1.3 ± 0.6	BHB	-1.20 ± 0.75	
S2 09	0.3 ± 0.2	BHB	-2.08 ± 0.39	
S2 12	0.8 ± 0.2	BHB	-2.00 ± 0.47	
S2 14	0.5 ± 0.2	BHB	-1.61 ± 0.45	
S2 20	0.6 ± 0.2	BS	-1.16 ± 0.35	
S2 29	1.0 ± 0.3	BHB	-1.32 ± 0.53	
S2 32	1.5 ± 0.2	BHB/BS	-1.35 ± 0.64	/ -1.08 ± 0.40
S2 34	0.5 ± 0.2	BHB	-1.97 ± 0.43	
S3 01	1.8 ± 0.3	BHB/BS	-0.57 ± 0.77	/ -0.17 ± 0.46
S3 02	0.5 ± 0.2	BHB	-1.93 ± 0.45	
S3 03	1.2 ± 0.2	BHB/BS	-1.41 ± 0.54	/ -1.12 ± 0.36
S3 05	0.5 ± 0.2	BHB	-1.79 ± 0.43	
S3 06	0.5 ± 0.2	BHB	-1.79 ± 0.43	
S3 09	0.9 ± 0.3	BS	-0.90 ± 0.43	
S3 10	1.4 ± 0.4	BS	-0.75 ± 0.51	
S3 16	1.0 ± 0.3	BS	-0.81 ± 0.43	
S3 19	0.9 ± 0.4	BHB/BS	-1.26 ± 0.64	/ -0.89 ± 0.52
S3 22	0.7 ± 0.2	BHB	-1.86 ± 0.47	
S3 24	0.9 ± 0.3	BHB/BS	-1.39 ± 0.57	/ -1.05 ± 0.44

Je recommande avec mon copier/coller qui n'était pas si mal. J'obtiens [table2-utf.tsv](#) :
 Cette fois-ci, je tente le `iconv` dès le départ.

`iconv -f UTF-8 -t ASCII//TRANSLIT table2-utf.tsv -o table2.out`

Oups, il faut remplacer les « - » par des « - » si on veut que la commande fonctionne... On obtient bien de l'ASCII.

`:set list`

```
S1 01 0.4 ± 0.2 BHB -1.84 ± 0.41 $
S1 02 0.7 ± 0.2 BHB -1.73 ± 0.43 $
S1 03 0.3 ± 0.1 BHB -2.03 ± 0.37 $
S1 16 1.2 ± 0.3 BHB/BS -0.99 ± 0.59 / -0.63 ± 0.42 $
S1 28 1.8 ± 0.3 BHB/BS -0.56 ± 0.77 / -0.15 ± 0.47 $
S1 32 1.7 ± 0.3 BHB/BS -0.81 ± 0.72 / -0.45 ± 0.45 $
S1 38 1.4 ± 0.4 BHB -0.95 ± 0.65 $
S1 43 1.1 ± 0.5 BHB -1.17 ± 0.69 $
S1 50 1.0 ± 0.4 BHB -1.39 ± 0.63 $
S1 63 0.5 ± 0.2 BHB -1.95 ± 0.42 $
S2 01 0.4 ± 0.3 BHB -1.85 ± 0.49 $
S2 08 1.3 ± 0.6 BHB -1.20 ± 0.75 $
S2 09 0.3 ± 0.2 BHB -2.08 ± 0.39 $
S2 12 0.8 ± 0.2 BHB -2.00 ± 0.47 $
S2 14 0.5 ± 0.2 BHB -1.61 ± 0.45 $
S2 20 0.6 ± 0.2 BS -1.16 ± 0.35 $
S2 29 1.0 ± 0.3 BHB -1.32 ± 0.53 $
S2 32 1.5 ± 0.2 BHB/BS -1.35 ± 0.64 / -1.08 ± 0.40 $
S2 34 0.5 ± 0.2 BHB -1.97 ± 0.43 $
S3 01 1.8 ± 0.3 BHB/BS -0.57 ± 0.77 / -0.17 ± 0.46 $
S3 02 0.5 ± 0.2 BHB -1.93 ± 0.45 $
S3 03 1.2 ± 0.2 BHB/BS -1.41 ± 0.54 / -1.12 ± 0.36 $
S3 05 0.5 ± 0.2 BHB -1.79 ± 0.43 $
S3 06 0.5 ± 0.2 BHB -1.79 ± 0.43 $
S3 09 0.9 ± 0.3 BS -0.90 ± 0.43 $
S3 10 1.4 ± 0.4 BS -0.75 ± 0.51 $
S3 16 1.0 ± 0.3 BS -0.81 ± 0.43 $
S3 19 0.9 ± 0.4 BHB/BS -1.26 ± 0.64 / -0.89 ± 0.52 $
S3 22 0.7 ± 0.2 BHB -1.86 ± 0.47 $
S3 24 0.9 ± 0.3 BHB/BS -1.39 ± 0.57 / -1.05 ± 0.44 $
S4 02 0.1 ± 0.2 BHB -2.01 ± 0.49 $
S4 04 0.7 ± 0.5 BS -1.57 ± 0.59 $
S4 06 1.8 ± 0.6 BS -0.77 ± 0.60 $
S4 12 2.5 ± 0.4 BS -0.29 ± 0.60 $
S4 16 0.8 ± 0.3 BS -0.90 ± 0.39 $
:set list
```

permet d'afficher les caractères de contrôle restants qui sont les tabulations (on veut bien les garder). En les remplaçant par un « | », on a nos champs. Sinon on utilise `anfile` pour réaligner tout correctement.

Commande join (man join)

Pour joindre les 2 tables, on peut faire un script awk ou bien utiliser la commande **join** grâce au premier champ qui est commun entre les deux tables. Cette commande ne fonctionne que si les champs de comparaison sont triés de la même manière dans les 2 tables. C'est le cas ici.

join -t '|' table1.dat table2.dat -a 1 > table.dat permet de joindre les deux tables dans une nouvelle.

Par défaut, la commande **join** prend la première colonne comme clef de comparaison entre les deux tables. Ça tombe bien.

Par contre, on utilise **-t** pour lui indiquer quel est le séparateur de champs dans notre table.

On indique avec l'option **-a** que l'on veut conserver toutes les lignes du premier fichier (table1.dat) même s'il n'y a pas de correspondance dans l'autre fichier.

IV. Table 1 de 2021AJ...162..276Z avec col. Class de Table 2

```
wget -c "https://iopscience.iop.org/1538-3881/162/6/276/suppdata/ajac1ce7t1_ascii.txt?doi=10.3847/1538-3881/ac1ce7" -O t1_ascii.txt
```

```
anafile -d -ccg t1_temp > prog_t1
```

J'obtiens la **table1.dat** :

```
J080544.13+113040.30|J0805+1130EN|121.43468|11.51171|1.50|3.3/10.8|19.60|1.4%|20110211|56.8
J0805+1130WS|121.43391|11.51120|2.33|18.13|1.2%|20060223|
J084809.69+351532.12|J0848+3515EN|132.04216|35.25953|1.47|5.7/6.3|16.16|0.0%|20030202|67.3
J0848+3515WS|132.04038|35.25892|1.91|16.18|0.0%|20110312|
J090714.61+520350.61|J0907+5203EN|136.81088|52.06406|2.06|7.4/8.5|17.83|0.0%|20010324|11.9
J0907+5203WS|136.81020|52.06206|2.34|16.93|0.0%|20010329|
J091448.94+085324.45|J0914+0853WN|138.70394|8.89013|1.83|3.8/9.4|18.35|0.2%|20111230|153.4
J0914+0853ES|138.70440|8.88920|1.58|18.83|0.1%|20031201|
J095559.35+395438.87|J0955+3954EN|148.99952|39.91302|1.85|10.1/9.7|16.52|0.0%|20030506|37.3
J0955+3954WS|148.99731|39.91079|2.31|16.61|0.0%|20120119|
J100602.50+071131.80|J1006+0711EN|151.51044|7.19217|2.28|5.6/12.5|16.88|0.0%|20030102|81.8
J1006+0711WS|151.50889|7.19195|4.06|16.98|2.1%|20030422|
J110639.56+433620.64|J1106+4336W|166.66297|43.60582|1.71|4.9/10.5|18.04|0.0%|20040226|93.4
J1106+4336E|166.66485|43.60574|1.67|17.33|0.0%|20040218|
J111519.98+542316.75|J1115+5423EN|168.83326|54.38799|1.94|8.8/11.8|16.33|0.0%|20030309|48.6
J1115+5423WS|168.83016|54.38639|1.97|17.64|0.0%|20030102|
J114411.74+102202.40|J1144+1022EN|176.04894|10.36735|2.17|6.5/14.7|17.69|0.0%|20030405|12.1
J1144+1022WS|176.04855|10.36558|2.08|18.48|0.0%|20120223|
J121418.25+293146.70|J1214+2931EN|183.57607|29.52964|1.45|7.9/9.3|16.08|0.0%|20060324|60.8
J1214+2931WS|183.57419|29.52873|1.50|16.87|0.0%|20130306|
J122217.85-000743.70|J1222-0007EN|185.57438|-0.12881|3.33|4.6/13.5|18.27|1.9%|20010322|80.5
J1222-0007WS|185.57308|-0.12903|2.29|18.73|0.2%|20010401|
J133031.98-003613.80|J1330-0036WN|202.63229|-0.60333|3.47|4.2/4.4|15.48|1.7%|20000428|116.9
J1330-0036ES|202.63329|-0.60383|2.64|16.93|5.0%|20010219|
J133817.27+481632.20|J1338+4816EN|204.57407|48.27806|2.35|10.2/5.6|14.48|0.0%|20030503|29.2
J1338+4816WS|204.57200|48.27561|2.10|14.43|0.0%|20090114|
J150134.72+544734.07|J1501+5447EN|225.39469|54.79280|1.87|3.8/10.7|18.62|0.3%|20020319|28.7
J1501+5447WS|225.39373|54.79179|1.78|18.23|0.2%|20020401|
J151751.77+252353.38|J1517+2523E|229.46573|25.39816|1.75|5.5/7.5|15.63|0.0%|20110412|96.8
J1517+2523W|229.46406|25.39834|2.29|19.15|0.2%|20070522|
J155344.31+302508.50|J1553+3025EN|238.43466|30.41904|1.13|3.5/13.1|18.30|0.0%|20050413|40.4
J1553+3025WS|238.43393|30.41830|1.82|21.12|1.1%|20110705|
J155850.44+272323.93|J1558+2723WN|239.70974|27.39106|1.85|4.1/7.2|19.22|1.6%|20110527|160.0
J1558+2723ES|239.71019|27.38998|2.43|16.11|0.1%|20030702|
J164507.91+205759.43|J1645+2057WN|251.28297|20.96651|1.50|4.2/9.7|18.11|0.0%|20040611|157.2
```

Si on veut répéter les noms de la première colonne :

```
.,,$!acut -c1-20B -c21-
```

Ajouter un pipe à la fin de la table (en vue de la suite) :

```
.,,$!acut -c1-92 -i'|'
```

Pour la table 2,

Dans **t2_ascii.txt** :

```
:g/^t/d pour supprimer les lignes blanches qui commencent par une tabulation.
```

anafile pour obtenir **table2.dat**

On a maintenant 2 tables de 40 lignes chacune avec les noms classés dans le même ordre.

Je peux sélectionner la colonne souhaitée de la table2.dat :

```
.,,$!acut -c100-
```

Et faire un paste de la table1 et de cette colonne – attention, il s’agit d’un collage recno à recno donc s’assurer avant que les lignes sont bien dans le même ordre dans les 2 tables !

```
paste table1.dat t2 > table12.dat
```

Dans la **table12.dat** obtenue, il y a la Table 1 formatée précédemment et la colonne voulue.

```
.,,$!expand pour enlever la tabulation de la table12.dat qui s’est ajoutée avec le paste.
```

Alternativement, on peut utiliser le **mode Visual Block** en faisant un **ctrl + V** pour :

sélectionner la colonne en Table 2 (avec les **flèches**), **y** pour copier

p en Table1 pour « pasteur » la colonne.

V. Table 1 de 2019MNRAS.488.1790L

(1) La récupération en copier/coller depuis le HTML ne donne pas grand-chose cette fois.

(2) Dans ce cas, il faut passer par le PDF mais un copier/coller depuis le PDF n'est pas terrible non plus...

Cela nécessite de rassembler les lignes. On peut le faire via une macro.

Créer une macro

q b pour lancer l'enregistrement de la macro « b »

@b pour rappeler la macro

20@b pour rappeler la macro vingt fois.

Macro : on se place en fin de ligne, on fait 2 x J pour joindre les lignes et on se place sur la ligne suivante.

Mais dès la 5^e ligne, 2 fois J n'est plus suffisant donc pas terrible, terrible... Il faut un peu de patience mais ça peut marcher.

(3) On peut récupérer le PDF en entier et utiliser :

La commande ps2ascii (man ps2ascii)

Dans certains cas, la commande **ps2ascii table.pdf > table.txt** peut nous sauver... Si on fait ps2ascii sur l'ensemble de l'article, il faut juste retrouver la table et la mettre en forme...

VI. Table A1 de 2020MNRAS.492.3728B

La table est en PDF uniquement et à la verticale.

Commencer par faire une rotation horaire dans le menu de droite pour la mettre dans le bon sens.
Pour voir si un copier/coller est possible.

On se rend compte que les signes « - » (encore eux !) bloquent la copie.

Mais en étant patient, on peut récupérer la table par petits bouts en copier-coller – je n’ai pas trouvé de meilleure solution...

`anafile -d' ' -ccg tablea1.txt > prog_ta1` permet de transformer la table en une fois.

VII. Substitutions par blocs : table 1 de 2022ApJS..258....8M

La table 1 est déjà alignée si on récupère la MRT. Ouf.

Par contre, il y a une colonne de noms Gaia, et une colonne de noms dans laquelle les noms Gaia ne sont pas répétés. Or, c’est cette colonne qui sert à faire les noms SIMBAD et les jointures entre tables...

Pour ajouter les noms Gaia lorsqu’ils manquent en conservant l’alignement de la table :

- faire de la place pour le nom Gaia complet afin de conserver l’alignement

`.,$,!acut -c1-44 -b5 -c45-`

- activer les expressions régulières

`:set magic`

- Remplacer le champ vide par Gaia DR2 + l’ID sans modifier les espaces.

`:%s^(^[0-9]\+)\(\ \+\)/1 Gaia DR2 \1\2`

30 espaces dans la première zone (soit le nombre de caractère qu’il faut pour écrire Gaia DR2 et 19 chiffres + 1 espace avant et après) et 2 espaces avant Gaia.

Liste des commandes utilisées

Commandes Unix & cds-pack :

wget	- récupération d'un fichier	(pages 2, 5, 12)
acut	- réaligner/transformer des colonnes	(pages 1, 3, 4, 7-9, 12)
trcol	- substitution sur des colonnes données	(pages 3, 4, 8, 9)
anafile	- pour transformer une table en 1 commande (entre autres)	(pages 3, 7, 12)
source	- pour exécuter l'ensemble des commandes d'un fichier	(page 3)
join	- pour joindre 2 tables avec un nombre de lignes différent	(pages 7, 11)
paste	- pour coller 2 tables qui ont le même ordre	(page 12)
file	- pour vérifier si on a de l'ASCII, de l'UTF-8, etc.	(pages 7, 9)
iconv	- pour convertir UTF-8 vers ASCII (par ex.)	(pages 9, 10)
ps2ascii	- pour convertir PDF vers ASCII	(page 13)
vimdiff	- pour comparer 2 fichiers avec VI	(page 4)
expand	- suppression des tabulations	(page 12)

Commandes VI :

:set list	- pour afficher les caractères de contrôle	(page 10)
:g	- pour appliquer une transformation sur un pattern	(pages 5, 12)
:s	- substitutions – dont substitutions par blocs	(pages 1, 3, 7, 14)
:set magic	- pour activer les expressions régulières	(pages 1, 5)
q + lettre	- Enregistrement d'une macro « lettre »	(page 13)
ctrl + v	- Passage en mode Visual Block	(page 12)

Raccourcis clavier :

* Supprimer du texte :

d1G	- Suppression ligne courante au début du fichier
dG	- Suppression ligne courante à la fin du fichier
d1j	- Suppression de la ligne courante et d'1 ligne en-dessous
d2j	- Suppression de la ligne courante et de 2 lignes au-dessous
d1k	- Suppression de la ligne courante et d'1 ligne au-dessus
x	- Suppression du caractère courant
D	- Suppression depuis le caractère courant jusqu'à la fin de la ligne
:10,20d	- Suppression des lignes 10 à 20

* Sauver et quitter :

:w	- enregistrer
:w file	- enregistrer avec le nom « file »
:q!	- quitter sans sauvegarder
:wq	- enregistrer et quitter

* Annuler une action et revenir en arrière :

u	- annule l'action précédente
ctrl + r	- annuler l'annulation
Flèche vers le haut	- Rappeler d'une commande (fonctionne aussi dans le shell)

* Copier, coller, insérer, modifier

- yy - copie la ligne courante (y copie toute la sélection en mode Visual Block)
- y4j - copie la ligne courante et les 4 lignes en-dessous
- y\$ - copie jusqu'à la fin de la ligne
- p - insérer après la ligne (à l'endroit du curseur s'il s'agit d'un bloc)
- o - insérer une ligne en-dessous de la ligne courante
- O - insérer une ligne au-dessus de la ligne courante
- r - Remplacer le caractère courant par ce qui suit
- J - Joindre deux lignes ensemble

* Macro automatique

- .
- Exécute l'ensemble des dernières commandes

Retrouver des exemples de commandes sur le Twiki :

<http://cloud-wiki.u-strasbg.fr/twiki/bin/view/Ressources/MAJVizier>

(Section I. sauf pour anafile qui est en Section II. – advanced stuff !, l'option -ccg est expliquée vers la fin)