Comparaison avec le fichier lu sur les cartes en 1985

Date : 13 octobre 2026.

Deux versions sont comparées :

- **Notre reconstitution** : lecture du listing imprimé (scan des pages 205 à 219), relectures comprises. Voir `etat_des_connaissances.md`.
- **Le fichier de 1985** : `ufo/sceau/poher/FIC2-Poher1976-JPRS-732cas_clean.txt`, présenté comme la sortie d'un lecteur de cartes perforées.

Ces deux versions viennent de deux canaux différents : l'encre du listing pour la nôtre, les trous des cartes pour celle de 1985. Un accord entre les deux est donc une confirmation forte.

1. Résultat en bref

- **731 cartes sont communes** aux deux versions. Sur ces cartes, **34 caractères diffèrent sur 58 480**, soit 0,058 %.
- **697 cartes sur 731 sont identiques** caractère pour caractère (95,3 %). Les 34 autres diffèrent par un seul caractère chacune.
- **Dans les 34 cas, c'est très probablement le fichier de 1985 qui a raison.** Je n'ai trouvé aucun écart où le scan contredit nettement la version de 1985.
- Un des 34 écarts révèle un caractère invalide sur la carte elle-même (`&`, carte 0314).
- Chaque version contient des cartes que l'autre n'a pas : 6 chez nous, 1 dans le fichier de 1985. On connaît donc **738 cartes** au total.

2. Le fichier de 1985

Propriété	Constat
Format	texte ASCII, fins de ligne Windows, 733 lignes de 80 caractères
Contenu	732 cartes de données et une ligne `A000` suivie de 76 zéros, en dernière position
Jeu de caractères	chiffres, majuscules, et un seul `&`
Doublons	aucun numéro d'ordre en double
Ordre	ni trié par numéro, ni par date, ni par source ; il suit en partie l'ordre du listing (110 ruptures de séquence)

Remarques :

- La ligne `A000…` ressemble à une carte de fin de paquet ou à un séparateur, pas à une observation. Je l'ai écartée.
- Le nom du fichier contient « JPRS » et « clean ». Cela peut indiquer un passage par une autre édition ou un nettoyage après lecture. Je n'ai aucun moyen de le vérifier ; la présence du `&` et de la ligne `A000` plaide plutôt pour une sortie brute.
- Le fichier ne contient que le fichier « mondial » : les 220 cartes du fichier français n'y sont pas répétées.

3. Cartes présentes d'un seul côté

Absentes du fichier de 1985 (6 cartes)

Carte	Dans le listing	Date	Lieu	Notre confiance
0001	p.213 ligne 2	1957-12-14	FU (Brésil / Amazonie)	0,986
0012	p.213 ligne 3	1964-05-18	IS (USA)	0,985
0014	p.213 ligne 4	1964-06-02	IQ (USA côte ouest)	0,997
0019	p.213 ligne 5	1964-06-24	IS (USA)	0,961
6004	p.216 ligne 14 et p.219	1954-09-10	JA (France nord)	0,9999
0137	p.218 ligne 47	1957-11-08	IA (France sud)	0,953

- Les cartes 0001, 0012, 0014 et 0019 sont quatre lignes consécutives, les premières de la partie « obs. à l'étranger – sources très crédibles ». Elles ont probablement été perdues ensemble (début d'un paquet).
- La carte 6004 (10 septembre 1954, atterrissage avec personnages, probablement Quarouble) est imprimée deux fois dans le listing : elle a bien existé.
- La carte 0137 est celle qui ne figurait déjà que dans le fichier français du listing.

Absente de notre listing (1 carte)

```
26585911081855970000JA000000050001 1ZZZZ022000000000000000000000000000000000000000000
```

- Carte 2658 : 11 août 1855, source 59 (Guieu), France nord, 1 disque rouge lumineux.
- Dans le fichier de 1985, elle précède immédiatement la carte 2659, qui est la toute première ligne de notre listing (p.205). Elle se trouvait donc sans doute en tête du paquet et n'a pas été imprimée, ou l'a été sur une page que nous n'avons pas.

Conséquence pour l'écart « 825 »

Aucune des deux versions n'atteint 825 : le listing a 736 cartes dans son fichier mondial, le fichier de 1985 en a 732, et leur réunion 738. L'écart avec le titre manuscrit reste inexpliqué, mais deux sources indépendantes s'accordent maintenant sur un fichier d'environ 735 cartes.

4. Les 34 écarts

Colonnes du tableau : « auto » est la lecture automatique avant relecture, « nous » la version finale, « conf. » notre confiance sur ce caractère, « impr. » le nombre d'impressions de la carte dans le listing.

Carte	Col.	Champ	Auto	Nous	1985	Conf.	Impr.	Listing	Relecture	Ce que montre le scan
0002	20	latitude	F	F	E	0.955	1	p.216 l.17	non relu	reste de barre basse visible : E
0200	20	latitude	F	F	E	0.955	1	p.213 l.36	non relu	reste de barre basse visible : E
0212	20	latitude	F	F	E	0.955	1	p.215 l.41	non relu	reste de barre basse visible : E
0247	20	latitude	F	F	E	0.988	1	p.214 l.40	non relu	le scan montre un F net ; barre basse du E non imprimée
0270	20	latitude	H	H	B	0.955	1	p.214 l.43	non relu	boucle haute fermée visible : B plausible (BU = Antarctique)
0298	20	latitude	F	F	E	0.955	1	p.215 l.54	non relu	reste de barre basse visible : E
0368	20	latitude	F	F	E	0.955	1	p.213 l.42	non relu	reste de barre basse visible : E
0545	20	latitude	F	F	E	0.988	1	p.206 l.29	non relu	le scan montre un F net ; barre basse du E non imprimée
0701	20	latitude	F	F	E	0.988	1	p.215 l.9	non relu	reste de barre basse visible : E
0702	20	latitude	F	F	E	0.955	1	p.216 l.29	non relu	reste de barre basse visible : E
2019	20	latitude	F	F	E	0.955	1	p.207 l.31	non relu	le scan montre un F net ; barre basse du E non imprimée
2614	20	latitude	F	F	E	0.955	1	p.208 l.51	non relu	le scan montre un F net ; barre basse du E non imprimée
2717	20	latitude	F	F	E	0.988	1	p.209 l.24	non relu	reste de barre basse visible : E
2727	20	latitude	F	F	E	0.955	1	p.209 l.34	non relu	reste de barre basse visible : E

Carte	Col.	Champ	Auto	Nous	1985	Conf.	Impr.	Listing	Relecture	Ce que montre le scan
2731	20	latitude	F	F	E	0.955	1	p.209 l.36	non relu	le scan montre un F net ; barre basse du E non imprimée
2742	20	latitude	F	F	E	0.988	1	p.209 l.45	non relu	le scan montre un F net ; barre basse du E non imprimée
4030	20	latitude	F	F	E	0.955	1	p.210 l.50	non relu	reste de barre basse visible : E
6056	20	latitude	F	F	E	0.955	1	p.215 l.24	non relu	reste de barre basse visible : E
0046	21	longitude	U	U	D	1.000	1	p.214 l.1	non relu	U avec des restes de trait en haut : D plausible
0144	21	longitude	U	U	Q	0.955	1	p.214 l.22	non relu	petite queue en bas à droite : Q plausible (IQ = côte ouest des USA)
0240	21	longitude	D	D	O	0.955	1	p.215 l.4	non relu	lettre ronde, nettement un O (HO = Hawaï) ; notre lecteur n'avait pas de gabarit O
0418	21	longitude	R	R	B	1.000	1	p.206 l.21	non relu	le scan ressemble à un R (bas ouvert) ; B dont le bas est mal imprimé
2645	21	longitude	R	R	B	1.000	1	p.209 l.13	non relu	le scan ressemble à un R (bas ouvert) ; B dont le bas est mal imprimé
2572	22	nombre de témoins	8	8	B	0.970	1	p.208 l.16	relu par vous (confiance 2)	illisible sur le scan (8 ou B)
2705	22	nombre de témoins	B	8	B	0.970	1	p.208 l.4	relu par vous (confiance 2)	illisible sur le scan (8 ou B)
2784	22	nombre de témoins	B	8	B	0.700	1	p.210 l.9	relu par vous (confiance 1)	illisible sur le scan (8 ou B)
1512	24	âge	1	1	3	0.955	1	p.206 l.63	non relu	chiffre cassé, 3 plausible
0314	31	méthode	6	B	&	0.350	1	p.205 l.13	relu par vous (confiance 0)	le scan montre le signe « ¬ » : caractère hors codage réellement présent sur la carte
0017	34	forme	B	8	B	0.700	1	p.213 l.62	relu par vous (confiance 1)	illisible sur le scan (8 ou B)
0246	34	forme	G	G	C	0.700	1	p.216 l.24	relu par vous (confiance 1)	C net, avec des taches
0704	34	forme	G	G	C	0.962	2	p.211 l.4	non relu	C net sur les deux impressions
6028	34	forme	G	G	C	0.962	2	p.212 l.55	non relu	C net sur les deux impressions

Carte	Col.	Champ	Auto	Nous	1985	Conf.	Impr.	Listing	Relecture	Ce que montre le scan
0536	36	dimension	8	6	8	0.700	1	p.215 l.60	relu par vous (confiance 1)	caractère très dégradé
2783	38	dimension	8	8	B	0.991	2	p.210 l.8	non relu	illisible sur les deux impressions (8 ou B)

Par type de confusion

Nous → 1985	Nombre	Colonnes	Explication
F → E	17	20 (latitude)	l'imprimante ne marque presque pas la barre basse du E
8 → B	5	22, 34, 38	formes très proches une fois dégradées
G → C	3	34 (forme)	taches d'encre prises pour la barre du G
R → B	2	21 (longitude)	bas du B non imprimé
U → D, U → Q, D → O, H → B	4	20, 21	lettres rares, mal imprimées
1 → 3, 6 → 8	2	24, 36	chiffres cassés
B → &	1	31	caractère invalide sur la carte

Le cas des latitudes E et F

C'est notre seule erreur systématique : la moitié des écarts.

- Sur les cartes communes, nous lisions 40 `F` et 5 `E` ; le fichier de 1985 donne 23 `F` et 22 `E`.
- Sur 11 des 17 cartes, un reste de barre basse est visible sur le scan. Sur les 6 autres, le scan montre un `F` propre.
- La géographie tranche en faveur de 1985 : `EV` et `EU` (15-30°S) couvrent Rio, São Paulo, le Paraguay et le nord de l'Argentine, alors que `FV` et `FU` (0-15°S) désignent l'Amazonie et le Nordeste. Une vingtaine de cas brésiliens en zone équatoriale était peu vraisemblable.
- Notre lecteur attribuait à ces caractères une confiance de 0,955 à 0,988 : il hésitait, mais juste au-dessus du seuil de relecture (0,95). Aucun n'a donc été relu.

Le caractère `&` de la carte 0314

- Colonne 31 (méthode d'observation), carte du 4 juillet 1961.
- Le scan montre le signe « ¬ », que nous avions tous jugé illisible (votre hypothèse était `B`, confiance 0).
- Le lecteur de 1985 donne `&`, c'est-à-dire une perforation seule en ligne 12, sans chiffre. Les lettres A à I se codent par la ligne 12 plus un chiffre : il s'agit vraisemblablement d'une lettre dont le second trou manque (`A` ou `B`, les deux seules valides ici).
- La carte porte donc une erreur de perforation d'origine. Les deux versions sont fidèles, chacune à sa manière ; la valeur voulue reste inconnue.

5. Ce que la comparaison confirme

Points sur lesquels nous avions un doute et que le fichier de 1985 confirme :

Point	Notre version	1985
Carte 0412, col. 45 (jugée illisible, hypothèse `3`)	3	3
Carte 2759, latitude corrigée d'après le codage	B	B
Carte 0339, latitude corrigée d'après le codage	D	D
Carte 2638, jour 23 avec mois 00	présent	présent
Carte 0540, source 72 dans un bloc de source 58	présent	présent
Col. 22, code `Z` non documenté	102 cartes	101 cartes
Col. 46, code `8` non documenté	19 cartes	19 cartes
Col. 73, code `5` non documenté	9 cartes	9 cartes
Col. 48	toujours 0	toujours 0
Couleur (col. 39-40)	jamais au-delà de 15	jamais au-delà de 15
Col. 26 renseignée sur 148 cartes	oui	oui (148)

Les incohérences de plausibilité (personnages sans débarquement, atterrissage sans trajectoire, grands nombres d'objets) sont identiques dans les deux versions : elles sont bien sur les cartes.

6. Bilan de notre méthode

Lecture automatique et relecture

Étape	Écarts avec 1985 sur les 731 cartes
Lecture automatique seule	72
Après relecture (vous et Claude) et contrôles de cohérence	34

Les 147 caractères relus

Le tableau compte les lectures justes d'après le fichier de 1985. Les 147 caractères relus sont tous sur des cartes communes.

Lecteur	Niveau	Justes / total
Lecture automatique	—	104 / 147
Claude	sûr	40 / 40
Claude	moyen	45 / 53
Claude	incertain	36 / 54
Vous	3 très confiant	27 / 27
Vous	2 confiant	75 / 77
Vous	1 peu confiant	37 / 41
Vous	0 illisible	1 / 2

- Vos lectures sont justes 140 fois sur 147, celles de Claude 121 fois, la machine 104 fois.
- Vos niveaux de confiance sont bien ordonnés : aucune erreur au niveau 3, 2 au niveau 2, 4 au niveau 1.
- Sept caractères relus restent faux : cinq confusions 8/B ou G/C où vous aviez choisi contre la machine ou contre Claude avec une confiance de 1 ou 2, un 6/8, et le `&`.
- Vous avez corrigé 16 caractères que la machine et Claude lisaient mal tous les deux.

Calibration de nos confiances

Confiance du caractère	Caractères	Écarts observés	Écarts attendus
< 0,50	2	1	1,3
0,50 à 0,80	17	4	5,1

Confiance du caractère	Caractères	Écarts observés	Écarts attendus
0,80 à 0,95	33	0	4,3
0,95 à 0,99	581	25	14,7
0,99 à 0,999	70	1	0,4
≥ 0,999	57 777	3	1,2

Confiance de la carte	Cartes	Identiques	Attendu
< 0,50	2	50 %	32 %
0,50 à 0,80	24	79 %	70 %
0,80 à 0,95	106	88 %	91 %
0,95 à 0,99	262	95 %	97 %
≥ 0,99	337	99 %	100 %

- Nous attendions 705 cartes identiques sur 731 ; il y en a 697. Nos confiances étaient donc un peu optimistes, mais du bon ordre de grandeur.
- Le défaut est concentré dans la tranche 0,95 à 0,99 : 25 écarts pour 15 attendus, presque tous des E/F. La calibration avait été faite sur les cartes imprimées deux fois, qui sont surtout françaises et ne contiennent presque aucune latitude `E` : elle ne pouvait pas voir ce biais.
- Les cartes imprimées deux fois sont presque parfaites : 215 identiques sur 218. Les cartes imprimées une fois le sont moins : 482 sur 513. C'est ce que nous avions prévu.
- Trois caractères lus avec une confiance supérieure à 0,999 sont faux (longitudes des cartes 0046, 0418 et 2645).
- Rétrospectivement, un seuil de relecture à 0,99 au lieu de 0,95 aurait fait relire environ 580 caractères de plus et aurait couvert 25 des 27 écarts non relus.

Les craintes sur 8 et B

Sur les colonnes 22, 34, 38, 43 et 45, nous avons 152 caractères `8` ou `B` : 147 sont justes, 5 sont des `B` lus `8`. Aucun `8` n'a été lu `B`.

7. Laquelle est la meilleure version ?

Mon avis : le fichier de 1985 est le meilleur texte pour les 731 cartes communes, et notre reconstitution le complète.

Arguments pour 1985 :

- un lecteur de cartes lit des trous, et les confusions qui nous ont posé problème (E/F, 8/B, C/G, B/R) correspondent à des perforations sans rapport entre elles ;
- les 34 écarts s'expliquent tous par un défaut d'impression ou de lecture de notre côté ;
- la géographie confirme les 17 latitudes `E`.

Réserves :

- la provenance exacte du fichier n'est pas documentée (« JPRS », « clean ») ;
- un lecteur de cartes peut aussi se tromper (carte abîmée, colonne mal lue) ; je n'en vois aucun signe ici, mais sur les 6 latitudes où le scan montre un `F` net et les 2 longitudes où il montre un `R`, la preuve repose sur le seul fichier de 1985 ;
- il lui manque 6 cartes.

Version mixte proposée

Élément	Source
731 cartes communes	fichier de 1985 (les 34 écarts tranchés en sa faveur)
Carte 0314, col. 31	garder `&` et signaler le caractère comme invalide sur la carte
Cartes 0001, 0012, 0014, 0019, 6004, 0137	notre reconstitution, avec leur confiance
Carte 2658	fichier de 1985
Ligne `A000`	écartée

Cela donnerait 738 cartes. Pour les 731 communes, la confiance deviendrait très élevée là où les deux versions s'accordent (58 446 caractères lus identiquement par deux canaux indépendants).

Décision (13 octobre 2026) : ce mix est la nouvelle version de référence.

- `resultats/listing_reference.txt` : 738 cartes, 80 colonnes, dans l'ordre du fichier de 1985 ; les 6 cartes ajoutées sont insérées à la place qu'elles occupent dans le listing imprimé (0001, 0012, 0014, 0019 en tête, 6004 au rang 218), la carte 0137 à la fin.
- `resultats/listing_reference.csv` : les mêmes cartes avec leur fonds et une note.

- `resultats/fichier_poher_reference.html` : version décodée ; une astérisque signale les 6 fiches du fonds Pierre Deransart (listing imprimé), les autres étant du fonds Jean-Pierre Rospars (lecture des cartes en 1985).
- Générateurs : `analyse/build_reference.py` et `analyse/gen_html_reference.py`.

Les anciens fichiers de `resultats/` (`listing_reconstitue.txt`, `lignes.csv`, `fichier_poher_decode.html`…) restent tels quels : ils décrivent notre seule reconstitution et ne sont plus la référence.

8. Fichiers

Fichier	Contenu
`comparaison_fichier_1985.md`	ce rapport
`analyse/compare_1985b.py`	calcul de la comparaison
`analyse/work/cmp1985b.json`	liste des écarts et calibration
`analyse/work/cmp_0.png`, `cmp_1.png`	agrandissements des 34 caractères en désaccord