



Séminaire du LaDHUL

29 avril 2014

Pierre-Yves Brandt, « Constitution d'une base de données de dessins de dieux réalisés par des enfants »

Lors de cette séance, Pierre-Yves Brandt (UNIL, ISSRC) nous présente le projet « Dessins de dieux » : ses origines, son protocole de récolte de données, ses analyses en cours ainsi que ses chantiers futurs.

Origines

Une double formation en psychologie du développement et en théologie n'est certainement pas étrangère à l'intérêt de Pierre-Yves Brandt pour l'étude du développement religieux chez l'enfant. Ce thème de recherche a une histoire, des praticiens (notamment Goldman, 1964 ; Fowler, 1981 ; Nye et Carlson, 1984 ; Oser, Gmünder et Ridez, 1991) et différentes techniques d'enquête, l'une d'elle étant la récolte et l'analyse de dessins. Cette technique fut développée pour la première fois par Ernest Harms (1941) qui récolta plus de 4000 dessins d'enfants représentant « l'être le plus grand que tu peux imaginer ». Cette étude, très intéressante, propose une catégorisation des dessins mais, manquant d'illustrations substantielles, offre peu de moyens pour se représenter les dessins sur la base desquels les catégories ont été construites. Il faut donc faire confiance à Harms et la façon dont il construit ses catégories à partir de dessins *qui ne sont pas accessibles au lecteur*. En effet, le format de publication de l'époque ne tolère que la forme discursive et ne supporte pas – de fait – les illustrations couleurs sur lesquelles repose pourtant l'analyse. Le lecteur n'a donc aucune prise et il est *contraint* de suivre Harms dans le façonnage de ses catégories psychologiques. Cette impossibilité d'accéder aux matériaux de base est également *délétère pédagogiquement* car l'enseignant qui parle des travaux de Harms est forcé de traiter d'analyses de dessins *alors même qu'il ne peut pas les montrer*. Evidemment, si l'on est convaincu de la pertinence de ce mode d'enquête et que l'on considère les conclusions de Harms comme satisfaisantes, ces deux problèmes peuvent tout à fait ne pas surgir. Mais lorsque les catégories proposées ne nous satisfont pas ; lorsque que l'on juge les conclusions de Harms *bancales* ; lorsque l'on se dit qu'on aurait certainement vu des choses différentes en analysant ces dessins dont l'accès nous est impossible, la question de reprendre à nouveaux frais l'analyse de dessins d'enfants prend tout son sens.

Le souhait de Pierre-Yves Brandt de pouvoir accéder directement à des dessins est relayé par



UNIL | Université de Lausanne

LADHUL - Laboratoire
de cultures et humanités
digitales de l'UNIL

une étudiante, Carole Herren, qui propose en 2004 de faire son travail de demi-licence en s'appuyant sur 20 dessins de dieux réalisés par des enfants de la région vaudoise. Le travail est encourageant et ses résultats concordent – avec toutes les limites que comporte un travail de demi-licence – avec les travaux de Hanisch (1996) qui montre que le taux de représentations anthropomorphiques de dieux tend à diminuer avec l'âge. Ce travail permet aussi à Pierre-Yves Brandt de montrer – enfin ! – des exemples de dessins aux étudiants de son cours sur le développement religieux chez l'enfant. Mais dans l'auditoire se trouve Yuko Kagata, une étudiante japonaise qui se pose des questions : « Au Japon, ça ne serait pas les mêmes dessins car les dieux ne sont pas *personnels* ; il y a le *Kami* de la montagne, de la rivière, de la lumière, etc. Les résultats provisoires de Carole Herren, qui montrent que le taux de représentations anthropomorphiques de dieux chez les enfants diminue avec l'âge, ne concernent certainement qu'une petite partie du globe : la vôtre. Au Japon, il n'est pas du tout certain qu'il y ait autant de représentations anthropomorphiques. Je fais même l'hypothèse qu'il sera très bas, même chez les enfants petits », dit-elle à Pierre-Yves Brandt. « Soit, ma fois, allez donc récolter des dessins au Japon », lui propose Pierre-Yves Brandt qui devient très vite son directeur de mémoire de DEA.

En 2005, surprise ! Après avoir récolté et analysé 142 dessins de dieux, le mémoire de Yuko Kagata montre que les dessins d'enfants japonais, relativement à leur âge, comportent davantage de représentations anthropomorphiques que ceux des enfants européens. Le thème semble terriblement intéressant et porteur de nouveautés : qu'en est-il des autres parties du globe ? De plus, contrairement aux années 1940, il est maintenant possible de *numériser* ces matériaux et de les rendre disponibles au regard des lecteurs et des étudiants en les déposant dans une base de données accessible par internet. Le problème lié à la construction privée de catégories problématiques voit là une tentative de réponse censée et relativement peu coûteuse. Mais la numérisation des dessins permettrait également d'extraire de nouveaux éléments intéressants en s'engageant – pourquoi pas, plus tard – dans des analyses automatisées d'images et de leur contenu. Oui, constituer une *base de données* qui rassemblerait ces dessins scannés permettrait 1) de rendre *consultables* les dessins, 2) de mieux *cadrer* l'arbitraire des interprétations et 3) de *tenter* des analyses automatisées.

En septembre 2009, Pierre-Yves Brandt et Yannick Meyer d'Unimédia étudient la possibilité de rendre *digitalement* disponible les dessins récoltés sans pour autant qu'ils puissent être exploités commercialement. Le système qui se met progressivement en place est une base de données *Filemaker* qui met à disposition les dessins selon deux niveaux d'autorisation. Le premier concerne toutes les personnes intéressées par la consultation des dessins mais ne permet pas de les télécharger en haute définition. Le deuxième niveau d'autorisation est dépendant d'une *clé d'accès* délivrée après contrôle des intentions mais autorise le téléchargement haute définition. Au niveau des fonctionnalités, la base de données peut classer les dessins en fonction de requêtes par pays, genre, âge, etc. en prenant appui sur la codification des fichiers TIFF (qui eux-mêmes ne sont pas stockés dans la base). Comme nous le verrons par la suite, cette architecture n'est pas suffisamment souple pour encaisser toutes les ambitions futures du projet « Dessins de dieux » qui va donc bientôt émigrer vers un système de type MySQL.

Toujours en 2009, parallèlement à la numérisation des dessins et à la construction de la base de données Filemaker, Pierre-Yves Brandt et son équipe poursuivent leur collecte, cette fois-ci en Suisse. À la suite d'une collaboration entre l'Université de Lausanne et l'Université de Saint-Petersbourg, Zhargalma Dandarova intègre l'équipe et parvient à récolter des dessins dans deux

régions de la Russie, à Saint-Pétersbourg et à Oulan-Oudé (Bouriatie), près du lac Baïkal. Grâce également à la collaboration de Camelia Soponaru Puzdriac qui a récolté des dessins en Roumanie, plus de 1200 numérisations sont aujourd'hui (début 2014) stockées dans les serveurs d'Unimédia et rendues accessibles sur l'Internet *via* la base de données Filemaker et les métadonnées sur lesquelles elle travaille.

Récolte des données

Le protocole de récolte des données a été convenu dès 2005 lors la collecte de dessins au Japon. C'est en fait un abus de langage de parler en nombre de *dessins* ; il faudrait davantage parler en nombre d'*enfants* puisque pour chaque enfant questionné, 3 documents sont récoltés. Le premier est évidemment le *dessin* couleur, de format A4, dont la consigne est : « Pourrais-tu dessiner dieu ? ». Au *verso* de ce document, seulement après avoir fait son dessin, l'enfant restitue la consigne afin de tester la façon dont il l'a interprétée (si l'enfant est trop jeune pour écrire, l'enquêteur restitue ses propos). En-dessous de la restitution de la consigne – sur le même document –, on demande à l'enfant de *décrire son dessin à un autre enfant* (imaginaire) qui se trouverait de l'autre côté de la terre. Le troisième document est un *questionnaire* adapté au contexte culturel de la récolte de données, rempli par l'enfant (ou l'enquêteur pour les enfants les plus jeunes) qui recense des informations sur l'enfant (âge, sexe, confession, nationalité, etc.) et le contexte de ses pratiques religieuses (p.ex. au Japon : y a-t-il a un autel chez lui ? est-il déjà allé dans un temple ?). **Ainsi donc, pour chacun des 1200 enfants questionnés, trois documents sont scannés** : 1) le dessin (en format TIFF), 2) le verso du dessin contenant la restitution de la consigne et la description du dessin (en format JPEG) et 3) la page de questionnaire (en format JPEG). Les documents sont ensuite stockés dans le serveur d'Unimédia et leur nom coïncide avec un code que la base de données Filemaker pourra interpréter pour répondre aux diverses requêtes effectuées depuis l'interface en ligne.. La base de données, en se basant sur les requêtes des utilisateurs, pourra ensuite *retrouver* les documents dans les serveurs d'Unimédia et les restituer selon certains critères d'ordonnancement¹.

Traduire le terme « dieu » en japonais n'est pas un exercice facile. Pour tenter de rendre les consignes (et donc les dessins) un maximum commensurables avec les pays à majorité chrétienne, il a été décidé d'utiliser le terme « Kami » qui désigne toute une pluralité de génies dans la tradition shinto. Evidemment, cette tradition diffère du monothéisme chrétien en ce qu'elle implique une pluralité de *Kami*. Pour autant, le terme « Kami » a l'avantage de n'être associé à aucune représentation typée et laisse donc une grande liberté de représentation du *divin*. A titre d'exemple, lorsque les chrétiens sont entrés en contact avec les Japonais, le terme inventé pour représenter le dieu unique chrétien fut « Kami-Sama » soit « Kami-personnifié »². L'échantillonnage des enfants s'est fait en fonction d'hypothèses propres au champ de la

¹ À noter que lors d'une utilisation sans clé d'accès, les requêtes ne peuvent concerner que les scans des dessins.

² Le suffixe « Sama » est utilisé après un nom propre pour dire « Monsieur ». « Mignolet-Sama » voudrait donc dire « Monsieur Mignolet ».

UNIL | Université de Lausanne

LADHUL - Laboratoire
de cultures et humanités
digitales de l'UNIL

psychologie du développement. Trois groupes d'âge ont ainsi été sélectionnés: 7-8 ans, 10-11 ans et 13-14 ans.

Base de données Filemaker, protocole de codage et feuilles Excel

Pour ce qui est de la base de données Filemaker, le codage s'applique au nom du fichier et comporte les indications suivantes : Nationalité+date_région_sexe_contexte(école publique ou bouddhiste, paroisse, etc.)_années_mois_prénom-document(recto, verso ou questionnaire). Par exemple, le nom de fichier du dessin de Marie, âgée de 8 ans et 10 mois, recueilli dans une paroisse du canton de Neuchâtel en 2010 sera codé : **ch10_ne_f_ps_08_10_mar-r**. On voit donc qu'à ce stade, la base de données ne peut pas traiter des informations relatives au *contenu* des dessins, descriptions et questionnaires. Il lui sera possible d' « extraire » tous les dessins faits par des garçons japonais âgés de 10 ans ou de lister les descriptions selon leur ordre chronologique. Par contre, elle ne pourra pas travailler quantitativement sur les couleurs, l'organisation des formes ou le contenu des questionnaires tout simplement parce que ces éléments n'ont pas été codés et qu'elle est donc incapable de les reconnaître. C'est en fait au logiciel *Excel* qu'est confié (pour l'instant) cette tâche d'analyse quantitative en prenant appui sur un certain nombre de **descripteurs**.

Les dessins possèdent des caractéristiques plus ou moins communes : certains ne contiennent qu'une figure, d'autre plusieurs ; certains représentent un animal, d'autres un être humain ou encore un individu possédant tous les attributs des anges ; certains contiennent une auréole, une barbe, un sceptre, etc., etc. Mais par la force des méninges (et en se basant sur les fichiers-descriptions en format JPEG), il est possible d'isoler certaines caractéristiques discriminantes et de les classer dans des fiches Excel relativement au nom de code des dessins auxquels ils sont rattachés. Ce travail aboutit à une fiche Excel relativement complexe qui recense tous les descripteurs (42 en tout) pour chaque dessin.

À ce stade, les données concernent plus de 1200 enfants qui ont à chaque fois produit trois documents : un dessin (recto), des descriptions (verso) et un questionnaire rempli par l'enfant ou l'enquêteur. Ces documents scannés disposent de noms de fichier permettant à la base de données Filemaker de les « extraire » selon plusieurs critères d'ordonnement. En ce qui concerne les fichiers-dessins, leur contenu est codé en vertu de 42 descripteurs classés dans des fiches Excel.

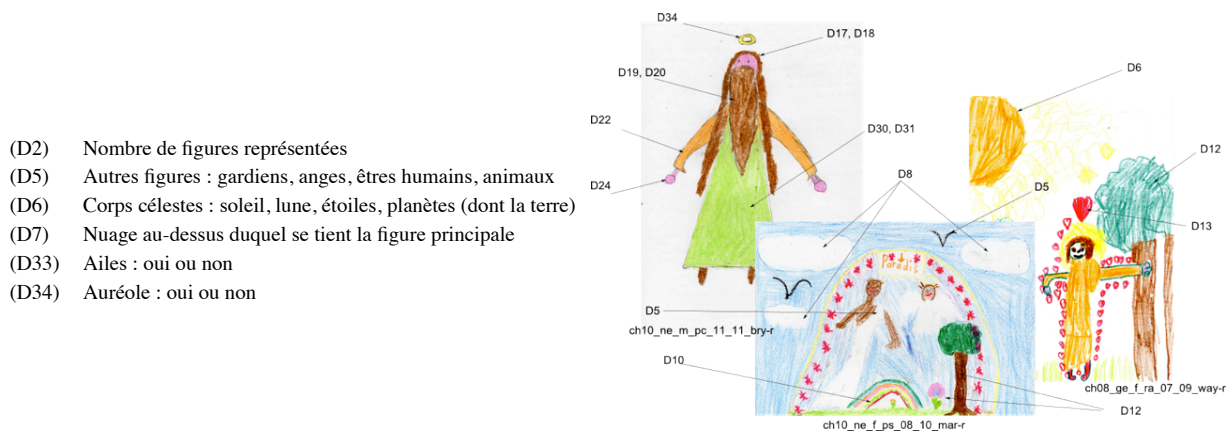


Figure 1: à gauche, quelques exemples de descripteurs (Dn). À droite, trois fichiers-dessins dont les contenus sont décrits en fonction de leurs descripteurs pertinents. Au pied de chaque dessin, leur nom de fichier codé.

Quelques exemples d'analyse

Certes, mais quoi faire de tout ça ? Calculer bien sûr ! C'est l'avantage indéniable de ce type de codage laborieux.

Représentations féminines et anthropomorphisme

Une première série d'analyses automatisées de contenu cherchent à comparer les dessins de dieux d'enfants japonais et européens aux résultats d'une vaste étude menée par Hanisch (1996) sur plus de 2500 dessins de dieux récoltés en Allemagne de l'Ouest (Heidenheim) et de l'Est (Leipzig). Un premier résultat intéressant montre une différence notable entre le taux de représentations féminines chez les filles japonaises *et le reste*. En effet, les dessins des enfants (garçons et filles) de Heidenheim et de Leipzig ont un taux de représentations féminines similaire de 6.5%. Ce taux est également très bas (4.5%) pour les garçons japonais. Par contre, le taux des filles japonaises est significativement plus élevé (48.2%) et indique **un curieux effet de genre**. Des analyses complémentaires ont ensuite été faites sur les dessins russes avec là-aussi des résultats intéressants : alors qu'à St-Petersbourg, le taux de représentations féminines est similaire à celui de l'Allemagne (6%), celui-ci grimpe jusqu'à 14% en Sibérie, à Oulan-Oudé.

Une deuxième série d'analyses vise à tester l'hypothèse d'un taux de représentations anthropomorphiques inférieur chez les enfants japonais que chez les enfants allemands du fait des influences de la tradition shinto et de ses divinités non personnelles.



Figure 2: exemples typiques d'une représentation anthropomorphique et non-anthropomorphique de dieu pour la collection japonaise.

Unil

UNIL | Université de Lausanne

LADHUL - Laboratoire
de cultures et humanités
digitales de l'UNIL

Or il n'en est rien ! Les tableaux croisés produits par Excel sont formels : même si le taux de représentations anthropomorphiques diminue avec l'âge (un tiers chez les enfants de 13-14 ans), il reste significativement haut chez les enfants japonais et même plus haut que chez les enfants suisses.

Age	Type de Kami		
	Anthropomorphique	Non anthropomorphique	Total
7-8 ans	45	4	49
10-11 ans	47	3	50
13-14 ans	31	12	43
Total	123	19	142

Chi 2 [2] = 11.33; p = 0.003

Tableau 1: distribution des représentations anthropomorphiques de dieux chez les enfants japonais en fonction de l'âge

Affiner l'analyse

Ces premières analyses sont encourageantes car elles produisent des résultats intéressants et significatifs. Mais elles font également émerger un certain nombre de problèmes relatifs à des catégorisations trop abruptes. Le cas des catégories « anthropomorphique »/ « non-anthropomorphique » est éclairant : quelle est la frontière entre les deux ? Prenons le cas du dessin d'une main : est-ce une représentation anthropomorphique ? Oui et non, puisqu'elle ne représente qu'une partie d'un individu. Plus la recherche avance, plus les cas limites apparaissent et incitent à creuser cette question de l'anthropomorphisme.

La comparaison des représentations de dieux issus d'écoles bouddhistes et non-bouddhistes dans la sous-collection japonaise était l'occasion de tenter un système de catégorisation plus fin. Il a ainsi fallu reconsidérer la fiche Excel et regrouper les dessins partageant un certain nombre de descripteurs afin de créer des catégories plus précises, appelées « **types** ».

Ligne	No. Sujet	Categorie	CodeSujet	Ecole	religion	sexe d'enfant	age			personnage	sexe de person.	barbe/moustache	couleur barbe	cheveux	couleur cheveu	Age	nombre de Kami	cadrage	Contact de Kami	décor	Contact marginal de décor	format	Manga	Z	AA	AB		
							ans	mois	décimal																			
L4	12	56	112	JFNB7,6HI	10	NB	NS	F	7	6	0.5	7.50	A	f	0	0	Chl	nr	je	1	en	K	Ce	D	D	L	F	Bd5
L5	13	67	112	JFB10,2MI	8	B	B	F	10	2	0.2	10.17	A	f	0	0	Chl	j	je	1	en	K	Islo	Dhdbg	D	L	F	Bd2
L6	14	120	112	JFNB7,6MA	10	NB	NS	F	7	6	0.5	7.50	A	f	0	0	Chl	nr	je	3	en	K	S	np	P	F	Bdca	
L7	15	2	121	JGB11,0NT	2	B	NS	G	11	0	0	11.00	A	m	1	bl	Chc	bl	vx	1	en	K	S	np	L	F	Bd3	
L8	16	3	121	JGNB14,4WS	6	NB	NS	G	14	4	0.3	14.33	A	m	1	bl	Chb	bl	vx	1	en	K	Cns	Dhdg	L	Fn	Bd1	
L9	17	9	121	JFNB8,9OM	10	NB	NS	F	8	9	0.8	8.75	A	m	1	bl	Chab	np	vx	1	en	K	Cn	D	P	Fn	Bd2	
L10	18	33	121	JGNB13,8KA	6	NB	NS	G	13	8	0.7	13.67	A	m	1	g	Chab	np	vx	1	en	Kb	Cn	D	P	Fn	Bd2	
L11	19	58	121	JGNB13,0SHU	6	NB	NS	G	13	0	0	13.00	A	m	1	bl	Chab	vx	1	en	K	S	np	L	F	Bd1		
L12	20	75	121	JFNB10,6MO	11	NB	S	F	10	6	0.5	10.50	A	m	1	g	Chc	g	vx	1	en	K	Cn	D	L	Fn	Bd1	
L13	21	82	121	JFNB13,1MH	12	NB	NS	F	13	11	0.9	13.92	A	m	1	bl	chl	j	vx	1	en	Khdg	Ca	D	L	F	Bd2	
L14	22	98	121	JFNB11,5IM	11	NB	NS	F	11	5	0.4	11.42	A	m	1	g	chl	j	vx	1	en	K	Ca2en	Dhdg	P	Fn	Bd4	
L15	23	99	121	JFNB11,4TS	11	NB	NS	F	11	4	0.3	11.33	A	m	1	bl	chab	g	vx	1	en	K	Ca2nr	Dhdg	L	Fn	Bd3	
L16	24	108	121	JGNB8,2HJ	4	NB	NS	G	8	2	0.2	8.17	A	m	1	bl	Chc	nr	vx	2	en	K	Cns	Dhd	P	Fn	Bd4	
L17	25	119	121	JFNB13,0RE	12	NB	NS	F	13	0	0	13.00	A	m	1	bl	chl	j	vx	1	en	K	S	np	L	F	Bd2	
L18	26	8	201	JGB8,2MN	1	B	NS	G	8	2	0.2	8.17	H	m	1	n	Chc	nr	vx	1	en	K	S	np	P	F	Bdca	
L19	27	22	201	JGB11,8KS	2	B	NS	G	11	8	0.7	11.67	H	m	0	0	Chab	vx	1	en	K	Cns	Dhd	L	F	Bd2		
L20	28	44	201	JFB10,8RE	8	B	B	F	10	8	0.7	10.67	H	m	1	g	Chl	g	vx	1	en	Kb	Ce	Dhdg	L	NF	Bd1	
L21	29	45	201	JFNB7,11YY	10	NB	NS	F	7	11	0.9	7.92	H	m	1	bl	Chab	np	vx	1	en	K	S	np	L	F	Bd1	
L22	30	47	201	JGNB13,1YO	6	NB	NS	G	13	6	0.5	13.50	H	m	1	bl	Chab	np	vx	1	en	K	S	np	L	F	Bd2	
L23	31	55	201	JGNB7,6HI	4	NB	NS	G	7	6	0.5	7.50	H	m	0	0	Chl	bl	vx	1	en	K	S	np	P	F	Bd3	
L24	32	61	201	JGB11,7KK	2	B	NS	G	11	7	0.6	11.58	H	m	1	nr	Chl	nr	vx	1	en	K	Cn	Dhdg	L	Fn	Bd2	
L25	33	64	201	JGNB14,2AT	6	NB	NS	G	14	2	0.2	14.17	H	m	1	nr	Chl	nr	vx	1	en	Kh	S	np	P	F	Bd2	
L26	34	69	201	JFB8,9YA	7	B	SB	G	11	8	0.7	11.67	H	m	1	g	Chc	g	vx	1	en	K	S	np	L	F	Bd2	
L27	35	87	201	JGNB11,8KT	2	B	B	F	8	9	0.8	8.75	H	m	1	g	Chc	n	vx	1	en	K	S	np	L	F	Bd4	
L28	36	90	201	JGNB11,9HT	5	NB	NS	G	11	9	0.8	11.75	H	m	1	g	Chab	np	vx	1	en	Kg	S	np	L	F	Bd4	
L29	37	95	201	JFB7,10AM	7	B	NS	F	7	10	0.8	7.83	H	m	1	bl	chab	vx	1	en	K	S	np	L	F	Bd2		
L30	38	101	201	JGB13,4YM	3	B	B	G	13	4	0.3	13.33	H	m	1	bl	chab	vx	1	en	K	Cn	D	L	Fn	Bd3		
L31	39	112	201	JFNB13,6RI	12	NB	NS	F	13	6	0.5	13.50	H	m	1	g	chl	g	vx	1	en	K	Cn	D	L	Fn	Bd1	
L32	40	129	201	JFNB8,5TH	10	NB	NS	F	8	5	0.4	8.42	H	m	1	bl	chl	g	vx	1	en	K	S	np	L	F	Bd2	
L33	41	130	201	JGB10,11OY	2	B	B	G	10	11	0.9	10.92	H	m	1	bl	chl	nr	vx	1	en	K	Cn	D	L	Fn	Bd8	
L34	42	136	201	JGNB8,3OH	4	NB	NS	G	8	3	0.3	8.25	H	m	1	g	chl	g	vx	1	en	K	Cn	D	P	Fn	Bd2	
L35	43	19	211	JGB8,1MT	1	B	NS	G	8	1	0.1	8.08	H	m	0	0	Chc	nr	vx	1	en	Khd	S	np	L	F	Bd3	
L36	44	86	211	JFB7,8YU	7	B	B	F	7	8	0.7	7.67	H	m	1	bl	chl	g	vx	1	en	K	Cchnst	Dhdg	L	Fn	Bd3	
L37	45	104	211	JGB9,4AY	1	B	NS	G	8	4	0.3	8.33	H	m	1	bl	chl	nr	vx	1	en	K	S	np	L	F	Bd3	
L38	46	105	211	JGB13,7MIK	3	B	NS	G	13	7	0.6	13.58	H	m	1	bl	chl	g	vx	1	en	Kb	S	np	L	NF	Bd1	
L39	47	83	220	JGB13,2YK	3	B	BC	G	13	2	0.2	13.17	H	ind	0	0	Chab	np	je	1	en	Kh	Cn	Dhdg	P	Fn	Bd1	
L40	48	13	221	JFB8,0CHI	7	B	B	F	8	0	0	8.00	H	m	0	0	Chc	nr	je	1	en	K	C	D	P	F	Bd2	
L41	49	28	221	JGNB11,9MS	5	NB	NS	G	11	9	0.8	11.75	H	m	0	0	Chl	n	je	1	en	Kb	Cn	Dhdg	P	Fn	Bd3	
L42	50	68	221	JGB8,2TA	1	B	NS	G	8	2	0.2	8.17	H	m	1	bl	Ch	nr	je	1	en	K	Cn	D	L	Fn	Bd2	
L43	51	93	221	JGNB7,5TA	4	NB	NS	G	7	5	0.4	7.42	H	m	0	0	chca	nr	je	1	en	K	Cn	Dhdg	L	Fn	Bd2	

Figure 3: illustration de la feuille Excel « augmentée » classant les différents dessins en fonction de « types » (appelés ici « catégories », à gauche). L'image représente le travail terminé mais le processus s'est déroulé comme suit: en coloriant les cases « descripteurs » (par exemple à la colonne O, « barbe/moustache ») en fonction de leurs contenus, des groupements ont peu à peu émergé. C'est à partir de ces groupements ayant en commun plusieurs cases de mêmes couleurs que se sont progressivement dessinés les différents « types ».



UNIL | Université de Lausanne

LADHUL - Laboratoire de cultures et humanités digitales de l'UNIL

Si la feuille Excel se complexifie jusqu'à la limite du raisonnable, elle permet tout de même à Brandt, Kagata, Spitteler et Gilléron (2009) d'identifier 17 types et sous-types de représentations de dieux pour leur collection japonaise :

9 types (17 types et sous-types)

- 900 lumière
- 800 relation / narration
- 700 entité non anthropomorphique
- 600 masquée
- 500 monstrueuse
- Bouddha traditionnel (401), modifié (411)
- Figure céleste (ange 100, 110, 121)
- Humain céleste 201, 211, 220)
- Figure terrestre (300, 310, 320, 330)

Figure 4: tiré de Brandt, Kagata, Spitteler et Gilléron (2009). À partir de descripteurs partagés, chaque « type » affine la distinction « anthropomorphique/non anthropomorphique ». Certains types sont uniques (lumière, masquée, monstrueuse), tandis que d'autres sont subdivisés en sous-types (figure céleste, humain céleste, Bouddha).

Certes, ces types sont dépendants de l'interprétation des contenus des dessins effectuée par Pierre-Yves Brandt et par son équipe. Mais ça ne les empêche pas d'être pertinents pour l'analyse qu'ils cherchent à produire. Car cette nouvelle catégorisation permet au logiciel Excel d'analyser quantitativement la distribution et de produire des tableaux croisés capables de comparer – de manière plus ou moins représentative – les dessins issus d'écoles bouddhistes et non-bouddhistes :

	Ecole bouddhiste			non bouddhiste			bouddhiste		non bouddhiste		
	8 ans	11 ans	14 ans	8 ans	11 ans	14 ans	♂	♀	♂	♀	
Type 100	1	-	-	-	2	1	-	1	-	3	4
Type 110	2	1	1	3	1	2	2	2	1	5	10
Type 120	-	1	-	2	3	5	1	-	4	6	11
Type 200	3	5	1	4	1	3	6	3	5	3	17
Type 210	3	-	1	-	-	-	3	1	-	-	4
Type 220	2	5	2	3	5	-	2	7	4	4	17
Type 300	1	1	-	1	1	2	1	1	-	4	6
Type 310	3	1	-	1	1	1	3	1	2	1	7
Type 320	2	2	1	2	1	1	2	3	2	2	9
Type 330	-	1	-	1	1	1	-	1	3	-	4
Type 400	3	4	2	-	2	-	6	3	1	1	11
Type 410	5	3	1	-	1	2	8	1	3	-	12
Type 500	3	2	1	-	1	-	6	-	1	-	7
Type 600	-	2	2	-	-	-	4	-	-	-	4
Type 700	1	-	2	1	-	-	3	-	1	-	4
Type 800	1	2	6	1	-	1	4	5	2	-	11
Type 900	-	-	3	-	-	1	2	1	-	1	4
	30	30	23	19	20	20	53	30	29	30	142

Ages et types (toutes écoles): $\tau_c = .170, z = 2.386, p = .009$
 Ecoles bouddhistes : $\tau_c = .263, z = 2.788, p = .003$
 Ecoles non bouddhistes : $\tau_c = .114, z = 1.009, p = .156$

Tableau 2: tableau croisé issu de Brandt et al. (2009). La moitié gauche du tableau croisé compare la distribution des types et sous-types entre école bouddhistes et non bouddhistes en fonction de l'âge; la moitié droite effectuée la même comparaison en fonction du sexe.

Là aussi, des résultats intéressants émergent. Il semble par exemple que les garçons de 14 ans issus d'écoles bouddhistes soient davantage enclins à se représenter des dieux en fonction des types 500 à 900 et que les filles d'écoles non bouddhistes tendent à se représenter des figures célestes selon les types 100 à 120.

Peu importe ici l'interprétation de ces résultats. L'important pour nous est de voir qu'en travaillant sur le contenu des données, certaines codifications sont abandonnées tandis que d'autres émergent. Le codage du contenu ne s'effectue donc pas *en amont* des analyses ; il se fait et se rattrape *au même moment*. Le codage du contenu apparaît donc comme un exercice périlleux : jusqu'à quel point ces « types » et « sous-types » sont-ils discriminants ? La combinaison des descripteurs suffit-elle à faire émerger des critères d'opposition pertinents ? En s'appuyant toujours sur le même référentiel de discrimination (les descripteurs), ne risquons-nous pas d'être entraînés malgré nous vers une direction d'analyse unique ? La méthode comporte donc des limites notables sur lesquels il est nécessaire de travailler.

Analyse par clustering

C'est dans cette optique que Pierre-Yves Brandt et son équipe se sont engagés dans une **tentative d'analyse de contenu par clustering automatisé**. À partir de l'ensemble de la collection de dessins suisses, Fournier, Aeschbach et Montandon (2011) ont extrait aléatoirement 202 dessins en veillant à respecter la distribution des différents âges, sexes et lieux de récolte. A partir de ce corpus, ils ont commencé par distinguer manuellement les dessins en fonction de deux grandes « classes » : figuratif et non figuratif. Toujours en prenant manuellement appui sur les *listings* des fiches Excel établis en fonction du codage des 42 descripteurs, ils ont ensuite distingué 5 « sous-classes » figuratives :

– 2 classes et 5 sous-classes

- 100 figuratif (n=169; 83.6%)
- 101 sans attributs
- 110 un ou plusieurs attributs « normaux »
- 120 un ou plusieurs attributs « célestes »
- 130 avec ailes
- 140 « Dieu crucifié »
- 200 non-figuratif (n=33; 16.4%)

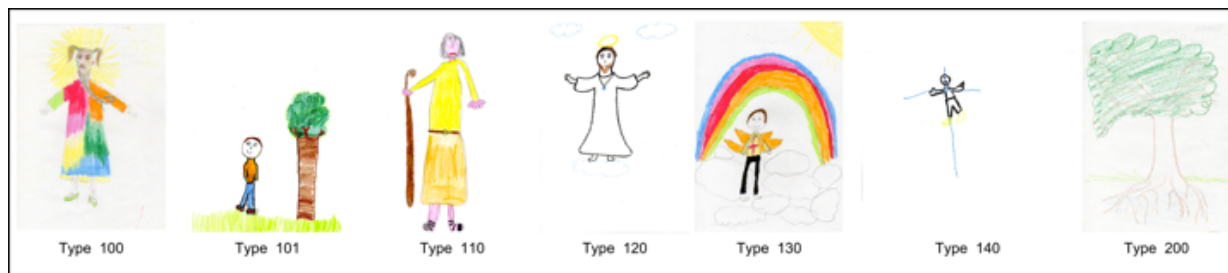


Figure 5: en-haut, récapitulatif des "classes" et "sous-classes". En bas, exemples typiques issus de l'échantillon (à noter que les termes "types" renvoient ici aux "classes" et "sous-classes" et non aux "types" précédemment rencontrés lors de l'étude sur les dessins japonais).

Unil

UNIL | Université de Lausanne

LADHUL - Laboratoire
de cultures et humanités
digitales de l'UNIL

Peu importe les détails de l'algorithme de *clustering* choisi. L'important est le résultat mitigé qui en résulte. En effet, en intégrant les descripteurs *et* les différentes classes et sous-classes, l'algorithme finit par produire 5 clusters dont les caractéristiques peinent à faire sens pour l'équipe de Pierre-Yves Brandt.

Le cluster qui se distingue le plus des autres est constitué de pages blanches : à n'en pas douter, l'élément central de ce cluster est celui de ne comporter aucun élément... Mais pour ce qui est des autres clusters, difficile d'interpréter les résultats de façon intéressante. En effet, il est très ardu d'extraire une quelconque similarité entre les dessins d'un même cluster tant ils rassemblent, au final, des caractéristiques observables très différentes. Dit autrement, en se basant sur les 42 descripteurs et les différentes classes et sous-classes, la méthode par cluster produit certes de nouveaux « paquets » de dessins mais ceux-ci sont non pertinents puisque trop différents lorsqu'on les observe séparément. La faute à qui ? Difficile à dire... La raison principale se trouve peut-être dans les 42 descripteurs qui ne parviennent pas à suffisamment distinguer les dessins : trop de descripteurs sont communs à trop de dessins et finissent en fin de compte par induire en erreur l'algorithme de cluster qui ne peut plus les distinguer. Il semble ainsi qu'il faille effectuer un nouveau travail sur les descripteurs afin de diminuer leur nombre tout en augmentant leur pouvoir discriminant. Il faut aussi en réinventer *de meilleurs* afin d'augmenter le pouvoir heuristique de la méthode par cluster.

Computational photography

Durant l'été 2013, Pierre-Yves Brandt contacte le laboratoire « Image and Visual Representation Group » (IVRG) de l'EPFL spécialisé dans le traitement d'images numériques. Car qui dit « scans de dessins », dit « pixels » ; et qui dit « pixels » dit « unités discrètes potentiellement traitables mathématiquement ». Cette collaboration pourrait être utile au projet « Dessins de dieux » car le savoir-faire de l'IVRG en matière de création d'algorithmes permettrait de dégager de nouveaux critères discriminants et non visibles à l'œil nu. Mais cette collaboration pourrait également être utile à l'IVRG car travailler sur un corpus de dessins scannés serait pour eux l'occasion de développer des outils de détection inédits. Sabine Süsstrunk – la directrice de l'IVRG – engage donc une doctorante intéressée par le challenge. Celle-ci commence bientôt à produire un certain nombre de prototypes de détection concernant notamment les centres de gravité, les palettes de couleur et la complexité des formes. Même si quelques unes des directions d'analyse sont avant tout des terrains de recherche propres au domaine de la *computational photography*, la détection des centres de gravité et des palettes de couleurs semblent suffisamment au point – c'est-à-dire *simples* – pour faire office de programme exécutables à même d'être inclus – pourquoi pas, plus tard – comme fonctionnalités d'une nouvelle base de données à l'architecture plus souple.

Cette collaboration avec l'IVRG fait également émerger une nouvelle problématique relative à la nécessaire uniformisation du matériel de dessin et de scannage. En effet, travailler sur des pixels exige une attention préalable à l'équivalence de ces pixels : un « rouge » produit par deux crayons de marques différentes sera très vite traduit comme deux couleurs différentes au cours du processus de pixellisation. Le problème est le même pour le papier et le matériel de scannage qui produisent vite des différences de pixel qu'il est coûteux (mais pas impossible) de corriger par la suite.

Déménagement et espérances

On l'a dit, les codages en termes de « descripteurs », « types », « sous-types », « classes » et « sous-classes » ne se situent pas dans la base de données Filemaker mais bien à côté, dans des fiches Excel. En l'état, la base de données met à disposition les dessins et les classe en fonction de requêtes faisant écho à la codification des fichiers TIFF ou JPEG.

Il s'ensuit que ce système n'est pas assez flexible : il faut sans cesse jongler entre la base de données et la feuille Excel et bien faire attention à ne pas modifier l'une sans modifier l'autre. Sans compter que le nombre de données augmente et que le projet intéresse de plus en plus de monde... À coup sûr, même si elle a rendu de fiers services par le passé, l'architecture de la base de données ne fait plus assez *honneur* à la richesse des données collectées et aux possibilités de recherche qu'elles encapsulent. Mais comment faire ? Filemaker est un logiciel de gestion de base de données trop isolé et aux fonctionnalités trop pauvres ; il apparaît nécessaire de le fuir pour émigrer vers les terres plus fertiles d'un système PHP MySQL.

Les idées foisonnent : si le déménagement se déroule bien, peut-être sera-t-il possible de créer une véritable *plateforme* avec différents *niveaux d'autorisation* permettant d'élargir la gamme des acteurs impliqués dans le traitement des données ! Ceux-ci pourraient lancer leurs propres analyses en usant des algorithmes de *clustering* ou d'analyse des pixels mis à disposition par la plateforme. Ça serait là une solution honorable à l'inéluctable arbitraire du codage et des interprétations. Moyennant également un système de dépôt et d'évaluation, il sera peut-être possible d'héberger d'*autres* dessins d'enfants ! À terme, les dessins de dieux pourraient ne devenir qu'une partie d'un plus grand ensemble : l'ensemble des dessins d'enfants capables de soutenir de nouvelles recherches en psychologie du développement.

Les espérances sont nombreuses mais après tout, un système PHP-MySQL serait tout à fait capable de les rendre réelles *pour autant que le déménagement ne se fasse pas à la hâte*. C'est là précisément que se situe le prochain défi (le plus grand ?) du projet « Dessins de dieux » : comment *bien* ré-agencer la base de données ? Comment effectuer un *bon* déménagement ?

Florian Jaton



UNIL | Université de Lausanne

LADHUL - Laboratoire
de cultures et humanités
digitales de l'UNIL