Table 6 - Appendix 1: Parameters values and results of the robustness test

Classification type: Simple - # of analyzed forms: 1,930 (>=22) - % of encompassed t. s.: 78.38%

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
data (603.29)	model (523.14)	firm (374.39)	term (339.51)	waste (1843.32)	to organize (519.82)	purchaser (720.98)	to sell (1127.2)	president (696.24)	euro (1879.28)
digital (381.26)	to change (413.18)	consultancy (370.81)	political (331.79)	recycling (1061.14)	theme (503.63)	public (669.32)	to rent (857.60)	jean (634.41)	funds (1780.05)
to connect (346.66)	new (203.64)	association (331.09)	social (321.40)	matter (899.97)	hour (465.98)	market (525.82)	renting (726.14)	françois (607.48)	nord (1401.95)
internet (303.37)	really (197.29)	development (289.14)	economical (298.56)	product (779.53)	tuesday (378.13)	purchase (525.57)	machine (500.41)	director (600.43)	calais (1108.06)
customer (282.02)	thing (153.69)	club (266.09)	climatic (269.96)	to recycle (635.24)	conference (363.19)	procur. (490.41)	tire (500.02)	michel (536.36)	third (1006.33)

Classification type: Simple - # of analyzed forms: 2,727 (>=13) - % of encompassed t. s.: 78.76%

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
social (542.79)	director (800.21)	firm (331.67)	model (244.08)	waste (1763.90)	emission (921.76)	to orga. (605.98)	data (629.56)	to sell (1060.3)	circular (741.65)
political (300.70)	president (590.68)	pole (248.52)	deep (167.27)	recycling (1167.80)	energy (901.76)	hour (507.00)	digital (448.75)	to rent (835.62)	law (725.95)
economical (297.45)	jean (330.09)	cluster (239.47)	crisis (167.07)	reuse (644.78)	gas (789.23)	thursday (399.27)	to conn. (400.05)	renting (794.97)	ademe (420.64)
term (272.69)	founder (285.42)	association (218.32)	to change (166.75)	to recycle (638.07)	greenhouse (533.17)	april (383.60)	digit (313.21)	car (485.17)	mp (400.01)
environm. (271.80)	general (265.33)	project (210.26)	change (141.57)	reemploy. (601.05)	co2 (464.24)	friday (304.33)	internet (289.50)	vehicle (457.01)	minister (395.61)

Classification type: Simple - # of analyzed forms: 4,369 (>=6) - % of encompassed t. s.: 85.20%

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
director (414.29)	term (429.49)	to implem. (338.48)	waste (1897.25)	emission (1122.23)	work (647.12)	circular (613.75)	to sell (894.98)	climatic (334.40)	data (904.06)
hub (358.04)	model (378.44)	actor (287.76)	recycling (1278.73)	gas (1017.59)	job (391.85)	minister (544.90)	renting (802.51)	country (285.84)	to conn. (713.93)
association (320.21)	economical (357.38)	implementa. (192.96)	to recycle (887.69)	energy (927.70)	value (337.28)	institute (493.07)	to rent (779.44)	change (228.95)	digital (555.59)
club (262.50)	long (186.52)	role (189.66)	reemploy. (690.68)	greenhouse (630.01)	relevance (314.43)	law (477.66)	car (456.10)	growth (193.66)	digit (520.30)
firm (247.83)	impact (173.66)	organization (177.29)	material (687.64)	co2 (533.46)	productivity (294.15)	mp (414.28)	sales (431.94)	inequality (173.14)	internet (464.11)

Classification type: Simple - # of analyzed forms: 5,436 (>=4) - % of encompassed t. s.: 90.41%

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
to implem. (252.73)	waste (1270.76)	car (816.30)	to sell (1549.45)	hub (352.57)	minister (542.84)	climatic (243.87)	funds (1779.95)	data (1012.80)	public (826.99)
economical (212.07)	recycling (871.06)	renting (554.68)	michelin (903.90)	firm (350.47)	circular (486.53)	growth (238.93)	euro (1520.87)	digital (787.30)	purchaser (693.80)
new (200.07)	matter (741.73)	vehicle (474.56)	tire (821.40)	director (343.09)	institute (439.24)	crisis (226.53)	nord (1385.74)	to conn. (632.89)	procurement (551.13)
actors (192.46)	product (554.90)	to rent (363.12)	use (709.81)	club (275.64)	françois (411.71)	change (160.65)	third (1338.36)	google (412.67)	nssd (510.26)
role (154.23)	to recycle (525.62)	shop (346.29)	kilometer (571.06)	association (272.91)	mp (378.84)	to face (146.89)	revolution (1088.89)	digit (406.56)	criterion (465.32)

Classification type: double (12-14) - # of analyzed forms: 3,679 (>=8) - % of encompassed t. s.: 57.08%

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
project (604.15)	consultancy (273.18)	renting (738.65)	new (150.87)	to orga. (581.54)	mp (471.96)	resource (312.25)	waste (389.16)
call (386.43)	cjd (250.97)	car (522.48)	value (124.02)	april (499.23)	nicolas (397.92)	natural (221.83)	law (363.43)
region (382.46)	firm (228.80)	to rent (513.73)	model (123.23)	hour (498.85)	lambert (366.66)	growth (177.41)	article (342.33)
territory (326.11)	club (202.70)	shop (312.18)	customer (83.09)	date (382.34)	michel (362.31)	emission (156.73)	public (322.12)
ademe (275.41)	manager (145.87)	vehicle (285.48)	digital (74.17)	registration (338.81)	françois (319.61)	energy (145.94)	product (316.49)

Classification type: double (12-16) - # of analyzed forms: 3,679 (>=8) - % of encompassed t. s.: 54.83%

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
data (294.58)	resource (720.62)	renting (653.46)	project (673.95)	consultancy (343.74)	to organ. (646.26)	mp (479.45)	article (390.79)
digital (216.16)	natural (447.51)	to rent (524.78)	call (398.80)	practice (184.34)	hour (506.39)	lambert (413.44)	public (373.75)
digit (191.45)	climatic (259.46)	to sell (449.53)	region (362.79)	firm (157.09)	april (440.69)	françois (378.36)	law (365.78)
value (139.35)	to extract (221.64)	car (380.07)	territory (342.83)	cjd (156.29)	registration (362.61)	nicolas (372.00)	waste (347.88)
new (110.70)	to consu. (214.20)	tire (265.13)	ademe (279.06)	master (146.74)	cci (323.16)	minister (353.09)	objective (285.93)

For every corpus, the number of clusters was adjusted so that the descending hierarchical classification encompassed a minimum of 75% of classified text segments – t.s. = text segments

Table 7 - Appendix 2: Parameters values for press-type corpus-related DHC

Press type	# texts	# text segments (t.s.)	# active forms	# analyzed forms (freq.)	average # forms/t.s.	# clusters	% classified text segments
Business & Economics	152	4,003	9,015	1,890 (>=6)	36.002	6	99.65%
Environmental	97	2,083	5,968	2,408 (>=3)	35.985	6	84.25%
General	249	4,713	10,193	2,093 (>=6)	35.641	6	81.14%
Government & Public Actors	36	1,954	5,474	2,267 (>=3)	36.252	6	95.29%
Professional	76	2,505	6534	2,542 (>=3)	35.853	6	75.89%

To facilitate the comparison, the number of clusters was adjusted to six for each press type while controlling that the descending hierarchical classification encompassed a minimum of 75% of the classified text segments.