

ARTICLE DE RECHERCHE

Anatomie d'un communiqué de presse

Un recensement préenregistré de 246 777 communiqués en anglais sur 1 151 sites

Émile Bolduc^{1*} 

¹ Centre de recherche PPN World, PPN World, Montréal (Québec), Canada

* Auteur correspondant: ppnw@ppnsource.com · ORCID <https://orcid.org/0009-0005-7882-4741>

Citation: Bolduc, É. (2026). Anatomie d'un communiqué de presse : un recensement préenregistré de 246 777 communiqués en anglais sur 1 151 sites (PPN World Research Paper No. 2, Edition 1). PPN World. <https://www.ppn-world.com/fr/research/anatomy-of-a-press-release>

Enregistrée: 11 octobre 2026

Données recueillies: 11 octobre 2026 – 13 octobre 2026

Publiée: 15 octobre 2026

Corrigée: 17 octobre 2026

Droits d'auteur: © 2026 PPN World. Article en libre accès distribué selon les termes de la licence Creative Commons Attribution 4.0 International (CC BY 4.0), qui permet l'utilisation, la distribution et la reproduction sans restriction sur tout support, à condition de citer l'auteur et la source.

Financement: Cette étude a été financée par Bollé Communications inc., qui exploite PPN World et PPN Source. Elle n'a reçu aucun financement externe.

Intérêts concurrents: L'auteur est affilié à PPN World (Bollé Communications inc.), qui a publié cette étude, exploite un service commercial de veille de communiqués et a fourni l'archive dont la population est tirée. PPN Source, un diffuseur qu'exploite la même entreprise, est l'un des sites de la population (6 communiqués).

Résumé

Contexte. Le communiqué de presse est l'un des textes les plus écrits de la communication professionnelle et la ressource que les journalistes disent trouver la plus utile, mais sa forme n'a guère été mesurée que sur des échantillons de quelques centaines à quelques milliers de textes d'un seul secteur, ou dans des décomptes commerciaux limités aux diffuseurs.

Méthodes. Nous avons entrepris de lire chaque communiqué en anglais portant une date de publication du 1er juillet 2026 au 6 octobre 2026 dans une archive de communiqués — 246 777 communiqués sur 1 151 sites de diffuseurs, d'entreprises, de gouvernements nationaux, régionaux et locaux, d'universités, de ligues sportives et d'organismes sans but lucratif — en respectant le robots.txt de chaque site, et en avons mesuré 176 492 sur 877 sites, avec des règles de détection enregistrées, avec le code d'analyse, avant la lecture de toute page, puis révisées une fois, publiquement, après un premier test de fiabilité. Les règles ont été vérifiées contre un codeur lisant les pages, et quatre hypothèses testées par un bootstrap par grappes de sites avec une correction de Holm.

Résultats. Le communiqué s'est scindé en deux. Sur les sites des diffuseurs, 64 % des communiqués se terminent par une section « À propos » et 76 % s'ouvrent sur une ligne de date ; sur les sites propres des organisations, 3 % et 17 % (4,3 % et 11 % sans les sites sportifs, dont les pages sont surtout des articles). Les communiqués des diffuseurs sont plus longs de 155 mots à la médiane (H1 confirmée), surtout hors du récit : le corps du texte, de l'attaque à la section « À propos », n'est pas plus long de façon décelable. Les citations de dirigeants portent un mot d'émotion plus souvent que celles d'officiels (11,8 % contre 4,1 % ; H4 confirmée). Les règles de la ligne de date (H2) et du contact médias (H3) n'ont pas atteint la fiabilité exigée par l'enregistrement : ces deux hypothèses sont rapportées, non tranchées. POUR DIFFUSION IMMÉDIATE survit sur 3,2 % des communiqués, dont 77 % sur un seul site, celui du département de la Justice des États-Unis.

Conclusions. La section « À propos », une convention de l'ère du fil de presse, appartient à la copie du diffuseur ; la ligne de date et le contact médias vont dans le même sens, mais leurs règles n'ont pas pu le trancher. La copie qu'une organisation publie sur son propre site a gardé le récit et peu de l'appareil qui l'entourait.

Mots-clés: press release; public relations; genre; content analysis; census; dateline; media contact; boilerplate; newswire; pre-registration

1. Introduction

Un communiqué de presse est une déclaration qu'une organisation publie sur elle-même, à l'adresse des médias et du public. C'est l'une des plus anciennes formes de relations publiques — on fait conventionnellement remonter le genre à la déclaration qu'Ivy Lee publia pour le Pennsylvania Railroad après un accident de train en octobre 1906 (Hiebert, 1966) — et c'est toujours le matériau que les journalistes disent trouver

le plus utile : dans un sondage commercial mené en 2025 auprès de plus de 3 000 journalistes de 19 marchés, 72 % le désignaient comme la ressource la plus utile qu'une équipe de communication puisse fournir (Cision, 2025).

La manière d'écrire un communiqué est enseignée dans tous les cours de relations publiques : une ligne de date, une attaque qui répond à qui, quoi, quand et où, la citation d'un dirigeant, un paragraphe « À propos », un contact médias,

une marque de fin. Les linguistes décrivent le résultat comme un genre hybride, informatif par sa forme et promotionnel par son but (Catenaccio, 2008), et comme un texte écrit pour être recopié — « préformulé » pour l'article du journaliste, avec des citations prêtes à reprendre (Jacobs, 1999). Que cette forme survive au passage du fil de presse au web, et qui la conserve, n'avait pas, à notre connaissance, été mesuré à grande échelle. Les analyses de contenu du genre portent surtout sur quelques centaines à quelques milliers de communiqués d'un seul secteur (Sumner et al., 2014 ; Bossema et al., 2019 ; McLaren et Gurău, 2005 ; Lassen, 2006) ; un corpus universitaire d'environ 24 600 communiqués d'un seul diffuseur servait une autre question — la reprise de leur texte par les médias (Nicholls, 2019) ; le plus grand décompte de la forme elle-même que nous ayons trouvé est commercial, l'analyse par un diffuseur de 138 901 communiqués de dix fils de presse, limitée aux diffuseurs (PR Gun, 2026). Cette étude entend mesurer chaque communiqué en anglais d'une population définie sur un trimestre, avec des hypothèses et une analyse enregistrées avant la lecture de toute page.

2. Matériel et méthodes

2.1 Population

La population comprend chaque communiqué de l'archive de PPN World portant une date de publication du 1er juillet 2026 au 6 octobre 2026, en anglais, provenant des sources servies par l'archive : 246 777 communiqués sur 1 151 sites. Un site est une origine — schéma, hôte et port — et prend le type d'éditeur des sources qui y renvoient. Les diffuseurs (fils de presse commerciaux) ont publié 142 073 des communiqués (58 %) ; le reste a été publié par des gouvernements nationaux et leurs organismes, des gouvernements d'États, de provinces et de villes, des ligues sportives, des entreprises, des universités, des organismes sans but lucratif, des organisations politiques et quelques autres sur leurs propres sites. Les gouvernements nationaux comprennent ici des organisations internationales comme les Nations unies et les institutions de l'Union européenne (13 % du poids de ce groupe), et les ligues sportives leurs clubs et d'autres organismes sportifs. Les sources de l'archive sont des salles de presse, et tout ce qu'elles publient n'est pas un communiqué au sens strict : les sites sportifs publient surtout des articles de leurs propres rédacteurs, quelques entreprises des billets de blogue, et 92 % du poids des organisations politiques est une archive des messages de Donald Trump sur Truth Social. Le recensement garde ce que les sources ont publié, comme enregistré ; les limites donnent les chiffres des sites propres sans les sites sportifs. L'étude est un recensement : chaque communiqué de la population a été inclus, aucun échantillonné.

2.2 Lecture et extraction

Les pages ont été lues du 11 octobre 2026 au 13 octobre 2026. Le robots.txt de chaque site a été lu d'abord et appliqué à chaque communiqué (Koster et al., 2022) ; les requêtes vers chaque site portaient une à une, à au moins une seconde d'intervalle, d'un client qui s'identifiait, puis une seconde fois comme navigateur en cas de refus. Aucun test anti-robot n'a été résolu, aucune erreur de certificat contournée. Le texte a

été extrait avec Mozilla Readability, l'algorithme du mode lecture de Firefox (Barbresi, 2021 ; Bevendorff et al., 2023). Un communiqué est mesuré si sa page a répondu avec au moins 50 mots d'anglais portant le titre de l'archive : 176 492 communiqués (72 %) sur 877 sites. Les autres sont comptés par motif — surtout des pages refusées ou disparues (19 795), des certificats invalides (17 428, presque tous sur le site d'un même diffuseur), des tests anti-robots (11 059) et des robots.txt injoignables (6 077). Dans chaque site, les communiqués mesurés sont pondérés pour que chaque site garde sa part de la population.

2.3 Mesures

Un seul programme enregistré lit chaque page : le titre (longueur, un chiffre, un verbe d'annonce) ; la ligne de date (un lieu en capitales, une date ou la mention d'un diffuseur, puis un tiret, en tête du texte) ; le sous-titre ; une signature ; la longueur en mots, totale et du corps, de l'attaque à la section « À propos » ; les citations de huit mots ou plus avec un verbe de parole, attribuées à un dirigeant quand un titre d'entreprise accompagne l'orateur et à un officiel quand c'est un titre public, et si elles contiennent un mot d'émotion (pleased, excited, thrilled, delighted, proud, honored, happy, glad, elated) ; la section « À propos » ; un contact médias — une mention de contact avec un courriel ou un numéro de téléphone ; une marque de fin (###, -30-) ; POUR DIFFUSION IMMÉDIATE (FOR IMMEDIATE RELEASE) ; images, vidéo et liens.

2.4 Hypothèses et analyse

Quatre hypothèses ont été enregistrées : les communiqués des diffuseurs sont plus longs que ceux des sites des organisations (H1, écart des nombres de mots médians pondérés) ; la ligne de date est une convention de diffuseur (H2, écart des parts) ; la plupart des communiqués ne nomment personne à joindre (H3, part avec un contact, comparée à la moitié) ; les mots d'émotion sont plus fréquents dans les citations de dirigeants que dans celles d'officiels (H4, communiqués commerciaux contre ceux des organismes publics, par citation). Intervalles et valeurs p unilatérales proviennent de 2 000 répliques bootstrap qui rééchantillonnent les sites à l'intérieur de chaque type d'éditeur ; les quatre tests portent une correction de Holm (Holm, 1979). Une mesure ne pouvait trancher une hypothèse que si elle s'accordait avec un codeur lisant la page sur au moins 90 % des communiqués, avec un α de Krippendorff $\geq 0,800$ (Krippendorff, 2004 ; Lombard et al., 2002). Ce qui avait été vu avant l'enregistrement — les identifiants, adresses et titres de l'archive, et deux pilotes de 100 et 1 001 communiqués hors de la population — est consigné dans l'enregistrement, comme le recommandent les analyses de données existantes (Weston et al., 2019).

2.5 Fiabilité des mesures

Fiabilité. Le codeur A, un modèle de langage (Claude Opus 5.5, Anthropic) — ce type de modèle a égalé ou dépassé des travailleurs de plateforme sur des tâches d'annotation comparables (Gilardi et al., 2023) — a codé 200 communiqués tirés au hasard — au moins 60 des diffuseurs et 60 des organismes publics, comme enregistré ; le tirage en a donné 106, 70 et 24 d'autres éditeurs — à partir de la page elle-même, selon le

guide de codage enregistré, sans voir le résultat du détecteur. Un second codeur, humain, était enregistré pour un sous-ensemble et s'est retiré après cinq communiqués ; l'étude repose donc sur un seul codeur, un modèle de langage, et le dit. Là où les règles enregistrées échouaient, elles ont été révisées sur ces 200 communiqués puis testées une seule fois, sans modification, sur 100 nouveaux communiqués tirés de la même façon (au moins 30 et 30 ; le tirage en a donné 56, 37 et 7) et codés à l'aveugle (figure 2). L'extraction a rendu le communiqué entier sur 78 % des pages du test, et une partie sur la plupart des autres.

2.6 Écarts par rapport à l'enregistrement

1. Cadence. L'étape de mesure a tourné en parallèle et plus de sites ont été lus à la fois que prévu ; les règles de courtoisie sont restées les mêmes, et remesurer le second pilote en parallèle a donné des résultats identiques. L'ordinateur de recherche s'est mis en veille pendant une partie du recensement ; les communiqués échoués pendant la veille ont été relus, ordinateur éveillé (les deux décomptes sont rapportés).
2. Les refus 429 Too Many Requests ont été relus une fois, à au moins 30 secondes d'intervalle par site ; les hypothèses sont tranchées sur le décompte enregistré, le décompte relu est une ligne de robustesse.
3. Le second codeur, humain, était enregistré pour 50 communiqués et douze questions, puis pour 20 et quatre ; il s'est arrêté après cinq, et aucun accord entre codeurs n'est calculé.

4. Les règles enregistrées du contact et de la citation de dirigeant ont échoué au test de fiabilité sur les 200 premiers communiqués. La règle du contact lit désormais aussi la page autour du texte extrait ; les titres de dirigeants sont lus quelle que soit la casse et reliés à un orateur nommé plus loin par son seul nom ; les guillemets sont appariés à l'intérieur d'un paragraphe. Les pages de GOV.UK qui ne sont pas des communiqués (avis statistiques, consultations, publications) ont été mises de côté : 3655 pages, comptées dans la population parmi celles non mesurées. Les règles révisées ont été figées avant un second test, à l'aveugle, sur 100 nouveaux communiqués, dont le résultat a été consigné avant la nouvelle mesure du recensement. Les résultats enregistrés sont rapportés à côté des résultats révisés.

3. Résultats

3.1 Deux communiqués

La figure 1 lit un communiqué de haut en bas sur les deux types de sites. Les éléments que le fil de presse a rendus standards — la ligne de date en tête, la section « À propos » et le contact en pied — sont la norme sur les sites des diffuseurs et rares sur ceux des organisations : une section « À propos » clôt 64 % des communiqués des diffuseurs contre 3 %, une ligne de date ouvre 76 % contre 17 %, un contact figure sur 65 % contre 14 %. La citation est l'élément que les deux gardent : 59 % des communiqués des diffuseurs et 42 % des autres citent quelqu'un.

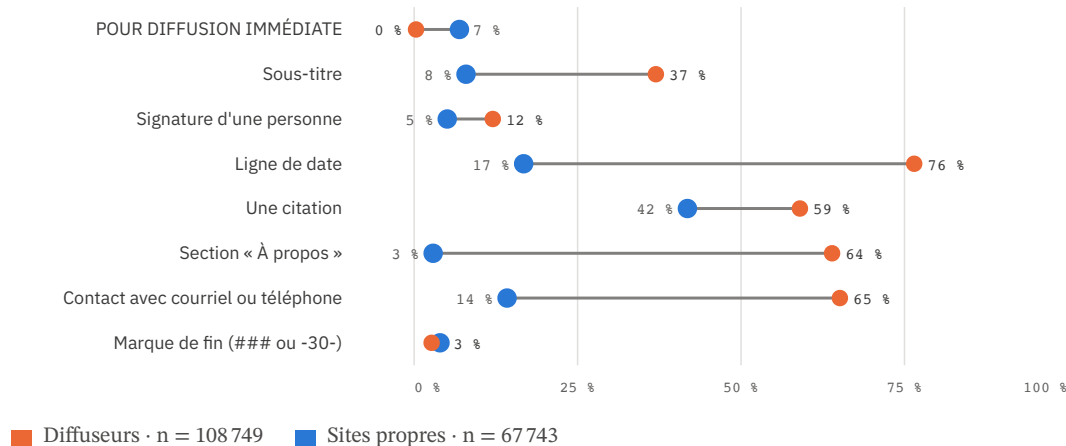


Figure 1. Deux communiqués. Part des communiqués portant chaque élément, sur les sites des diffuseurs et sur les sites propres des organisations, pondérée pour les communiqués qui n'ont pu être lus. Éléments dans l'ordre de lecture d'un communiqué. Les règles du sous-titre, de la marque de fin et de la signature n'ont pas atteint le seuil de fiabilité (figure 2) : ces lignes sont indicatives.

Le partage n'oppose pas les entreprises aux gouvernements. Les entreprises qui publient dans leur propre salle de presse gardent plus de cet appareil que les organismes publics — 21 % ont une section « À propos » et 24 % un contact — mais bien moins que le même genre de texte sur le site d'un diffuseur (tableau 1). Les gouvernements régionaux et locaux font

exception parmi les organismes publics : 46 % de leurs communiqués nomment un contact. Les ligues sportives, premier groupe de sites propres mesuré, ne le font presque jamais (0,6 %) : leurs pages sont surtout des articles, non des communiqués.

Tableau 1. Anatomie d'un communiqué, par éditeur. Communiqués mesurés, longueur médiane en mots, et part portant chaque élément, pondérée pour les communiqués qui n'ont pu être lus. Les diffuseurs sont les fils de presse commerciaux ; toutes les autres lignes publient sur leurs propres sites. Les gouvernements nationaux comprennent des organisations internationales ; les ligues sportives, des clubs et d'autres organismes sportifs, dont les pages sont surtout des articles ; l'essentiel du poids des organisations politiques est une archive des messages de Donald Trump sur Truth Social.

Éditeur	Mesurés	Mots (médiane)	Ligne de date	« À propos »	Contact	Citation	Pour diffusion immédiate
Tous les communiqués	176 492	588	51 %	38 %	43 %	52 %	3 %
Diffuseurs	108 749	637	76 %	64 %	65 %	59 %	0 %
Entreprises	4 120	718	17 %	21 %	24 %	36 %	0 %
Gouvernements nationaux	23 152	401	9 %	2 %	19 %	35 %	15 %
Gouv. régionaux et locaux	7 106	405	21 %	4 %	46 %	43 %	13 %
Universités	4 134	678	2 %	2 %	17 %	66 %	0 %
Ligues sportives	26 362	594	27 %	1 %	1 %	47 %	0 %
Organismes sans but lucratif	1 621	513	15 %	10 %	11 %	46 %	0 %
Organisations politiques	873	128	2 %	0 %	0 %	5 %	0 %
Tous les sites propres	67 743	482	17 %	3 %	14 %	42 %	7 %

3.2 La longueur

Le communiqué médian compte 588 mots, à un niveau Flesch-Kincaid de 15 (Kincaid et al., 1975) — écrit pour un lecteur universitaire. H1 est confirmée : les communiqués des diffuseurs sont plus longs que ceux des sites des organisations de 155 mots à la médiane (intervalle à 95 % : 55 à 244 ; p ajustée de Holm = 0,005).

L'enregistrement posait la même question pour le seul corps du texte, de l'attaque à la section « À propos », au contact ou à la marque de fin (R5). L'écart disparaît presque : 468 mots sur les sites des diffuseurs, 465 sur ceux des organisations (écart de 3, intervalle à 95 % : -84 à 96). Sans les sites sportifs, dont les articles sont longs, il est de 57 mots (intervalle à 95 % : -2 à 117 ; une vérification non enregistrée), contre 204 pour le communiqué entier. Le communiqué d'un diffuseur est plus long surtout à cause de ce qui entoure le récit — ligne de date, section « À propos », bloc de contact, mentions légales — et non, de façon décelable, parce que le récit l'est.

3.3 Qui appeler

Le bloc de contact est le seul élément d'un communiqué écrit pour le journaliste seul. H3 demandait si la plupart des communiqués ne nomment désormais personne à joindre. Tel que mesuré, 43 % des communiqués portent un courriel ou un numéro de téléphone sous une mention de contact (intervalle à 95 % : 24 % à 57 %) — 65 % sur les sites des diffuseurs, 14 % sur ceux des organisations.

La mesure n'a pas pu trancher. Sur 100 communiqués tirés pour la tester, la règle s'est accordée avec un codeur lisant la page dans 89 % des cas (α ,779), sous les 90 % enregistrés. Elle se trompe rarement quand elle trouve un contact (précision 96 %), mais a manqué 16 % de ceux trouvés par le codeur, le plus souvent un contact sous une formule qu'elle ne connaît pas (« Questions? », « Address for clarification ») ou un contact que l'extraction a laissé de côté. La part du recense-

ment est donc un plancher, assez proche de la moitié pour que la question enregistrée reste ouverte : H3 est rapportée, non tranchée. Des manques à ce taux ne combleraient toutefois pas l'écart entre les deux types de sites.

3.4 La ligne de date

Une dépêche envoyée de l'extérieur s'ouvre sur le lieu où elle a été écrite, en capitales (Associated Press, 2026) ; les diffuseurs appliquent la convention aux communiqués, souvent avec la date et leur propre nom. 51 % des communiqués s'ouvrent ainsi : 76 % sur les sites des diffuseurs et 17 % sur ceux des organisations, un écart de 60 points (intervalle à 95 % : 23 à 73).

H2 est rapportée, non tranchée. La règle de la ligne de date a réussi le test de fiabilité sur les 200 premiers communiqués codés (90,5 % d'accord, α ,810) et l'a échoué sur les 100 tirés pour tester les règles révisées (87 %, α ,739) ; l'enregistrement applique le test au second tirage. Ses erreurs vont surtout dans un sens : quand elle trouve une ligne de date, elle a presque toujours raison (précision 98 %) ; elle a manqué 20 % de celles trouvées par le codeur.

3.5 Citations et émotions

52 % des communiqués citent quelqu'un dans une phrase de huit mots ou plus. H4 compare les dirigeants cités dans les communiqués commerciaux aux officiels cités dans ceux des organismes publics. Les citations de dirigeants portent un mot d'émotion — pleased, excited, thrilled, delighted, proud, honored, happy, glad, elated — dans 11,8 % des cas (132 328 citations) ; celles d'officiels dans 4,1 % (13 313). H4 est confirmée : l'écart est de 7,6 points (intervalle à 95 % : 4,9 à 8,8 ; p ajustée de Holm = 0,002).

L'émotion est une formule plutôt que la règle : la plupart des citations de dirigeants ne contiennent aucun des neuf mots. Quand l'un d'eux apparaît, il relève du registre que décrit Jacobs (1999) — une citation écrite par l'équipe de communication, pour être reprise telle quelle dans un article.

Tableau 2. Les quatre hypothèses enregistrées. Estimations avec les règles de détection révisées, intervalles bootstrap à 95 % par grappes et valeurs p ajustées de Holm ; la dernière colonne donne l'estimation avec les règles enregistrées. H2 et H3 reposent sur des mesures restées sous la règle de fiabilité enregistrée et ne sont pas tranchées.

Hypothèse	Estima- tion	Intervalle à 95 %	p (Holm)	Verdict	Enregis- trée
H1 · Les communiqués des diffuseurs sont plus longs (mots, médiane)	155	55 à 244	0,005	Confirmée	169
H2 · La ligne de date est une convention de diffuseur (écart, points)	60	23 à 73	0,002	Non tranchée	60
H3 · La plupart des communiqués ne nomment personne (part avec contact)	43 %	24 % à 57 %	0,185	Non tranchée	29 %
H4 · Plus de mots d'émotion chez les dirigeants (écart, points)	7,6	4,9 à 8,8	0,002	Confirmée	8,1

3.6 Les marques de l'ère du télécopieur

POUR DIFFUSION IMMÉDIATE indiquait à une rédaction qu'un communiqué sur papier pouvait paraître tout de suite. Sur une page web, il n'a plus de lecteur à instruire, et il est rare : 3,2 % des communiqués le portent, 0,3 % sur les sites des diffuseurs. Il survit chez les pouvoirs publics, et surtout sur un site : le département de la Justice des États-Unis l'imprime sur chaque communiqué mesuré, et ses communiqués font 77 % de tous ceux qui le portent — une estimation qui repose sur les 365 de ses 4862 communiqués qui ont pu être lus (presque tous les autres étaient derrière un test anti-robot), que les pondérations font valoir pour tous. La part des gouvernements nationaux est de 14,9 % avec lui et de 0,5 % sans lui ; les gouvernements régionaux et locaux le gardent sur 13,0 % des communiqués. La marque de fin — ### ou -30-, « fin du texte » — figure sur 3,2 % des communiqués, sur le quart de ceux des gouvernements régionaux et locaux (26 %) et peu d'autres ; la règle de la marque de fin, comme celles du sous-titre et de la signature, n'a pas atteint le seuil de fiabilité (figure 2) : ces parts sont indicatives.

Les titres sont longs. Le titre médian compte 81 caractères, 93 sur les sites des diffuseurs et 67 sur ceux des organisations ;

75 % de tous les titres et 86 % de ceux des diffuseurs dépassent 60 caractères, près des 90,2 % trouvés sur les fils de presse par le décompte d'un diffuseur (PR Gun, 2026). Un verbe d'annonce (announces, launches, unveils...) ouvre 31 % des titres des diffuseurs et 8 % des autres.

3.7 Vérifications de robustesse

Robustesse. Pondérer chaque site également, retirer les cinq plus grands sites, retirer les articles signés (tels que les trouve la règle de la signature, qui n'a pas atteint le seuil de fiabilité), les pages lues seulement après rendu ou les pages que le client identifié n'a pu lire, supprimer les pondérations, et compter les communiqués d'abord refusés avec 429 Too Many Requests (1853 mesurés de plus) ou garder tels que d'abord enregistrés les communiqués perdus pendant la mise en veille de l'ordinateur de recherche (917 mesurés de moins) : chaque variante refait les quatre tests. H1 et H4 sont confirmées dans chacune ; l'écart de H1 va de 94 à 163 mots, celui de H4 de 5,5 à 7,7 points. Avec les règles enregistrées, 179606 communiqués étaient mesurés et les estimations étaient de 169 mots (H1), 60 points (H2), 29 % avec contact (H3) et 8,1 points (H4) ; chaque test enregistré était alors nominalement confirmé.

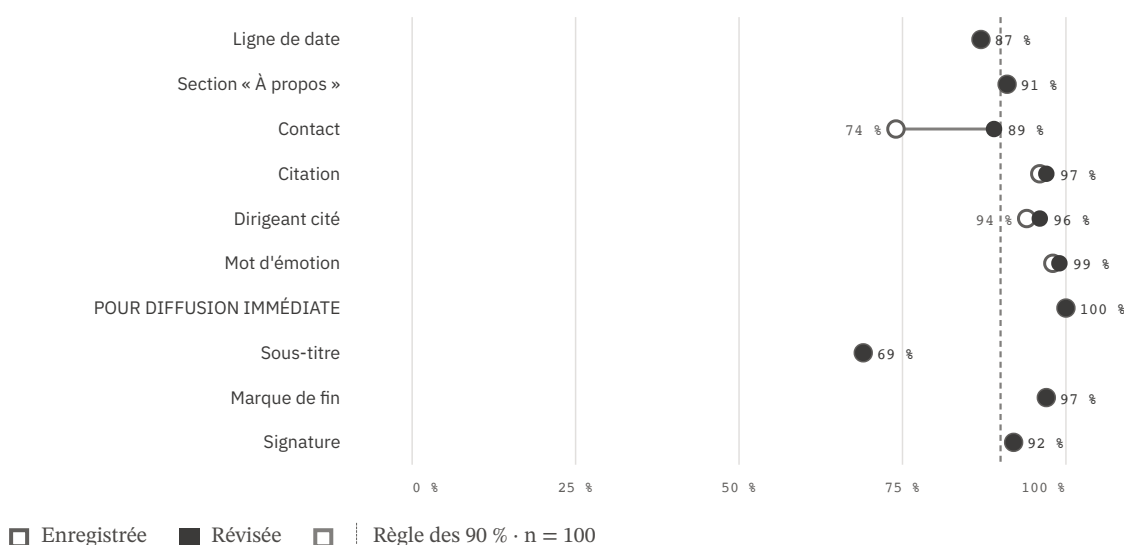


Figure 2. Jusqu'où chaque mesure est fiable. Accord entre la règle de détection et un codeur lisant la page, sur 100 communiqués tirés après la révision des règles (creux : la règle enregistrée ; plein : révisée). Une mesure ne tranche une hypothèse qu'à 90 % d'accord avec un α de Krippendorff $\geq 0,800$.

4. Discussion

4.1 Principaux résultats

1. **La section « À propos » appartient aux diffuseurs.** 64 % des communiqués publiés sur les sites des diffuseurs se terminent par un paragraphe type décrivant l'organisation ; 3 % de ceux publiés sur les sites des organisations. Le même partage vaut pour la ligne de date (76 % contre 17 %).
2. **Les communiqués des diffuseurs sont plus longs — surtout autour du récit (H1).** À la médiane, un communiqué compte 637 mots sur le site d'un diffuseur et 482 sur celui de l'organisation (intervalle à 95 % de l'écart : 55 à 244 ; H1 confirmée). Le corps du texte, de l'attaque à la section « À propos », n'est pas plus long de façon décelable : 468 contre 465 mots (468 contre 411 sans les articles des sites sportifs). L'écart tient surtout à l'appareil qui l'entoure.
3. **Le communiqué type fait 588 mots ; celui d'une entreprise, 718 ; celui d'un gouvernement national, 401.** En médiane, un communiqué fait 588 mots : 637 sur les sites des diffuseurs et 482 sur les sites des organisations. Selon qui le publie, celui d'une entreprise fait 718 mots, d'une université 678, d'une ligue sportive 594, d'un gouvernement régional ou local 405 et d'un gouvernement national 401. L'écart tient surtout au texte lui-même : le corps du communiqué d'une entreprise fait 637 mots, celui d'un gouvernement national 387.
4. **Les dirigeants se disent ravis plus souvent que les officiels (H4).** Un mot d'émotion — pleased, excited, thrilled, delighted, proud — figure dans 11,8 % des citations de dirigeants des communiqués commerciaux et dans 4,1 % des citations d'officiels des organismes publics (écart de 7,6 points, intervalle à 95 % : 4,9 à 8,8 ; H4 confirmée).
5. **POUR DIFFUSION IMMÉDIATE survit sur un site fédéral — et chez les gouvernements régionaux et locaux.** La mention écrite pour le télécopieur survit sur 3,2 % des communiqués, et 77 % d'entre eux sont sur un seul site : le département de la Justice des États-Unis l'imprime sur chaque communiqué que nous avons pu lire (365 sur 4862). Sans lui, la part des gouvernements nationaux tombe de 14,9 % à 0,5 %, et la part globale à 0,8 %. Les gouvernements régionaux et locaux la gardent sur 13,0 % des communiqués ; les diffuseurs sur 0,3 %.
6. **Les titres sont longs, surtout chez les diffuseurs.** Le titre médian compte 81 caractères — 93 sur les sites des diffuseurs, 67 sur ceux des organisations. 75 % des titres dépassent 60 caractères, 86 % chez les diffuseurs.

4.2 Ce que cela signifie pour les communicateurs

Quelle est la longueur d'un communiqué de presse typique ? 588 mots à la médiane, sur 176492 communiqués en anglais mesurés en 2026 : 637 sur les sites des diffuseurs, 482 sur ceux des organisations. Le récit lui-même — de l'attaque à la section « À propos » — compte 468 mots sur les sites des diffuseurs et 465 sur ceux des organisations.

Un communiqué a-t-il encore besoin d'une ligne de date ? Sur le site d'un diffuseur, c'est la norme : 76 % des communiqués s'ouvrent ainsi. Sur le site de l'organisation, 17 % (11 % sans les sites sportifs). Ce sont des estimations : la règle de la ligne de date n'a pas atteint la fiabilité exigée par l'enregistrement, et elle manque des lignes de date plus souvent qu'elle n'en invente. La ligne de date appartient à la copie qui part sur le fil.

Faut-il mettre un contact médias dans un communiqué ? C'est le seul élément écrit pour le journaliste seul, et la copie publiée sur le site de l'organisation l'omet le plus souvent : tel que mesuré, 14 % portent un courriel ou un numéro de téléphone (23 % sans les sites sportifs), contre 65 % sur les sites des diffuseurs. Ce sont des estimations : la règle du contact n'a pas atteint la fiabilité exigée par l'enregistrement. Une salle de presse sans nom à appeler renvoie le journaliste vers la copie du diffuseur — ou ailleurs.

La mention POUR DIFFUSION IMMÉDIATE est-elle encore utilisée ? Rarement : 3,2 % des communiqués la portent, et 77 % d'entre eux viennent d'un seul site, celui du département de la Justice des États-Unis. Sans lui, la part tombe à 0,8 %, et ce sont alors les gouvernements régionaux et locaux qui la gardent le plus (13,0 %) ; sur les sites des diffuseurs, elle est presque absente (0,3 %).

Les citations de dirigeants sont-elles plus émotives que celles des officiels ? Oui, sans que ce soit la règle : 11,8 % des citations de dirigeants des communiqués commerciaux contiennent pleased, excited, thrilled, delighted, proud ou un mot semblable, contre 4,1 % des citations d'officiels des organismes publics.

4.3 Limites

- La population est celle des sources d'une archive, choisies pour un service commercial de veille, sur un trimestre, en anglais. Les diffuseurs portent plus de la moitié de ses communiqués ; les organisations qui publient seulement sur leurs propres sites, et celles que l'archive ne suit pas, sont sous-représentées. Les résultats sont donnés par type d'éditeur pour qu'aucun groupe n'en cache un autre.
- Tout ce que publient les sources de l'archive n'est pas un communiqué au sens strict. Les sites sportifs portent 38 % du poids des sites propres et publient surtout des articles ; sans eux, 4,3 % des communiqués des sites propres ont une section « À propos », 11 % une ligne de date et 23 % un contact (contre 3 %, 17 % et 14 %), 52 % de tous les communiqués portent un contact (intervalle à 95 % : 33 % à 64 %), et l'écart de H1 est de 204 mots (122 à 269). Le partage entre les deux types de sites tient. Ces vérifications ont été faites après la publication et n'étaient pas enregistrées.
- Deux sites de diffuseurs portent 68 % du poids des diffuseurs : les chiffres des diffuseurs les décrivent avant tout ; sans les cinq plus grands sites (R4), H1 et H4 restent confirmées.
- Les communiqués derrière des tests anti-robots, des certificats invalides, refusés à notre client ou sur des sites dont le robots.txt n'a pu être lu sont comptés, pas devinés. Les pondérations rendent à chaque site sa part, mais ne peuvent rendre un site dont rien n'a été mesuré — 17 %

des communiqués de la population sont sur de tels sites : 19 % de ceux des diffuseurs, presque tous sur deux sites, et 20 % de ceux des gouvernements nationaux, dont la moitié sur un site d'actualités militaires — et quand peu de communiqués d'un site ont pu être lus, ce petit nombre vaut pour tous.

- Les types d'éditeur sont ceux de l'archive, et quelques sites sont mal classés : le gouvernement de l'Alberta et le Bureau australien de la statistique parmi les entreprises (5 % du poids de ce groupe), le CERN et le SANS Internet Storm Center parmi les universités. Une page jointe avec et sans www compte sur deux sites (245 pages mesurées). Les deux sont laissés tels qu'enregistrés.
- Les mesures sont des règles. Celles de la ligne de date et du contact n'ont pas atteint la fiabilité exigée sur le tirage de test, et les hypothèses qui en dépendent ne sont pas tranchées.
- Le seul codeur est un modèle de langage. Ses codages seront déposés avec les données pour que chacun puisse les vérifier contre les pages.
- L'extraction est imparfaite : sur environ une page de test sur cinq, le texte extrait n'était qu'une partie du communiqué.
- L'étude est financée et publiée par une entreprise qui vend de la veille de communiqués ; les résultats rapportent les diffuseurs en agrégat et ne nomment aucun de leurs sites, sauf celui de l'éditeur.

5. Conclusions

Il y a désormais deux communiqués. La copie publiée sur le site d'un diffuseur garde les éléments du fil de presse : une ligne de date, une section « À propos », un contact à joindre. La copie publiée sur le site de l'organisation, le plus souvent, non. L'écart de longueur tient surtout à cet appareil, pas au récit.

Préenregistrement. Les hypothèses, la population, les mesures et les analyses ont été enregistrées le 11 octobre 2026, avant la lecture de toute page de la population (commit 4b769927). Le manifeste des empreintes des fichiers enregistrés a été soumis à OpenTimestamps au début du recensement ; le bloc Bitcoin qui l'ancre (970 932) a été miné 21 minutes après la première page demandée, et le manifeste et sa preuve seront déposés avec les données. L'enregistrement, son addenda, le guide de codage et le journal des écarts sont publiés sans modification. <https://www.ppnworld.com/research/anatomy-of-a-press-release/registration/press-release-anatomy.md>

Disponibilité des données. Le tableau d'anatomie par type d'éditeur est publié en CSV sous licence CC BY 4.0 ; ses groupes gardent les noms de l'archive (wire : diffuseurs ; government : gouvernements nationaux ; municipality : gouvernements régionaux et locaux ; league : sport), et ses colonnes de ligne de date, de contact, de sous-titre, de signature et de marque de fin portent les limites de fiabilité exposées plus haut. Les données par communiqué — chaque communiqué de la population avec son adresse, son site, son type d'éditeur, son résultat, sa pondération et ses mesures, sans texte de page, citation ni coordonnée — seront déposées sur Zenodo avec le code de chaque étape ; d'ici là, le Centre de recherche les envoie sur demande. L'enregistrement, son addenda, le guide de codage, les consignes du codeur A et le journal des écarts sont publiés sans modification. <https://www.ppnworld.com/research/anatomy-of-a-press-release.csv>

Déclaration éthique. L'étude a lu des pages web publiques et n'a impliqué aucun participant humain ; aucun examen éthique n'a été demandé. Elle a respecté le robots.txt de chaque site, s'est d'abord identifiée puis a réessayé une fois comme navigateur en cas de refus (une ligne de robustesse exclut ces pages), a limité sa cadence, et n'a jamais résolu un test anti-robot ni contourné une erreur de certificat. Les blocs de contact contiennent des données personnelles : seuls leur présence et leur type entrent dans les données ; les pages conservées pour refaire les mesures ne sont pas publiées.

Utilisation d'outils d'IA. Claude (Anthropic) a été le codeur A de la validation et, utilisé par Claude Code avec Claude Opus 5.5, a préparé les logiciels de collecte et d'analyse, des ébauches du texte et cette mise en page, sous la supervision de l'auteur. L'auteur est responsable du contenu de l'étude.

Corrections. Corrigée le 17 octobre 2026, après la publication du 15 octobre 2026 ; les corrections sont listées, datées, sur la page web de l'étude, et aucune estimation des quatre hypothèses n'a changé. <https://www.ppnworld.com/fr/research/anatomy-of-a-press-release>

Références

- Associated Press. (2026). The Associated Press stylebook 2026–2028 (58th ed.). Basic Books.
- Barbaresi, A. (2021). Trafilatura: A web scraping library and command-line tool for text discovery and extraction. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 122–131. <https://doi.org/10.18653/v1/2021.acl-demo.15>
- Bevendorff, J., Gupta, S., Kiesel, J., & Stein, B. (2023). An empirical comparison of web content extraction algorithms. Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. <https://doi.org/10.1145/3539618.3591920>
- Bossema, F. G., Burger, P., Bratton, L., Challenger, A., Adams, R. C., Sumner, P., Schat, J., Numans, M. E., & Smeets, I. (2019). Expert quotes and exaggeration in health news: A retrospective quantitative content analysis. Wellcome Open Research, 4, 56. <https://doi.org/10.12688/wellcomeopenres.15147.2>
- Catenaccio, P. (2008). Press releases as a hybrid genre: Addressing the informative/promotional conundrum. Pragmatics, 18(1), 9–31. <https://doi.org/10.1075/prag.18.1.02cat>
- Cision. (2025, May 7). 2025 State of the Media report [Survey of more than 3,000 journalists in 19 markets].
- Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. Proceedings of the National Academy of Sciences, 120(30), e2305016120. <https://doi.org/10.1073/pnas.2305016120>
- Hiebert, R. E. (1966). Courtier to the crowd: The story of Ivy Lee and the development of public relations. Iowa State University Press.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2), 65–70. <https://www.jstor.org/stable/4615733>
- Jacobs, G. (1999). Preformulating the news: An analysis of the metapragmatics of press releases. John Benjamins. <https://doi.org/10.1075/pbns.60>
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas for Navy enlisted personnel (Research Branch Report 8-75). Naval Technical Training Command.
- Koster, M., Illyes, G., Zeller, H., & Sassman, L. (2022, September). Robots Exclusion Protocol (RFC 9309). Internet Engineering Task Force. <https://doi.org/10.17487/RFC9309>
- Krippendorff, K. (2004). Content analysis: An introduction to its methodology (2nd ed.). Sage.
- Lassen, I. (2006). Is the press release a genre? A study of form and content. Discourse Studies, 8(4), 503–530. <https://doi.org/10.1177/1461445606064830>
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability.

- Human Communication Research, 28(4), 587–604.
<https://doi.org/10.1111/j.1468-2958.2002.tb00826.x>
- McLaren, Y., & Gurău, C. (2005). Characterising the genre of the corporate press release. *LSP and Professional Communication*, 5(1), 10–30.
- Nicholls, T. (2019). Detecting textual reuse in news stories, at scale. *International Journal of Communication*, 13, 4173–4197.
- PR Gun. (2026, August 25). [Analysis of 138,901 press releases from ten newswires, 2021–2026] [Press release].
- Sumner, P., Vivian-Griffiths, S., Boivin, J., Williams, A., Venetis, C. A., Davies, A., Ogden, J., Whelan, L., Hughes, B., Dalton, B., Boy, F., & Chambers, C. D. (2014). The association between exaggeration in health related science news and academic press releases: Retrospective observational study. *BMJ*, 349, g7015. <https://doi.org/10.1136/bmj.g7015>
- Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for increasing the transparency of analysis of preexisting data sets. *Advances in Methods and Practices in Psychological Science*, 2(3), 214–227.
<https://doi.org/10.1177/2515245919848684>

Annexe. Données des figures

Chaque tableau donne les chiffres tracés dans la figure du même numéro.

Tableau A1. Deux communiqués

Élément	Diffuseurs	Sites propres
POUR DIFFUSION IMMÉDIATE	0,3 %	6,9 %
Sous-titre	37,0 %	7,9 %
Signature d'une personne	12,0 %	5,1 %
Ligne de date	76,5 %	16,7 %
Une citation	59,0 %	41,8 %
Section « À propos »	63,9 %	2,9 %
Contact avec courriel ou téléphone	65,1 %	14,2 %
Marque de fin (### ou -30-)	2,7 %	4,0 %

Tableau A2. Jusqu'où chaque mesure est fiable

Mesure	Accord, règle enregistrée	Accord, règle révisée	α , révisée	Peut trancher
Ligne de date	87 %	87 %	,739	rapportée seulement
Section « À propos »	91 %	91 %	,811	tranche
Contact	74 %	89 %	,779	rapportée seulement
Citation	96 %	97 %	,940	tranche
Dirigeant cité	94 %	96 %	,910	tranche
Mot d'émotion	98 %	99 %	,928	tranche
POUR DIFFUSION IMMÉDIATE	100 %	100 %	1,000	tranche
Sous-titre	69 %	69 %	,326	rapportée seulement
Marque de fin	97 %	97 %	,558	rapportée seulement
Signature	92 %	92 %	,162	rapportée seulement