# BioSTransformers : nouveaux modèles neuronaux siamois pour l'apprentissage sans exemple de textes biomédicaux

Safaa Menad *, Saïd Abdeddaim*, Lina F. Soualmia*

* TIBS-LITIS UR4108, Université de Rouen Normandie, 76000 Rouen, France
{safaa.menad1, said.abdeddaim, soualfat}@univ-rouen.fr

**Résumé.** L'entraînement de modèles transformeurs de langages sur des données biomédicales a permis d'obtenir des résultats prometteurs. Cependant, ces modèles de langage nécessitent pour chaque tâche un affinement (fine-tuning) sur des données supervisées très spécifiques qui sont peu disponibles dans le domaine biomédical. Nous proposons d'utiliser des modèles neuronaux siamois (sentence transformers) qui plongent des textes à comparer dans un espace vectoriel pour deux tâches: la classification d'articles scientifiques et les réponses aux questions biomédicales. Nos modèles optimisent une fonction objectif d'apprentissage contrastif auto-supervisé sur des articles issus de la base de données bibliographique MEDLINE associés à leurs mots-clés MeSH (Medical Subject Headings). Les résultats obtenus sur plusieurs benchmarks montrent que les modèles proposés permettent de résoudre ces tâches sans exemples (zero-shot) et sont comparables à des modèles transformeurs biomédicaux affinés sur des données supervisés spécifiques aux problèmes traités [1].

## 1 Introduction

Le développement de modèles transformeurs pré-entraînés, tels que BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019), a permis d'améliorer les performances du traitement automatique du langage (TAL). L'abondance de données biomédicales disponibles, comme les articles scientifiques, a aussi rendu possible l'entraînement de ces modèles sur des corpus pour des applications biomédicales (Alsentzer et al., 2019; Lee et al., 2020; Liu et al., 2021). Ces modèles de langage nécessitent cependant un affinement (fine-tuning) pour chaque tâche sur des données supervisées très spécifiques et rarement disponibles, ce qui limite fortement leur usage en pratique. Comme la plupart des tâches de TAL biomédical (e.g., extraction de relations, classification de documents, questions-réponses) peuvent se réduire au calcul d'une mesure de similarité sémantique entre deux textes (p. ex. catégorie/résumé d'un article, requête/résultats, question/réponse), nous proposons dans cet article de construire un nouveau modèle transformeur siamois BioSTransformers (sentence transformer) pré-entraîné qui plonge des paires de textes sémantiquement liés (longs et courts) dans un même espace de représentation vectoriel. En plus d'être applicable à plusieurs types de tâches de TAL, un modèle siamois a aussi l'avantage de permettre de gagner du temps lors

1. Modèles et données disponibles : `https://github.com/arieme/BioSTransformers.git`

de son utilisation en précalculant les représentations vectorielles des textes. Par exemple en recherche documentaire, un modèle siamois peut permettre de précalculer et d'indexer les représentations vectorielles des textes du corpus ciblé pour n'en calculer que la représentation des requêtes lorsqu'elles sont soumises au moteur, contrairement aux modèles transformeurs affinés qui prennent en entrée la combinaison de toutes les paires de textes à comparer. Grâce à ce modèle, nous souhaitons : i) éviter les coûts engendrés par l'étiquetage des données, les calculs d'entraînement et d'affinement ; et ii) réduire considérablement ceux de la prédiction en proposant un modèle auto-supervisé de référence directement applicable à un large éventail de tâches biomédicales.

Dans ce cadre, nous comparons plusieurs modèles transformeurs siamois que nous avons entraînés sur des paires de textes formées, d'une part, de résumés du corpus d'articles biomédicaux PubMed [2], et d'autre part, des mots-clés MeSH (Medical Subject Headings) [3] qui leur sont associés. Nous utilisons une fonction objectif d'apprentissage contrastif auto-supervisé. Étant donné une paire de textes (résumé, mots-clés), le modèle doit prédire laquelle, parmi un ensemble d'autres paires de textes échantillonnées au hasard, lui est réellement associée dans PubMed. Nous montrons ensuite expérimentalement sur plusieurs benchmarks biomédicaux que sans affinement pour une tâche spécifique, notre meilleur modèle siamois pré-entraîné permet de résoudre sans exemples d'apprentissage (zero shot) deux tâches de TAL avec des résultats comparables aux modèles transformeurs biomédicaux ou encore généralistes affinés sur des données supervisées spécifiques aux problèmes traités. Deux tâches sont ciblées : la classification de documents et les réponses aux questions.

La section 2 présente les modèles transformeurs pré-entraînés ainsi que leur utilisation dans des modèles siamois. La section 3 décrit les modèles siamois que nous proposons dans ce travail. Enfin, la section 4 présente les résultats obtenus sur des benchmarks de référence.

## 2 Les transformeurs

Les transformeurs sont des réseaux neuronaux basés sur le mécanisme d'auto-attention multi-têtes qui améliore considérablement l'efficacité de l'apprentissage des modèles de grande taille. Il est composé d'un encodeur qui transforme le texte d'entrée en vecteur, et d'un décodeur qui transforme ce vecteur en texte en sortie. Le mécanisme d'attention fournit de meilleures performances dans ces modèles grâce à la modélisation des liens entre les éléments d'entrée et de sortie. Un modèle de langage pré-entraîné (MLP) est un réseau neuronal entraîné sur une grande quantité de données non annotées de manière non supervisée. Le modèle est ensuite transféré pour une tâche de TAL cible (downstream task), où un ensemble de données annotées plus petit et spécifique à la tâche est utilisé pour affiner le MLP permettant ainsi de construire le modèle final capable d'exécuter la tâche cible. C'est ce qu'on appelle l'ajustement d'un MLP.

2. `https://ftp.ncbi.nlm.nih.gov/pubmed/baseline/`

3. Le MeSH (`https://www.nlm.nih.gov/mesh/`) est un thésaurus spécialisé du domaine biomédical composé de 30000 termes utilisés pour l'indexation d'articles PubMed.

### 2.1 Modèles pré-entraînés

Les modèles de langage pré-entraînés, tels que BERT, ont conduit à des gains impressionnants dans de nombreuses tâches de TAL. Les travaux existants se concentrent généralement sur les données généralistes. Dans le domaine biomédical, le pré-entraînement sur les textes de PubMed permet d'obtenir de meilleures performances dans les tâches du TAL biomédicales (Beltagy et al., 2019; Lee et al., 2020; Peng et al., 2019a). L'approche standard de pré-entraînement d'un modèle biomédical commence avec un modèle généraliste et poursuit le pré-entraînement en utilisant un corpus biomédical. BioBERT (Lee et al., 2020) utilise pour cela les résumés extraits de PubMed et les articles en texte intégral de PubMed Central (PMC). BlueBERT (Peng et al., 2019b) utilise à la fois le texte de PubMed et les notes cliniques MIMIC-III (Medical Information Mart for Intensive Care) (Johnson et al., 2016). SciBERT (Beltagy et al., 2019) constitue une exception, le pré-entraînement est fait à partir de zéro, en utilisant la littérature scientifique.

### 2.2 Modèles siamois

Les transformeurs de paires de phrases (sentence-transformers) ont été développés pour la tâche de calcul d'un score de similarité entre deux phrases. C'est des modèles qui utilisent des transformeurs pour des tâches liées aux paires de phrases : calcul de similarité sémantique entre phrases, recherche d'informations, reformulation de phrases etc. Ces transformeurs sont basés sur deux architectures : les cross-encodeurs qui traitent la concaténation de la paire et les modèles siamois bi-encodeurs qui encodent en vecteur chacun des éléments de la paire. Sentence-BERT (Reimers et Gurevych, 2019) est un bi-encodeur basé sur BERT permettant de générer des plongements de phrases sémantiquement significatifs à utiliser dans des comparaisons de similarité textuelle. Pour chaque entrée, le modèle produit un vecteur de taille fixe ($u$ et $v$). La fonction objectif est choisie de façon à ce que l'angle entre les deux vecteurs $u$ et $v$ est d'autant plus faible que les entrées sont similaires. Plus précisément la fonction objectif utilise le cosinus de l'angle : $cos(u, v) = \frac{u.v}{||u||||v||}$, si $cos(u, v) = 1$, les phrases sont similaires et si $cos(u, v) = 0$, les phrases n'ont aucune relation sémantique.

D'autres modèles de plongements de phrases ont été développés (Gao et al., 2021; Wang et al., 2021; Cohan et al., 2020), parmi eux MiniLM-L6-v25 [4] est un bi-encodeur basé sur une version simplifiée de MiniLM (Wang et al., 2020). Ce modèle rapide et de petite taille a donné de bonnes performances sur différentes tâches pour 56 corpus (Muennighoff et al., 2022).

## 3 Modèles de langage proposés

Les transformeurs siamois donnent de bons résultats dans des domaines généralistes, mais pas dans les domaines de spécialité, comme le domaine biomédical (Muennighoff et al., 2022). Nous proposons ici de nouveaux modèles siamois pré-entraînés sur le corpus PubMed. Les transformeurs siamois ont été initialement conçus pour transformer des phrases (de taille similaire) en vecteurs. Nous proposons dans notre approche de transformer dans le même espace vectoriel les termes MeSH, les titres et les résumés des articles PubMed en entraînant un modèle de transformeur siamois sur ces données. Nous voulons nous assurer qu'il y a une

4. https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

correspondance dans cet espace vectoriel entre le texte court et le texte long. Nous avons donc entraîné nos modèles avec des paires d'entrées (titre, terme MeSH) et (résumé, terme MeSH).

Sur ces données nous avons construit deux types de modèles : le premier type est notre propre transformeur siamois (BioSTransformer) construit à partir d'un transformeur pré-entraîné sur des données biomédicales et le second est un transformeur siamois déjà pré-entraîné sur des données généralistes (BioS-MiniLM).

**BioSTransformers**. Pour construire les BioSTransformers, nous nous sommes inspiré du modèle Sentence-BERT (Reimers et Gurevych, 2019) en remplaçant BERT par d'autres transformeurs. Nous avons utilisé des transformeurs qui ont été entraînés sur des données biomédicales (bio-transformeurs) pour créer des transformeurs siamois en ajoutant une couche de pooling et en changeant la fonction objectif. La couche de pooling calcule le vecteur moyen des vecteurs de sortie du transformeur (token embeddings). Les deux textes en entrée passent successivement dans le transformeur produisant deux vecteurs $u$ et $v$ en sortie du pooling qui sont par la suite utilisés par la fonction objectif. Nous avons sélectionné trois bio-transformeurs BlueBERT (Peng et al., 2019b), PubMed BERT (Gu et al., 2022) et BioELECTRA (Kanakarajan et al., 2021). Ces modèles ont été entraînés sur PubMed à part BlueBERT qui a également été entraîné sur les notes cliniques.

**BioS-MiniLM**. Pour ce modèle nous avons utilisé un transformeur siamois pré-entraîné sur des données généralistes puis nous l'avons entraîné sur nos données. Plusieurs modèles généraux de sentence-transformer déjà pré-entraînés sont disponibles [5]. Ils diffèrent en taille, vitesse et performance. Dans ceux qui obtiennent les meilleures performances, nous avons utilisé MiniLM-L6-v2 (voir section 2) qui a été pré-entraîné sur 32 corpus généralistes (Reddit comments, S2ORC, WikiAnswers etc.).

**Fonction objectif**. Pour un transformeur de paires de phrases classique on dispose de données supervisées sous forme de triplets (phrase 1, phrase 2, score de similarité entre les deux phrases). Cependant, dans notre cas, nous ne disposons d'aucun score pour les résumés ou les titres et leurs termes MeSH correspondants. Nous considérons donc que :

— un résumé, un titre et les termes MeSH associés à un même article (identifié par un PMID) sont similaires et que le score est égal à 1 ;
— un résumé ou un titre avec des termes MeSH qui ne sont pas associés au même article ne sont pas similaires et que le score est donc égal à 0.

Nous utilisons une fonction objectif d'apprentissage contrastif auto-supervisé basée sur la fonction de perte de classement négatif multiple (Henderson et al., 2017) dite MNRL (Multiple Negative Ranking Loss) dans le package Sentence-Transformers [6]. La MNRL n'a besoin que des paires positives en entrée (le titre ou le résumé et un terme MeSH associé à l'article dans notre cas). Pour une paire positive (titre_$i$ ou résumé_$i$, MeSH_$i$), la MNRL considère que chaque paire (titre_$i$ ou résumé_$i$, MeSH_$j$) avec $i \neq j$ dans le même batch est négative. Comme un article peut être associé à plusieurs termes MeSH, nous avons fait en sorte dans la génération des batchs qu'un résumé (ou un titre) associé à un terme MeSH dans PubMed ne soient jamais pris comme une paire négative.

5. https://huggingface.co/sentence-transformers
6. https://www.sbert.net/docs/package_reference/losses.html\#multiplenegativesrankingloss

# 4 Expérimentations et résultats

## 4.1 Expérimentations

Dans un premier temps pour tester rapidement les différents transformeurs et la fonction objectif à choisir nous n'avons utilisé que les titres et nous avons réduit le nombre de termes MeSH. Nous avons sélectionné au total 1 402 termes MeSH et 3,79 millions de paires (titre, MeSH) et nous avons utilisé 18 940 articles avec leurs titres et termes MeSH pour la validation.

Dans un second temps, une fois sélectionnés les modèles transformeurs et la fonction objectif MNRL, nous avons évalué nos modèles BioSTransformers et BioS-MiniLM sur les paires (titre, MeSH) et (résumé, MeSH) générés à partir de tous les termes MeSH utilisés dans PubMed. Ayant constaté qu'il n'est pas nécessaire d'utiliser toutes les paires des 35 millions d'articles de PubMed, nous avons sélectionné 6,75 millions de paires pour le fine-tuning. 18 557 articles ont été utilisés pour la validation.

## 4.2 Résultats

Les deux tâches de TAL ainsi que les données utilisées sont décrites ci-après :

1. La classification de documents : le corpus Hallmarks of Cancer (HOC) est constitué de 1852 résumés de publications PubMed annotés manuellement par des experts selon une taxonomie qui est composée de 37 classes. Chaque phrase du corpus se voit attribuer zéro à plusieurs classes (Hanahan et Weinberg, 2000).
2. Les réponses aux questions (QA) :
   (a) PubMedQA : un corpus pour les réponses aux questions spécifiques à la recherche biomédicale. Il contient un ensemble de questions, ainsi qu'un champ annoté indiquant si le texte contient la réponse à la question de recherche (Jin et al., 2019).
   (b) BioASQ : un corpus qui contient plusieurs tâches de QA avec des données annotées par des experts, y compris des questions oui/non, de liste et de résumés. Nous nous concentrons sur le type de questions oui/non (tâche 7b) (Nentidis et al., 2019).

Nous considérons les deux tâches (classification de documents et réponse aux questions) comme un problème de similarité de textes et nous cherchons à retrouver pour chaque requête les résultats les plus proches. Nous considérons les $k$ résultats les plus proches de chaque requête, $k$ étant le nombre de résultats attribués à la requête par l'expert. La similarité entre la requête et les résultats est mesurée par la similarité cosinus entre le vecteur de la requête et les vecteurs des résultats. Dans une tâche de classification, la requête est la catégorie et les résultats sont les documents classés dans cette catégorie. Dans une tâche de réponse aux questions, la requête est la question et les résultats sont une réponse. Nous avons évalué nos modèles selon le score F1 utilisé dans les benchmarks : Hallmarks of Cancer (HoC) (Hanahan et Weinberg, 2000), PubmedQA (Jin et al., 2019) et BioASQ (Nentidis et al., 2019) dans (Gu et al., 2022). Les résultats obtenus par nos modèles transformeurs siamois sans exemple (sans fine-tuning) sont donnés dans le Tableau 1.

Le Tableau 2 montre les résultats obtenus sur les mêmes tâches par des modèles affinés spécifiquement à ces tâches (Gu et al., 2022). Pour chaque benchmark, ces modèles sont affinés avec les données supervisées disponibles dans chaque cas. Ces résultats montrent que les modèles que nous proposons permettent de résoudre ces tâches de façon comparable à des

| Corpus / Modèle | BioS-MiniLM | S-BioELECTRA | S-PubMedBERT | S-BlueBERT |
|---|---|---|---|---|
| HoC | 0,492 | **0,499** | 0,489 | 0,468 |
| PubMedQA | 0,649 | 0,675 | **0,729** | 0,652 |
| BioASQ | 0,747 | 0,694 | **0,751** | 0,713 |

TAB. 1 – *Résultats d'évaluation de nos modèles sur différents benchmarks selon le F1 score.*

modèles biomédicaux affinés sur des données supervisés spécifiques aux problèmes traités que nous n'avons pas utilisées dans notre approche sans exemple.

| Corpus / Modèle | BERT +affinement | RoBERTa +affinement | BioBERT +affinement | SciBERT +affinement | ClinicalBERT +affinement | BlueBERT +affinement | PubMedBERT +affinement |
|---|---|---|---|---|---|---|---|
| HoC | 0.802 | 0.797 | 0.815 | 0.812 | 0.807 | 0.805 | **0.823** |
| PubmedQA | 0.516 | 0.528 | **0.602** | 0.574 | 0.491 | 0.484 | 0.558 |
| BioASQ | 0.744 | 0.752 | 0.841 | 0.789 | 0.685 | 0.687 | **0.876** |

TAB. 2 – *Résultats d'évaluation des modèles affinés spécifiquement à ces tâches sur différents benchmarks selon le F1 score (Gu et al., 2022).*

Pour le benchmark HoC, les résultats obtenus par notre meilleur modèle S-BioELECTRA sont très en dessous des résultats obtenus par PubMedBERT+affinement (0,499 vs. 0,823). En effet, les modèles de (Gu et al., 2022) ont été affinés spécifiquement pour chaque tâche, notamment la classification des documents, en modifiant l'architecture du modèle et en ajoutant des couches spécifiques pour chaque cas.

En revanche pour le benchmark PubMedQA, les résultats obtenus par notre meilleur modèle S-PubMedBERT sont meilleurs que les résultats obtenus par BioBERT+affinement (0,729 vs. 0,602). Enfin, pour le benchmark BioASQ, les résultats obtenus par notre meilleur modèle S-PubMedBERT sont comparables aux résultats obtenus par les modèles affinés même si PubMedBERT+affinement donne de meilleurs résultats (0,751 vs. 0,876). Et tout cela, sans réadapter l'architecture de nos modèles pour chaque tâche et sans les affiner sur les données spécifiques aux benchmarks cités.

# 5 Conclusion

Dans cet article, nous avons proposé de nouveaux modèles siamois BioSTransformers et BioS-MiniLM qui permettent de résoudre des tâches sans exemple dans des textes biomédicaux. Ces modèles siamois plongent les paires de textes dans un même espace de représentation et permettent de calculer la proximité sémantique entre textes de différentes longueurs. Nos résultats montrent sur plusieurs corpus qu'avec un apprentissage sans exemple, nos BioSTransformers et particulièrement S-PubMedBERT arrivent à surpasser des modèles de l'état de l'art qui sont déjà entraînés sur ces données et pour ces tâches spécifiques. Nous envisageons d'améliorer nos modèles pour qu'ils puissent mieux répondre à la tâche de classification. Nos modèles auto-supervisés sont des modèles de référence qui pourraient être directement appliqués à d'autres tâches de TAL.

# Références

Alsentzer, E., J. Murphy, W. Boag, W.-H. Weng, D. Jindi, T. Naumann, et M. McDermott (2019). Publicly available clinical BERT embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, Minneapolis, Minnesota, USA, pp. 72–78. Association for Computational Linguistics.

Beltagy, I., K. Lo, et A. Cohan (2019). SciBERT : A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 3615–3620.

Cohan, A., S. Feldman, I. Beltagy, D. Downey, et D. S. Weld (2020). Specter : Document-level representation learning using citation-informed transformers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2270–2282.

Devlin, J., M.-W. Chang, K. Lee, et K. Toutanova (2019). BERT : Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pp. 4171–4186.

Gao, T., X. Yao, et D. Chen (2021). Simcse : Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 6894–6910.

Gu, Y., R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, et H. Poon (2022). Domain-specific language model pretraining for biomedical natural language processing. *ACM Transactions on Computing for Healthcare 3*(1), 1–23.

Hanahan, D. et R. A. Weinberg (2000). The hallmarks of cancer. *Cell 100*(1), 57–70.

Henderson, M., R. Al-Rfou, B. Strope, Y.-H. Sung, L. Lukács, R. Guo, S. Kumar, B. Miklos, et R. Kurzweil (2017). Efficient natural language response suggestion for smart reply. *arXiv preprint arXiv :1705.00652*.

Jin, Q., B. Dhingra, Z. Liu, W. Cohen, et X. Lu (2019). PubMedQA : A dataset for biomedical research question answering. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 2567–2577.

Johnson, A. E., T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, et R. G. Mark (2016). MIMIC-III, a freely accessible critical care database. *Scientific data 3*(1), 1–9.

Kanakarajan, K. r., B. Kundumani, et M. Sankarasubbu (2021). BioELECTRA : Pretrained biomedical text encoder using discriminators. In *Proceedings of the 20th Workshop on Biomedical Language Processing*, Online, pp. 143–154. Association for Computational Linguistics.

Lee, J., W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, et J. Kang (2020). BioBERT : a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics 36*(4), 1234–1240.

Liu, F., E. Shareghi, Z. Meng, M. Basaldella, et N. Collier (2021). Self-alignment pretraining for biomedical entity representations. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language*

*Technologies*, pp. 4228–4238.

Muennighoff, N., N. Tazi, L. Magne, et N. Reimers (2022). Mteb : Massive text embedding benchmark. *arXiv preprint arXiv :2210.07316*.

Nentidis, A., K. Bougiatiotis, A. Krithara, et G. Paliouras (2019). Results of the seventh edition of the BioASQ challenge. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 553–568. Springer.

Peng, Y., S. Yan, et Z. Lu (2019a). Transfer learning in biomedical natural language processing : An evaluation of BERT and ELMo on ten benchmarking datasets. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pp. 58–65.

Peng, Y., S. Yan, et Z. Lu (2019b). Transfer learning in biomedical natural language processing : An evaluation of BERT and ELMo on ten benchmarking datasets. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, Florence, Italy, pp. 58–65. Association for Computational Linguistics.

Reimers, N. et I. Gurevych (2019). Sentence-BERT : Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, pp. 3982–3992. Association for Computational Linguistics.

Wang, K., N. Reimers, et I. Gurevych (2021). Tsdae : Using transformer-based sequential denoising auto-encoderfor unsupervised sentence embedding learning. In *Findings of the Association for Computational Linguistics : EMNLP 2021*, pp. 671–688.

Wang, W., F. Wei, L. Dong, H. Bao, N. Yang, et M. Zhou (2020). Minilm : Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Advances in Neural Information Processing Systems 33*, 5776–5788.

## Summary

Training language transformers on biomedical data has shown promising results. However, these language models require fine-tuning on very specific supervised data for each task, which are rarely available in the biomedical domain. We propose to use siamese neural models (sentence transformers) that embed texts to be compared in a vector space, and apply them on two main tasks: biomedical classification of articles and question answering. Our models optimize an objective self-supervised contrastive learning function on articles from the MEDLINE bibliographic database associated with their MeSH (Medical Subject Headings) keywords. The obtained results on several benchmarks show that the proposed models can solve these tasks without examples (zero-shot) and are comparable to biomedical transformers fine-tuned on supervised data specific to the problems treated.